**UNIVERSIDAD DE EXTREMADURA**

**Departamento de Tecnología de los Computadores y de las Comunicaciones**

TESIS DOCTORAL

# Percepción dinámica del entorno en un robot móvil

Autora: Pilar Bachiller Burgos
Director: Pablo Bustos García de Castro

· Abril de 2008 ·

D. Pablo Bustos García de Castro, Titular de Universidad del Departamento de Tecnología de los Computadores y de las Comunicaciones de la Universidad de Extremadura

**CERTIFICA**

Que Dª Pilar Bachiller Burgos, Ingeniera en Informática, ha realizado en el Departamento de Tecnología de los Computadores y de las Comunicaciones, bajo mi dirección, el trabajo de investigación correspondiente a su Tesis Doctoral titulada:

**"Percepción dinámica del entorno en un robot móvil"**

Revisado el presente trabajo, estimo que puede ser presentado al tribunal que ha de juzgarlo y autorizo la presentación de esta Tesis Doctoral en la Universidad de Extremadura.

Cáceres, a 25 de abril de 2008

Fdo. D. Pablo Bustos García de Castro
Titular de Universidad
Departamento de Tecnología
de los Computadores
y de las Comunicaciones
Universidad de Extremadura

*A mis grandes maestros, mis padres*

# Agradecimientos

Aunque el escribir estas líneas me hace ser más consciente de que el trabajo de muchos años por fin va a ver la luz, he de reconocer que aún me cuesta creerlo del todo. Estoy segura que los que han estado a mi lado durante el proceso de elaboración de esta tesis entienden muy bien que tenga esa sensación. El camino ha sido apasionante, pero no ha estado libre de dificultades y es el momento de dar las gracias a todos aquellos que me han dado su apoyo de una u otra forma.

Como no podía ser de otra manera, la primera persona a la que debo dar las gracias es a Pablo, director de esta tesis, pero, además, maestro, compañero y amigo. El trabajo a su lado durante estos años ha sido mucho más que el desarrollo de este proyecto. Para mí ha sido sobre todo un aprendizaje continuo y una inquietud constante y maravillosa por todo lo relacionado con la robótica, la inteligencia y la visión.

A la gente del laboratorio, los que ya no están (Lito, Álvaro, Juan, Ricardo) y los que aún siguen (Pepe, Luis, Agustín). Muy en especial a Pepe, por sus ánimos, sus consejos y su apoyo continuo, y a Agustín, sin el que me encontraría completamente perdida ante cualquier problema relacionado con el hardware. Gracias también a Félix porque el tiempo que estuvo con nosotros en el laboratorio fue para mí una oportunidad de aprender como hay pocas y porque sus consejos son siempre de gran ayuda.

A mis compañeros, además de amigos, que no han parado de darme ánimos durante este tiempo: Adolfo, Fernando, Juan, Elena, Antonio, José Miguel, Julia, Encarna y Pedro, entre otros. Debo mencionar especialmente a Alberto, gran amigo con el que sé que cuento siempre, por su apoyo, su interés constante y por su cuidadosa revisión de este documento.

A mi familia, padres, hermanos, sobrinos y cuñados, a los que sé que este momento en mi vida les ilusiona tanto como a mí.

A mis amigos, muy en especial a Mª Luisa, por estar siempre a mi lado y permitirme desahogarme en tantos momentos.

Finalmente, mil gracias a Fran porque sin él esto habría sido mucho más duro, porque con su amor y su comprensión los momentos de agobio han sido más llevaderos y por ser la inspiración que necesitaba para finalizar este proyecto.

## Resumen

Durante los últimos años, la atención se ha convertido en una cuestión de gran interés en visión artificial. Los estudios sobre los mecanismos atencionales en la visión biológica han inspirado numerosos modelos computacionales. La mayoría de ellos se basan en la hipótesis de capacidad limitada asociada con la función de la atención desde algunas propuestas psicológicas. Estas teorías suponen que el sistema visual tiene una capacidad de procesamiento limitada y que la atención actúa como un filtro que selecciona qué información debe ser procesada en cada instante. Dicha suposición ha sido criticada por muchos autores que afirman, a partir de diversos estudios, que la capacidad de procesamiento de los sistemas perceptivos en los humanos es enorme. Desde este punto de vista, la función de un subsistema que seleccione qué información debe ser procesada no es necesaria. En su lugar, justifican el papel de la atención desde la perspectiva de la selección para la acción. De acuerdo con esta nueva concepción, la función de la atención es evitar la desorganización conductual mediante la selección de la información apropiada para la ejecución de una tarea. Esta noción de la atención resulta de gran interés en el campo de la robótica, donde se persigue, como objetivo fundamental, la construcción de robots autónomos que interactúen con entornos complejos, manteniendo múltiples objetivos conductuales. La selección atencional para la acción permitiría guiar los comportamientos del robot centrándose en objetivos visuales de interés y evitando las posibles distracciones. Dando un paso más, podríamos concebir esta forma de atención como un mecanismo de coordinación, dado que permitiría serializar las acciones de los posiblemente múltiples comportamientos activos. Para explotar estas ideas, en esta tesis se propone un sistema de atención visual basado en la teoría de la selección para la acción. El sistema ha sido diseñado y probado en un robot móvil dotado de una torreta de visión estéreo.

El sistema propuesto es modelado como un grupo de componentes de procesamiento específico que colaboran para seleccionar, fijar y mantener objetivos visuales de acuerdo con diferentes necesidades de actuación. Los componentes de bajo nivel están relacionados con la adquisición de imágenes y el control motor, así como con la detección y el mantenimiento de regiones de interés del entorno. Los componentes del nivel intermedio se encargan de extraer grupos de características de cada región de interés relacionadas con diversas cuestiones sobre el "qué" (información de aspecto) y el "cómo" (información espacial). Estas características son utilizadas por componentes de alto nivel, a los que denominamos selectores de objetivo, para dirigir la atención en función de ciertas especificaciones conductuales. El control atencional no está centralizado, sino que se encuentra distribuido entre varios selectores de objetivo. Cada uno de ellos dirige la atención a partir de diferentes especificaciones para centrarse en diferentes tipos de objetivos visuales. En un momento determinado, la atención abierta es dirigida por un único selector, mientras que los restantes atienden de manera encubierta a sus correspondientes objetivos. La frecuencia de adquisición del control abierto de la atención en cada selector es modulada por unidades de comportamiento de alto nivel de acuerdo con sus necesidades de información. Tras su selección, un objetivo es foveatizado a partir de dos movimientos independientes de las cámaras: un movimiento sacádico y de seguimiento en una de las cámaras y un movimiento asimétrico de vergencia en la otra. Esto permite mantener una fijación binocular estable, aunque el control atencional esté basado en información monocular. Una vez que este proceso senso-motor se completa, el objetivo que constituye el foco abierto de atención es enviado a las unidades de comportamiento. Sólo las acciones compatibles con el foco de atención son entonces ejecutadas, resolviendo el problema de coordinación entre comportamientos. El sistema completo actúa como una arquitectura de control que es atraída hacia diferentes objetivos visuales con finalidad ejecutiva. El intercalado específico entre las acciones viene dado por una relación de tiempos que emerge a partir de parámetros internos y características externas del mundo.

## Abstract

During the last few years, attention has become an important issue in machine vision. Studies of attentional mechanisms in biological vision have inspired many computational models. Most of them follow the assumption of limited capacity associated to the role of attention from psychological proposals. These theories hypothesize that the visual system has limited capacity of processing and that attention acts as a filter that selects the information that should be processed at each time. This assumption has been criticized for many authors who affirm that processing capacity of human perceptual systems is enormous. From this view, there is no need for an stage of selection of the information to be processed. Instead, they claim the role of attention from the perspective of selection for action. According to this new conception, the function of attention is to avoid a behavioral disorganization by selecting the appropriate information to drive task execution. Such a notion of attention is very interesting in robotics where the aim is to build autonomous robots that interact with complex environments, keeping multiple behavioral objectives. Attentional selection for action can guide robot behaviors by focusing on relevant visual targets while avoiding distracters. Moreover, it can be conceived as a coordination mechanism, since it allows serializing the actions of, potentially, multiple active behaviors. To exploit these ideas, we propose a visual attention system based on the selection for action theory. It has been design and tested in a mobile robot endowed with a stereo vision head.

The proposed system has been modeled as a collection of components collaborating to select, fix and track visual targets according to different task requirements. The low level components are related to image acquisition, motor control, as well as computation and maintenance of regions of interest (ROI). Components of intermediate level are in charge of extracting sets of ROI features related to "what" (appearance information) and "how" (spatial information) matters. These features are used by high level components, called target selectors (TS), to drive attention according to certain top-down behavioral specifications. Attention control is not centralized, but distributed among several target selectors. Each of them drives attention from different top-down specifications to focus on different types of visual targets. At a given time, overt attention is driven by one TS, while the rest attends covertly to their corresponding targets. The frequency of overt control of attention of each TS is modulated by high level behavioral units according to their information requirements. The fixation of a selected target is accomplished by two independent camera movements: a saccadic and tracking movement in one of the cameras and a vergence movement in the other. This allows controlling attention from monocular information while keeping stable binocular fixation. Once this perceptual-motor process is completed, the foveated target is sent to the behavioral units. Only actions compatible with the focus of attention are then executed, solving the behavior coordination problem. The whole system works as a control architecture that is attracted towards different visual targets to keep several behavioral goals. The specific interleaving between actions is given by an implicit time relation that links internal parameters and external world features.

# Índice general

# Índice de figuras

# Índice de tablas

# Capítulo 1

# Introducción

La visión es el sentido más eficaz en los humanos, proporcionando sobre el 80 % de la información que recibimos del mundo exterior. Sin lugar a dudas, un robot con esta capacidad sensorial podrá superar retos de mayor envergadura que otro que carezca de ella. Tradicionalmente, ha existido cierta disociación entre los estudios sobre visión artificial y los relacionados con el control de robots. En los primeros, el principal esfuerzo se ha centrado en producir interpretaciones simbólicas y/o geométricas del mundo a partir de la información sensorial obtenida por una o más vistas de una escena. Aunque en este sentido los avances son numerosos, esta concepción representacionalista de la percepción visual no ha producido los resultados deseados en robótica. Tal y como apunta Arkin, seguramente el problema no está en los medios, sino en los objetivos (Arkin, 1998). Desde la propuesta ecológica de la percepción visual de Gibson (Gibson, 1979), esta cuestión se clarifica aún más a través de la paradoja del "pequeño hombre sentado en el cerebro". Según esta teoría, si la imagen percibida se transmite al cerebro como un todo, debe haber un pequeño hombre, un homúnculo, sentado en el cerebro que mire esta imagen con su pequeño ojo, el cual, a su vez, transmitirá la información percibida a su pequeño cerebro. Esto lleva a una situación paradójica desde la que la percepción sólo puede ser explicada mediante la existencia de una serie infinita de homúnculos, percibiendo la imagen del anterior. De igual forma, cabe preguntarse: si la finalidad de la percepción es construir una representación detallada del mundo, ¿quién mira esa representación?

Existen evidencias psicológicas de que la generación de una representación interna del mundo no es más que una mera impresión. Ciertas teorías, como la ceguera al cambio (Simons y Levin, 1998) (O'Regan et al., 1999), demuestran que no vemos todo a nuestro

alrededor. Sólo percibimos lo que procesamos en cada momento. La sensación de una percepción completa se explica por la capacidad para acceder a cualquier parte de una escena visual a través de movimientos oculares, de la cabeza y/o del cuerpo. En este sentido, la percepción está estrechamente relacionada con la acción y, por lo tanto, no debe ser considerada como un proceso aislado.

En robótica, la importancia de la relación entre percepción y acción ha sido destacada por varios autores desde diferentes propuestas. Existen dos líneas fundamentales. Las que plantean las estrategias perceptivas a partir de las necesidades conductuales (percepción orientada a la acción) (Arbib, 1981) y en las que las acciones están dirigidas por los requerimientos perceptivos (percepción activa) (Bajcsy, 1988)(Ballard, 1991). El trabajo desarrollado en esta tesis se sitúa entre ambos enfoques, abordando dos cuestiones clave: ¿cómo incluir la influencia de las acciones en el procesamiento perceptivo?¿cómo regular las acciones en función de la percepción? La hipótesis fundamental de este trabajo es que ambas cuestiones pueden ser resueltas a través de la atención. Más concretamente, la base de nuestra propuesta es que la atención actúa como un medio de conexión entre percepción y acción que permite orientar el proceso perceptivo en función de las acciones y modular las acciones de acuerdo con el resultado perceptivo del control atencional. En este sentido, nuestro planteamiento se encuentra muy cerca de las teorías que conciben la atención como medio de selección para la acción. Desde este enfoque, la atención emerge de los mecanismos selectivos del propio sistema visual para la programación y ejecución de la acción.

## 1.1. Objetivos y aportaciones de la tesis

Como ya se ha esbozado anteriormente, esta tesis aborda el problema de la atención visual y su relación con la acción, dentro del marco de la robótica móvil y el control autónomo. Nuestra principal hipótesis es que la atención actúa como intermediario fundamental entre la percepción visual y el control de la acción. Para explorar nuestra propuesta, se ha establecido como principal objetivo el diseño de un sistema de atención visual orientado a la selección para la acción. El sistema es parte integrante de una arquitectura de control desde la que emergen diferentes comportamientos en base al enlace entre atención y acción.

Los diferentes aspectos que se han tenido en cuenta para obtener el sistema final son

variados. Por un lado, se plantea un método de análisis de imágenes que permita detectar zonas de potencial interés de una escena. Se definen distintas propiedades visuales de los elementos de una escena que determinen su relevancia en el proceso de selección atencional. Se establecen los mecanismos atencionales que dan lugar al comportamiento visio-motor deseado desde la perspectiva de la selección para la acción. Se desarrollan métodos de control de movimientos oculares que acompañen la selección atencional de una forma real y estable de foveatización de un objetivo visual.

Aunque los aspectos anteriores puedan considerarse las principales aportaciones de esta tesis, su desarrollo sólo es posible si existen dos cuestiones resueltas de antemano: un robot real sobre el que funcione el sistema y una arquitectura software flexible que proporcione una ejecución en tiempo real. Por este motivo, dentro del trabajo desarrollado se han incluido como objetivos previos los siguientes:

- Diseño y construcción de un robot móvil dotado de una torreta motorizada de visión estéreo.
- Diseño de una arquitectura software que proporcione una implementación modular del sistema y un funcionamiento en tiempo real.

Por último, para completar nuestro estudio y probar el sistema propuesto, se han realizado una serie de experimentos reales dirigidos a resolver tareas de navegación en el robot.

Aunque esta tesis no se plantea como una solución definitiva a las cuestiones principales de la percepción visual en la robótica móvil, creemos que abre nuevas vías de estudio, proporcionando las bases iniciales para exploraciones más profundas sobre las conexiones entre atención y acción en el control de robots.

## 1.2. Estructura de la tesis

La tesis está estructurada en dos partes. En la primera, se realiza una revisión sobre distintos aspectos de la visión biológica, destacando fundamentalmente los relacionados con la atención visual, y se analiza el estado de la cuestión en cuanto a modelos computacionales de atención visual y arquitecturas de control en robots. Este primer bloque comprende los capítulos 2, 3 y 4. La segunda parte, compuesta por los capítulos del 5 al 8, constituye la propuesta de esta tesis y en ella se exponen las diferentes cuestiones teóricas y prácticas

del sistema desarrollado.

A continuación se muestra un breve resumen de los contenidos de cada capítulo:

- Capítulo 2 - *El sistema visual humano*: en este capítulo se recogen aspectos de la visión biológica que resultan de interés para el desarrollo de nuestra propuesta. Se dedica parte del capítulo a realizar una revisión de los estudios psicológicos y neurológicos sobre atención visual, tratando diferentes cuestiones como tipos de atención, funciones de la atención y modelos teóricos de los mecanismos atencionales.

- Capítulo 3 - *Modelos computacionales de atención visual*: en este capítulo se describe un grupo de propuestas representativas de los modelos computacionales de atención visual surgidos en los últimos años. Este análisis pretende mostrar cuáles han sido los planteamientos principales que han dado lugar a las diferentes propuestas, con el fin de justificar la necesidad de definir un nuevo modelo computacional para cumplir con los objetivos establecidos en esta tesis.

- Capítulo 4 - *Arquitecturas de control en robots*: revisamos dentro de este capítulo los principales enfoques desde los que nacen las propuestas existentes sobre arquitecturas de control en robots. No se trata de un análisis exhaustivo, sino de una descripción general del estado de la situación actual que permitirá enmarcar nuestro trabajo en una aproximación concreta.

- Capítulo 5 - *Arquitectura hardware y software del sistema propuesto*: en este capítulo se describen las diferentes decisiones de diseño hardware y software que han permitido llevar a la práctica las ideas planteadas.

- Capítulo 6 - *Sistema de control basado en la atención*: se presentan en este capítulo los aspectos concretos de nuestra propuesta. Tras una descripción general del sistema, se detallan los diferentes mecanismos que en él intervienen, exponiendo diferentes cuestiones de tipo teórico, metodológico y algorítmico.

- Capítulo 7 - *Experimentos*: para validar nuestra propuesta, en este capítulo se presentan una serie de experimentos en los que se evalúan los diferentes módulos del sistema de manera independiente y también de forma global a través de un conjunto de pruebas de navegación.

- Capítulo 8 - *Conclusiones*: este capítulo resume las principales conclusiones de esta tesis, incluyendo además varias líneas de actuación que permitan mejorar el trabajo desarrollado en el futuro.

# Parte I

# Fundamentos

# Capítulo 2

# El sistema visual humano

## 2.1. Anatomía del sistema visual humano

El sistema visual humano es extraordinario en cuanto a la cantidad y calidad de la información que proporciona sobre el mundo. Un vistazo es suficiente para describir la posición, tamaño, forma y textura de los objetos. Los primeros pasos en este proceso de visión son la transmisión y refracción de luz a través de la óptica de los ojos, la transformación de energía luminosa en señales eléctricas por los fotoreceptores y el refinamiento de estas señales mediante las interacciones sinápticas de los circuitos neuronales de la retina. La información proporcionada por la retina inicia un proceso de interacción entre diversas zonas cerebrales que dan lugar a la percepción de una escena visual, al mismo tiempo que estimulan otros reflejos como ajustar el tamaño de la pupila, dirigir los ojos a zonas de interés y regular comportamientos homeostáticos (Purves et al., 2004).

### 2.1.1. El ojo humano

El ojo humano es un órgano esférico que se encuentra rodeado de 3 capas de tejido fino: una externa, formada por la esclerótica y la córnea; una intermedia, dividida en dos partes, una anterior (iris y cuerpo ciliar) y una posterior (coroides); y una interna o porción sensorial del ojo, la retina.

La esclerótica es la capa más externa del ojo y está formada por un tejido fibroso blanco y resistente, cuya función principal es la de cubrir la mayor parte del globo ocular. En el frontal del ojo, esta capa opaca se transforma en la córnea, una estructura transparente

que permite el paso de la luz hacia el interior protegiendo al iris y al cristalino. Posee propiedades ópticas de refracción, representando cerca de 2/3 de la capacidad de enfoque del ojo. Es uno de los pocos tejidos que no poseen irrigación sanguínea, dado que se nutre de la lágrima y del humor acuoso.

La capa de tejido intermedio está compuesta por tres estructuras que conjuntamente reciben el nombre de tracto uveal. La mayor de ellas es el coroides que se extiende hasta el cuerpo ciliar, situado cerca de la parte frontal del ojo. Esta segunda estructura es un anillo de tejido que rodea al cristalino y que se encarga de ajustar su curvatura y de producir el fluido que cubre la parte frontal del ojo (humor acuoso). El tercer componente del tracto uveal es el iris cuya estructura muscular permite regular la cantidad de luz que pasa a través de la pupila.

En la parte posterior de la superficie interna del ojo se encuentra la retina, zona sensorial del ojo. Contiene receptores sensibles a la luz (fotorreceptores) que convierten la luz en impulsos eléctricos, que se transmiten a las zonas cerebrales encargadas de la visión. Los fotorreceptores se clasifican en conos y bastones. Los bastones son sensibles a la oscuridad y a la luz sin color, mientras que los conos responden a la luz de color. El número de bastones supera con creces el número de conos, por lo que la densidad de bastones es mucho mayor que la de conos sobre la mayor parte de la retina. No obstante, la relación entre ambos tipos de fotorreceptores cambia drásticamente en función de la distancia al centro de la retina, lo que da lugar a una distinción entre dos partes de la superficie retiniana, la fóvea y la periferia. La fóvea es la zona central de la retina (tiene un diámetro de aproximadamente 1.2 milímetros) y presenta una alta densidad de conos, llegando incluso, en la región más interior (foveola), a incluir únicamente este tipo de receptores. Además, la relación entre los receptores y las células ganglionares, responsables de enviar la información al nervio óptico, es de 1 a 1, lo que dota a esta zona de una máxima agudeza visual. La región periférica, sin embargo, apenas contiene conos y está principalmente compuesta por bastones. La información procedente de múltiples receptores en esta zona se envía a una única célula ganglionar, lo que da lugar a una menor resolución que en la fóvea, que disminuye con el aumento de excentricidad.

### 2.1.2. Los caminos visuales

La información procedente de la retina es procesada por múltiples regiones del cerebro gracias a diferentes conexiones, denominadas caminos visuales. Las células ganglionares del tracto óptico alcanzan distintas estructuras del diencéfalo y del cerebro medio. La mayor de ellas es el núcleo genicular lateral del tálamo, que transmite información visual procedente de ambos ojos al córtex visual primario (V1 o córtex estriado). La segunda estructura, situada entre el tálamo y el cerebro medio, es el pretectum, encargado de controlar los ajustes de tamaño de la pupila. Existen, además, conexiones con otras zonas importantes como el colículo superior, situado en el cerebro medio, que permite coordinar los movimientos de la cabeza y de los ojos.

La información visual llega al córtex visual primario (V1) a través de las vías procedentes del núcleo genicular lateral. El área V1 presenta una organización retinotópica, es decir, la estimulación de una determinada zona de la retina excita una zona específica del córtex visual primario. Las células de esta región cortical se agrupan en varios sistemas funcionales que permiten responder a diferentes estímulos de orientación, color o predominio ocular. El área V1 envía información a otras regiones corticales a través de dos vías principales, el flujo dorsal y el flujo ventral. Estas regiones se encuentran situadas en el córtex periestriado y se conocen como V2, V3, V4, V5/MT. Cada una de estas áreas contiene un mapa del espacio visual y su activación depende del área V1. Las propiedades de reacción de las neuronas de algunas de estas regiones sugieren que son zonas especializadas en diferentes aspectos de una escena visual. Por ejemplo, el área MT contiene neuronas que responden selectivamente a la dirección de un borde en movimiento con independencia de su color. Por el contrario, V4 responde selectivamente al color de un estímulo visual sin tener en cuenta su dirección de movimiento. De cualquier forma, todas estás áreas juegan un papel importante en la detección de objetos y en la percepción de las relaciones espaciales de los objetos del campo visual.

## 2.2. Movimientos oculares

Los movimientos oculares son especialmente importantes en la visión humana porque permiten suprimir los efectos de la reducida extensión de la fóvea, zona de máxima agudeza visual. Mediante dichos movimientos es posible dirigir la fóvea a nuevos objetos de interés o compensar los desplazamientos de un objeto que está siendo focalizado.

Los movimientos oculares son posibles gracias a seis músculos, cada uno de los cuales es responsable de un determinado ajuste de la posición del ojo: el recto medio y lateral (movimientos hacia los lados), el recto superior e inferior (movimientos hacia arriba y hacia abajo), y los oblicuos superior e inferior (movimientos de torsión). A través de los grados de libertad proporcionados por estos músculos, se producen diferentes movimientos oculares, cada uno con su propio circuito de control y con una función especifica. Éstos son los sacádicos, las persecuciones lentas, los movimientos de vergencia y los movimientos vestibulares. En los siguientes apartados se describen las funciones principales de estos cuatro movimientos fundamentales.

### 2.2.1. Sacádicos

Son movimientos rápidos de los ojos que modifican bruscamente el punto de fijación. La amplitud de estos movimientos comprende un rango intermedio entre los pequeños movimientos que se realizan durante una lectura y los que se llevan a cabo mientras se mira alrededor de una habitación. Pueden producirse voluntariamente, pero también se realizan de manera refleja aunque el observador fije la mirada intencionadamente sobre una determinada zona de interés. Durante la duración de un movimiento sacádico se produce un fenómeno denominado supresión sacádica, que consiste en la interrupción casi absoluta de recogida de información. La duración de un sacádico depende de la amplitud del movimiento (entre 30 y 120 milisegundos), aunque la velocidad que alcanza aumenta a medida que crece el ángulo del desplazamiento. Se dice que los movimientos sacádicos son movimientos balísticos porque el sistema de generación de sacádicos no puede responder a cambios en la posición del objetivo mientras dura el movimiento ocular. Si el objetivo se desplaza durante este tiempo, el sacádico no permite alcanzar la posición correcta y es necesario realizar un segundo sacádico para corregir el error.

### 2.2.2. Persecuciones lentas

Son movimientos lentos de seguimiento que permiten mantener en la fóvea un estímulo en desplazamiento. Son voluntarios en el sentido de que el observador decide si mirar o no el objetivo en movimiento, pero, en ausencia de movimiento externo, es muy difícil llevar a cabo un movimiento de este tipo de manera intencionada.

### 2.2.3. Vergencias

Los movimientos de vergencia permiten situar la fóvea de los dos ojos sobre el mismo objetivo visual. Mediante estos movimientos es posible mantener una fijación binocular sobre un estímulo que varía en distancia con respecto al observador. Existen dos tipos de vergencias: movimientos convergentes y divergentes. Los movimientos convergentes se producen cuando el objetivo visual se acerca. La divergencia es el movimiento contrario, provocado cuando el objetivo se aleja del observador.

Los estudios sobre las relaciones entre movimientos sacádicos y movimientos de vergencia en la fijación binocular han conducido a dos posturas alternativas. Por un lado, algunos investigadores defienden la existencia de un movimiento simétrico post-sacádico de vergencia que se ajusta a la distancia del punto de fijación (Yarbus, 1967). Desde esta perspectiva, el cambio de dirección producido por un movimiento sacádico se programaría conjuntamente en ambos ojos. El segundo enfoque, sin embargo, apunta hacia una programación monocular de los movimientos sacádicos (Enright, 1998). En este sentido, proponen que la posición de un objetivo visual percibida desde uno de los dos ojos produce un sacádico en dicho ojo que proporciona un alineamiento preciso con el objetivo visual, mientras que el otro ojo alcanza la posición correcta posteriormente a través de un movimiento asimétrico de vergencia.

### 2.2.4. Movimientos vestibulares

Son movimientos involuntarios que permiten compensar los movimientos del tronco o la cabeza. A través de ellos la imagen en la retina se mantiene estable aunque varíe la posición de la cabeza. Así, cuando la cabeza se mueve hacia un lado, los ojos se mueven una distancia similar en la dirección opuesta. De esta forma, la imagen de un objeto en la retina permanece en una posición similar a la que se encontraba antes del movimiento.

## 2.3. Atención visual

A lo largo de los años, la atención se ha asociado con diferentes aspectos tales como consciencia, selección o control. Son numerosos los estudios realizados sobre la atención desde diferentes niveles (fisiológico, neurológico, psicológico, etc.). No obstante, no parece haber una definición claramente explicativa de la atención, ni tan siquiera un consenso

sobre los diferentes aspectos de la función atencional.

Con el fin de dar una visión global de las distintas cuestiones sobre la atención, en los siguientes apartados se describen varios estudios desde diferentes perspectivas: tipos de atención, funciones de la atención y modelos teóricos de los mecanismos atencionales. Por ser el área de interés de está tesis, nos centraremos fundamentalmente en la atención visual.

### 2.3.1. Tipos de atención

Hay muchas maneras de clasificar el sistema de atención visual de acuerdo con varios aspectos. Desde el punto de vista del estímulo, dicho estímulo puede atraer la atención mediante mecanismos endógenos o exógenos (Posner et al., 1980). En la atención endógena, el control depende principalmente de las intenciones y acciones del sujeto. Este tipo de atención es conocido también como atención *top-down* o guiada por objetivo (Yantis, 1998). Por otro lado, la atención puede ser dirigida por componentes exógenos, esto es, por estímulos externos que atraen la atención hacia una posición determinada. A este segundo tipo de atención se le denomina atención *bottom-up* o dirigida por estímulo.

Desde el punto de vista del sujeto, el punto de fijación de la mirada puede variar a un nuevo punto de atención (atención abierta). Por otro lado, también puede variar el procesamiento de atención a una nueva posición dentro del campo visual mediante una focalización mental, sin que se produzca ningún movimiento (atención encubierta). Una forma de atención encubierta es la localización previa de la posición a la que se mueve el ojo en la atención abierta.

Combinando algunos modelos neurológicos de atención, Perry y Hodges (Perry y Hodges, 1999) han dividido la atención en 3 categorías:

1. Atención selectiva y navegante: se caracteriza por el enfoque sobre un único estímulo relevante o el procesamiento en el momento en que se ignoran estímulos irrelevantes.

2. Atención sostenida: mantiene un foco de atención en periodos largos de tiempo.

3. Atención dividida: comparte la atención entre más de un estímulo relevante.

Por otra parte, James (James, 1890) propone varios criterios de clasificación de la atención tras distinguir diversos aspectos funcionales de la misma:

1. En función del objeto al que se dirige: atención sensorial (cuando la atención se dirige a la entrada perceptiva); atención intelectual (focalización mental sin influencia de la entrada sensorial).

2. En función del interés que la causa: atención inmediata (si el objeto atendido es interesante por sus propias características); atención derivada (cuando el interés del objeto es consecuencia de su asociación con otros intereses).

3. En función del modo en que es controlada: atención pasiva (involuntaria); atención activa (voluntaria).

### 2.3.2. Funciones de la atención

Los estudios iniciales sobre la atención enfatizan su función selectiva. James apunta que la atención permite controlar el acceso a la consciencia de únicamente el estímulo atendido: "*Todo el mundo sabe lo que es la atención. Es la toma de posesión por la mente, de un modo claro y vívido, de uno de entre varios objetos o cadenas de pensamiento simultáneamente posibles*"(James, 1890).

Esta misma idea se mantiene por varios autores que destacan la función selectiva de la atención bajo el supuesto de *capacidad limitada de la mente* (Broadbent, 1958)(Eriksen, 1990)(Deutsch y Deutsch, 1963)(LaBerge, 1995). De acuerdo con esta idea, la mente sería un sistema de capacidad limitada y la atención actuaría como un filtro que permitiría el procesamiento de un único elemento, evitando la "sobrecarga" del sistema (Broadbent, 1958)(Deutsch y Deutsch, 1963). Este tipo de teorías dio lugar a que los estudios se centraran en cuestiones como la ubicación de este filtro, surgiendo dos modelos: modelo de filtro pre-categorial y modelo post-categorial. En el modelo pre-categorial, el filtro se sitúa en las etapas iniciales del procesamiento (Broadbent, 1958). Todos los estímulos son analizados según ciertas características, pero sólo uno es seleccionado para pasar a las siguientes etapas de procesamiento. Sin embargo, otros autores consideran que el filtro se ubica en etapas tardías (post-categorial), esto es, proponen que el filtro atencional no actúa hasta que todos los estímulos son procesados y categorizados (Deutsch y Deutsch, 1963).

La idea de la capacidad limitada ha sido criticada por muchos autores que sugieren que la capacidad de los sistemas perceptivos es enorme y cuestionan la necesidad de los mecanismos atencionales desde la perspectiva de la selección para el procesamiento (Neumann et al., 1986) (Allport, 1987). Bajo este enfoque, plantean la dificultad, no para procesar gran cantidad de información al mismo tiempo, sino para determinar qué información es necesaria a la hora de actuar. Según esta concepción, la necesidad de la atención reside en la selección para la acción (Allport, 1987), lo que permite evitar el desorden conductual mediante la selección de información adecuada para dirigir la ejecución de una tarea. Dentro de este mismo enfoque, algunos autores (Riddoch et al., 2000b)(Riddoch et al., 2000a) sugieren, a partir de estudios experimentales, que los estímulos evocan acciones a partir de los *affordances*[1] derivados de sus propiedades visuales. Proponen la existencia de rutas cerebrales hacia la acción, controladas por los atributos de los objetos, que dan lugar a la secuencia de acciones apropiadas. Esta hipótesis surge de diversos estudios con pacientes que presentan ciertas disfunciones a la hora de actuar como, por ejemplo, los estudios sobre el "síndrome de desorganización de la acción", en la que los sujetos no pueden realizar tareas cotidianas de manera organizada. Los autores afirman que esta disfunción parece estar causada por una alta influencia del entorno para actuar sobre los objetos, que no permite centrar la acción en las restricciones de una tarea (Humphreys y Forde, 1998). Así, se plantea la necesidad de un proceso de control que permita modular los efectos de los *affordances* en la selección de la acción de manera que la secuencia de pasos que dan lugar a la consecución de una tarea se lleven a cabo en el orden adecuado. En este sentido, proponen que la selección de la acción viene precedida de la selección del objeto sobre el que se va a actuar entre los múltiples que evocan una respuesta potencial.

### 2.3.3. Modelos psicológicos y neurológicos de la atención

El interés creciente sobre la estrategia atencional de los sistemas biológicos ha llevado a muchos investigadores a proponer modelos de funcionamiento de los mecanismos atencionales desde diferentes áreas de conocimiento. En este apartado se muestran algunos de estos modelos desde los enfoques psicológico y neurológico. En el siguiente capítulo, se describen con más detalle los modelos computacionales surgidos en los últimos años.

[1] El termino *affordance* fue introducido por Gibson (Gibson, 1979) para describir las posibilidades para la acción ofrecidas por los objetos

Desde la perspectiva psicológica se han propuesto varios modelos, la mayoría de los cuales se encuentran apoyados en la hipótesis de capacidad limitada de la mente. Este es el caso de la "teoría de integración de características", la "búsqueda guiada" y el modelo "FeatureGate".

La teoría de integración de características, propuesta por Treisman (Treisman y Gelade, 1980), surge para explicar los diferentes tiempos de respuesta de un sujeto en función del número de características que definen un objetivo en una tarea de localización. Treisman observó que el tiempo de respuesta se incrementaba linealmente con el número de estímulos cuando la descripción del objetivo se realizaba a partir de varias características, mientras que dicho tiempo no se veía afectado por la complejidad de la escena cuando la búsqueda se realizaba a través de una única característica. Para explicar esta circunstancia, en la teoría de integración de características se propone la existencia de varios mapas separados de características de la escena visual tales como color, orientación, brillo y dirección de movimiento, que se obtienen en paralelo y de manera preatentiva. Las posiciones de las características se extraen por separado y se almacenan en un mapa maestro de posiciones (Treisman, 1988). La función de la atención en esta teoría es la de combinar las distintas características del estímulo seleccionado para formar un único objeto. La selección de cada estímulo es el resultado de la activación de una posición del mapa maestro, que permite la búsqueda serializada de un objetivo definido a través de varias características.

En la teoría de la búsqueda guiada (Wolfe, 1994), los estímulos son procesados a través de varios canales categorizadores que permiten obtener mapas de activación de las zonas de una escena visual. Algunos de estos canales se encargan de calcular niveles de activación de los estímulos mediante diferencias locales de características. Otros, sin embargo, reflejan la activación de cada estímulo en función de los objetivos del sujeto. A través de los mapas obtenidos por los diferentes canales, se obtiene un mapa de activación global resultante de la suma ponderada de los mapas individuales. La distribución atencional se lleva a cabo en orden decreciente de activación.

El modelo "FeatureGate" (Cave, 1999) utiliza una estructura jerárquica de mapas espaciales de características. En cada nivel existen varios elementos que responden a diferentes características, tales como color u orientación. La información pasa de un nivel al siguiente controlada por bloqueos atencionales, mediante mecanismos *bottom-up* y *top-down*. A

medida que la información sube de nivel, el número de elementos del nivel disminuye, aumentando el campo receptivo de cada elemento. El bloqueo de cada posición se lleva a cabo mediante un mecanismo que regula la competición entre diferentes posiciones para encontrar la que contiene la información más relevante. El mecanismo de selección *bottom-up* actúa activando aquellas posiciones que difieren en una característica dentro de su entorno local. La selección *top-down* se lleva a cabo inhibiendo aquellas posiciones cuyas características no mantienen relación con las del objetivo visual. Cuantas menos correspondencias existan entre las características del objetivo y las de una determinada posición, mayor será el nivel de inhibición asociado a dicha posición. A medida que la información fluye ascendentemente por la jerarquía, el número de objetivos visuales se reduce, obteniendo un candidato global que es seleccionado en el último nivel.

Como alternativa a los modelos que consideran la atención como un sistema unitario de selección de información para su posterior procesamiento, Posner propone un modelo de varias redes atencionales (Posner y Dehaene, 1994) que realizan funciones asociadas con diferentes aspectos tales como la detección, la alerta, la orientación, el control y la consciencia. De acuerdo con este modelo, la atención se materializa en tres redes neuronales denominadas red anterior, posterior y de vigilancia. La red anterior está relacionada fundamentalmente con la detección y selección de objetivos. Los autores afirman que la cuestión de la capacidad limitada de procesamiento está asociada con la capacidad de esta red para mantener varios objetivos. La red atencional de vigilancia es responsable del mantenimiento de un estado de alerta que permita al sujeto reaccionar adecuadamente en una situación determinada. Por último, la red posterior está asociada con la orientación visuo-espacial de la atención. El término orientación se refiere en este caso tanto al desplazamiento motor como a la atención encubierta y se considera diferente de la detección.

Existe otro grupo de modelos que centran la función de la atención en la representación y el reconocimiento de objetos (Olshausen et al., 1993)(Heinke y Humphreys, 1997)(Postma et al., 1997). La mayoría de ellos se basan en la idea de los circuitos de encaminamiento dinámico (Anderson y Van Essen, 1987)(Olshausen et al., 1993), en la que se propone que el foco de atención es el resultado de encaminar selectivamente la información desde una región del córtex visual primario hacia áreas corticales de mayor nivel. Este encaminamiento dinámico se produce por la reconfiguración dinámica de los recursos neuronales mediante la modulación de sus conexiones. Las ideas centrales de esta propuesta son: (1) el flujo

de información es controlado dinámicamente a través de una jerarquía piramidal de varios niveles de procesamiento neuronal; (2) las relaciones espaciales en el interior de la región visual atendida se mantienen explícitamente; y (3) el control de atención se lleva a cabo mediante la modulación dinámica de los pesos de las conexiones neuronales. El principal objetivo del modelo es obtener una representación de la región atendida que sea invariante a la escala y a la posición para su posterior reconocimiento. Esto se lleva a cabo mediante un conjunto de neuronas de control que se encargan de trasladar la representación topográfica de la región seleccionada de un nivel al siguiente, de manera que, en el nivel más alto, se obtenga una representación invariante de la región atendida.

Los modelos anteriores son, en su mayoría, *monárquicos* u *oligárquicos* (Roselló et al., 2001), es decir, se basan en la existencia de un mecanismo de control supramodal y unitario que ejerce su función de manera independiente a los sistemas perceptivos y motores (*modelos monárquicos*), o bien defienden que el control atencional es el resultado de la acción coordinada de varios mecanismos de este tipo (*modelos oligárquicos*). Existen otro tipo de modelos que plantean la naturaleza *anárquica* de la atención, esto es, niegan la existencia de un mecanismo de control atencional superior y sugieren que la atención emerge de los propios circuitos sensomotores. Este último grupo de modelos se sitúa muy cerca de las teorías que defienden la función atencional desde la perspectiva de la selección para la acción. Dentro de ellos, destaca la teoría premotora de la atención (Rizzolatti et al., 1987) en la que se plantea un modelo de atención integrado en los circuitos senso-motores que no incluye ningún centro de control supramodal. Se basa en la existencia de una serie de circuitos frontoparietales, denominados *mapas pragmáticos*, que codifican el espacio de diferentes formas según las acciones en las que dicha información espacial pueda ser utilizada. Ninguno de dichos circuitos posee un mapa espacial que pueda servir para todos los propósitos motores, por lo que los autores defienden la idea de que cada uno posee su propia forma de atención. En este sentido, afirman que la diversidad en las codificaciones espaciales hace difícil concebir un sistema unitario de control de atención debido a la multiplicidad de codificaciones con las que dicho sistema debería ser capaz de actuar. A partir de esta idea, proponen una atención modular senso-motora, estableciendo un estrecho vínculo entre la atención espacial y la selección para la acción. Según esta hipótesis, la activación de las neuronas de los mapas pragmáticos que codifican una cierta ubicación espacial se enlaza directamente con la activación de las neuronas que programan una acción determinada. Sostienen además que la atención encubierta es el resultado de la activación de un

programa motor que no se llega a ejecutar. Esta afirmación es congruente con resultados experimentales que demuestran que los mapas pragmáticos se activan aunque la acción final no se lleve a cabo.

Otra de las teorías de la atención incluida dentro del grupo de modelos anárquicos es la hipótesis de la competencia integrada de Duncan (Duncan, 1996). Según dicha teoría, la atención es el efecto emergente de la competencia entre las representaciones neuronales de los estímulos en múltiples sistemas. Los autores proponen que las neuronas del córtex visual primario, que responden preferentemente a diferentes estímulos, son mutuamente inhibitorias. Esto implica que los diferentes estímulos de una escena compiten por tomar el control neuronal, de forma que los distintos sistemas tiendan a converger en el procesamiento del mismo estímulo para el control de la acción. Se considera además que los mecanismos competitivos a favor de un estímulo están controlados por la tarea actual y se originan en las áreas cerebrales donde se calculan los atributos relevantes para la tarea.

# Capítulo 3

# Modelos computacionales de atención visual

Los estudios sobre los mecanismos atencionales en los sistemas biológicos han llevado a muchos investigadores a centrar un especial interés en el desarrollo de modelos de atención artificial que imiten al modelo biológico. La mayoría de estos modelos están motivados por la limitación de recursos de procesamiento de los sistemas computacionales e inspirados en las teorías de la atención basadas en la hipótesis de capacidad limitada de la mente. En este capítulo se revisan los modelos computacionales más representativos, comenzando por los que emplean una estrategia de control *bottom-up* (guiada por estímulo) pura y, continuando, por los que incluyen información *top-down* (objetivos e intenciones del sujeto) para llevar a cabo la selección atencional.

## 3.1. Modelo de Koch y Ullman

Se trata de un modelo de atención *bottom-up* (Koch y Ullman, 1985) que, aunque no fue implementado, tuvo gran influencia en otros modelos computacionales. Los autores introducen el concepto de *mapa de saliencia*, como soporte preatentivo de selección, en términos muy similares a los asociados con el *mapa maestro* propuesto por Treisman (Treisman, 1988). El mapa de saliencia es un mapa topográfico que cuantifica el grado en el que cada posición de la imagen llama la atención (nivel de saliencia). Este mapa se construye combinando la información procedente de varios mapas de características que codifican diferentes propiedades del espacio visual (color, orientación de bordes, disparidad...). Cada característica está representada a través de varios mapas que permiten codificar diferentes

dimensiones de dicha característica (por ejemplo, diferentes colores u orientaciones). Las relaciones de vecindad de la escena visual son preservadas en estos mapas, de manera que las posiciones próximas en la imagen original se proyectan en posiciones próximas en cada mapa. Existen, además, conexiones locales inhibitorias que permiten identificar las posiciones que difieren significativamente de su entorno otorgándoles una mayor activación. Cada mapa, por lo tanto, permite seleccionar individualmente las posiciones llamativas con respecto a una determinada dimensión de una característica concreta. Para obtener un valor de saliencia global de cada posición, las medidas individuales obtenidas de cada mapa de características se combinan en el mapa de saliencia. Los autores señalan que la forma en que esta combinación se lleve a cabo es irrelevante siempre y cuando el incremento de activación de una posición en un mapa de características implique un incremento de saliencia en el mapa global.

La selección de una posición a la que dirigir la atención se lleva a cabo a partir de la activación de un elemento del mapa de saliencia. Esta selección implica el almacenamiento de las propiedades de la posición atendida en una representación central utilizada en niveles más altos de procesamiento. Estas operaciones son realizadas por dos redes complementarias: la red WTA (*winner-take-all*), encargada de localizar la posición de mayor activación del mapa de saliencia, y una segunda red que transmite las propiedades de la región seleccionada a la representación central. La red WTA actúa activando la posición de máxima saliencia e inhibiendo el resto de manera que sólo dicha posición alcanza la representación central.

El modelo incluye además mecanismos de desplazamiento atencional que permiten distribuir la atención entre las distintas posiciones salientes del espacio visual. Para introducir esta dinámica proponen dos posibles implementaciones, una local y otra central. En la implementación local, la posición activa del mapa de saliencia adapta su nivel de saliencia de manera que éste decae transcurrido un cierto tiempo. En el esquema central, una vez que la información es transmitida a la representación central, se envía una señal inhibitoria a la posición seleccionada. En ambos esquemas, la red WTA responde a la nueva configuración desviando la atención de la posición seleccionada a la siguiente posición más llamativa.

## 3.2. Modelo basado en saliencia de Itti y Koch

El modelo propuesto por Itti y Koch (Itti y Koch, 2000) está estrechamente relacionado con el modelo de Koch y Ullman descrito en el apartado anterior. A nivel general, el esquema de control atencional presenta 4 puntos fundamentales (figura 3.1): (1) la entrada visual se representa en varios mapas topográficos de características; (2) la información de estos mapas se combina en un único mapa que representa la saliencia de cada posición con respecto a sus vecinos; (3) el máximo de este mapa de saliencia determina la posición a la que dirigir la atención; (4) el mapa de saliencia está dotado de una dinámica interna que permite modificar el foco de atención visitando diferentes zonas en orden decreciente de saliencia.

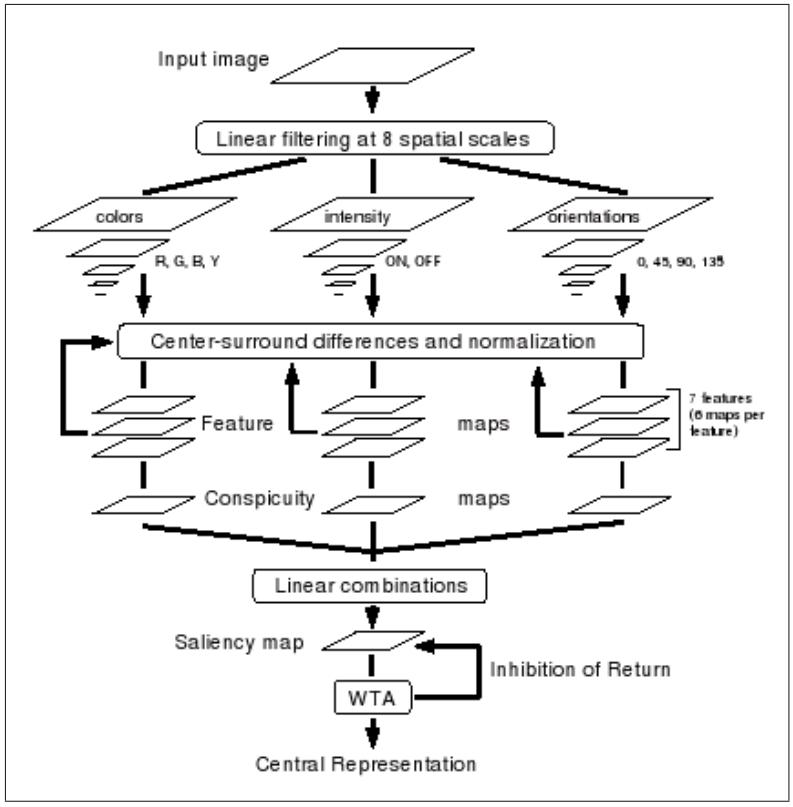



Figura 3.1: Esquema del modelo basado en saliencia de Itti y Koch (Itti y Koch, 2000)

El modelo utiliza 7 tipos de características para representar diferentes propiedades de la escena visual. La primera codifica el contraste en intensidad (ON/OFF) de cada posición (centro oscuro en entorno claro o centro claro sobre fondo oscuro). Otras dos características permiten representar los canales de doble oposición de color: RG (rojo/verde) y

BY(azul/amarillo). Las restantes 4 se encargan de codificar diferentes rangos de orientación local. Cada característica es extraída a partir de la imagen original en varias escalas. Dichas escalas se obtienen mediante filtros paso-bajo y submuestreos sucesivos de la imagen de entrada. La activación de las posiciones en función de cada característica, resultante de la comparación entre centro y entorno local, se calculan mediante diferencias de píxel entre varios niveles del espacio de escala. En concreto, la operación se realiza entre 6 pares de niveles, lo que da lugar a 6 mapas de características por cada tipo de característica y, por lo tanto, a un total de 42 mapas.

Tras la obtención de los diferentes mapas de características, la información almacenada en ellos es combinada en el mapa global de saliencia. Previamente, cada mapa es normalizado para que los valores de activación sean independientes de los mecanismos de extracción de cada característica. Además, cada mapa es convolucionado iterativamente mediante un filtro bidimensional de tipo DoG (diferencia de gaussianas) que permite resaltar las posiciones de mayor activación e inhibir las restantes. Tras este proceso, los mapas relacionados con intensidad, color y orientación son sumados por separado dando lugar a tres mapas de activación. Cada nuevo mapa es sometido al proceso iterativo de filtrado DoG y, finalmente, los 3 mapas son sumados linealmente para obtener un único mapa de saliencia.

La posición de mayor activación del mapa de saliencia constituye el estímulo de mayor saliencia al que dirigir la atención. La búsqueda de dicha posición se lleva a cabo mediante un proceso WTA (*winner-take-all*) implementado a través de una capa bidimensional de neuronas de "integración y disparo" con alta inhibición global. Cuando una neurona de esta capa se activa, provoca el desplazamiento del foco de atención a esa posición. Además, esta activación produce la inhibición de todas las celdas de la capa, dando lugar a una inicialización de la red de neuronas que vuelve a su estado inicial.

Para distribuir la atención entre los diferentes estímulos de alta saliencia del espacio visual, el modelo incluye un mecanismo de inhibición de retorno que actúa entre la capa WTA y el mapa de saliencia. Esta dinámica consiste en el envío, por parte de la red WTA hacia el mapa de saliencia, de información adicional sobre la posición seleccionada que permita su inhibición. Dicha información consiste en una superficie con la estructura espacial de una diferencia de gaussianas centrada en el foco de atención y con desviación

típica igual a la mitad del radio del foco. Esta información permite inhibir, en el mapa de saliencia, el centro y el entorno de la posición previamente seleccionada, provocando el cambio de atención a la siguiente posición de mayor saliencia. La inhibición sobre cada posición se mantiene durante un tiempo determinado, tras el cual decae completamente permitiendo de nuevo la selección de posiciones atendidas con anterioridad.

## 3.3. Modelo STM de Tsotsos

El modelo STM (*Selective Tuning Model*) de Tsotsos (Tsotsos et al., 1995) se basa en la aplicación de un proceso WTA jerárquico que proporciona la selección atencional. La estructura STM consiste en una jerarquía de procesamiento en la que cada nivel representa un cierto grado de abstracción de la entrada (por ejemplo, cada capa está asociada con partes cada vez mayores de un objeto). Cada unidad de esta red piramidal se encuentra conectada a un número determinado de unidades de sus niveles anterior y posterior. Cuando se aplica un estímulo a la capa de entrada de la jerarquía, éste activa de manera ascendente todas las unidades de la pirámide a las que se encuentra conectado (figura 3.2). El modelo supone que la intensidad en las respuestas de las unidades es una medida de la correspondencia entre un estímulo y un modelo, y refleja la importancia de dicho estímulo dentro de los contenidos de una escena.

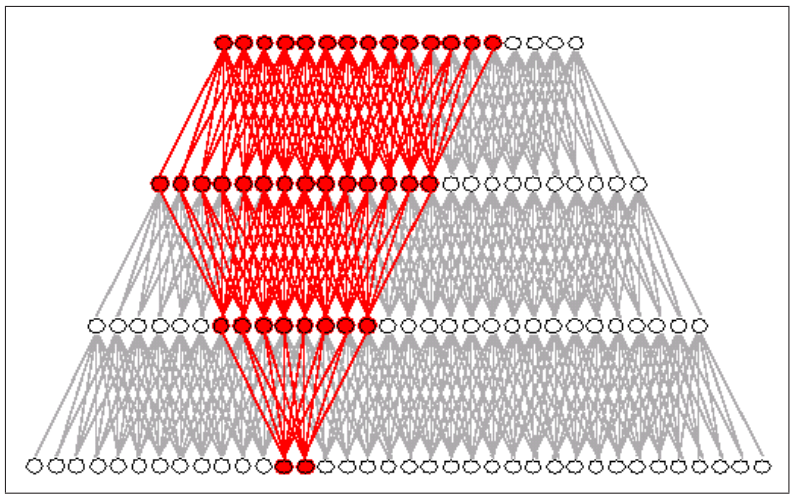


Figura 3.2: Activación ascendente de la estructura en la recepción de un estímulo (Tsotsos, n.d.)

La selección de la posición a la que dirigir la atención se lleva a cabo a partir de una jerarquía de procesos WTA (figura 3.3). El primer proceso WTA opera sobre toda la escena visual en el mayor nivel de la estructura. Este primer paso permite determinar cuáles son las unidades de mayor respuesta y comenzar un proceso de búsqueda hacia las unidades del nivel inferior mediante la activación de una jerarquía de procesos WTA. El ganador global del nivel más alto arranca un proceso WTA que opera únicamente sobre sus entradas directas. Esto permite localizar las unidades de mayor respuesta de entre las que constituyen su campo receptivo. A continuación, todas las unidades que no contribuyen al ganador son inhibidas. Con esta inhibición, las entradas de las unidades del nivel superior varían, modificando así sus salidas. Este paso constituye un refinamiento en las respuestas de las unidades de la estructura que actúa reduciendo o eliminando posibles interferencias entre señales. La estrategia empleada para el primer nivel se repite recursivamente a lo largo de la estructura. Como resultado, se obtiene la localización de las unidades que proporcionan mayor respuesta en el nivel conectado al campo sensorial, completando así la selección atencional. Las conexiones no inhibidas a lo largo de todos los niveles constituyen la zona de paso del estímulo atendido. La inhibición de las conexiones que forman esta zona de paso permite desplazar la atención hacia la siguiente zona de mayor saliencia del espacio visual.

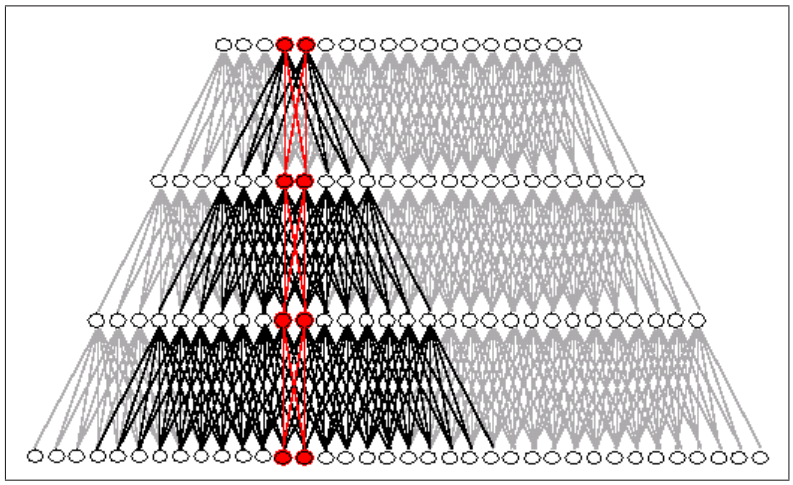


Figura 3.3: Configuración de la estructura tras la aplicación del proceso WTA jerárquico (Tsotsos, n.d.)

Todo el proceso descrito se lleva a cabo a través de diferentes tipos de unidades presentes

en la estructura:

- *Unidades interpretativas*: encargadas de calcular las características visuales.
- *Unidades de paso*: calculan el resultado de aplicar el proceso WTA a las entradas de una unidad interpretativa determinada y permiten el paso de la entrada ganadora hacia las siguientes unidades interpretativas.
- *Unidades de control de paso*: controlan el flujo de selección descendente a través de la pirámide y son responsables de activar o desactivar los procesos WTA.
- *Unidades de sesgo*: proporcionan un medio de selección *top-down*, guiada por tarea, a través de inhibición multiplicativa de las unidades interpretativas.

El modelo se basa en una serie de predicciones sobre la atención biológica que, en los últimos años, han encontrado apoyo a partir de estudios psicológicos y neurológicos (Tsotsos et al., 2001). Las más relevantes son las siguientes:

- La atención proporciona una supresión espacial alrededor del estímulo atendido, así como de características irrelevantes.
- El control atencional está integrado en la jerarquía de procesamiento visual.
- El procesamiento visual preatentivo y atentivo tiene lugar en la misma capa neuronal.
- La modulación atencional es necesaria en cualquier nivel de procesamiento en el que se deba realizar una correspondencia de muchos a uno entre procesos neuronales.
- La distancia topográfica entre los elementos atendidos y los elementos de distracción afecta a la modulación atencional.

## 3.4. Modelo de Sun y Fisher

Sun y Fisher (Sun y Fisher, 2003) proponen un modelo de atención apoyado en la *hipótesis de la competencia integrada* de Duncan (Duncan, 1996). Plantean un esquema atencional que incluye mecanismos de control basados en objetos en los que la atención es dirigida no a una región del espacio (atención espacial), sino a un objeto o grupo de objetos. La fusión entre atención espacial y atención basada en objetos se lleva a cabo a partir del concepto de agrupación, que envuelve objetos, características relacionadas y

posiciones. El término agrupación dentro de este modelo hace referencia a un punto, a un objeto o una característica, a un grupo de objetos o de características, o a una región. Los desplazamientos atencionales están guiados, en este esquema, por un mecanismo de selección jerárquica que permite dirigir la atención hacia un área del espacio, un objeto, una característica o una agrupación de varios elementos.

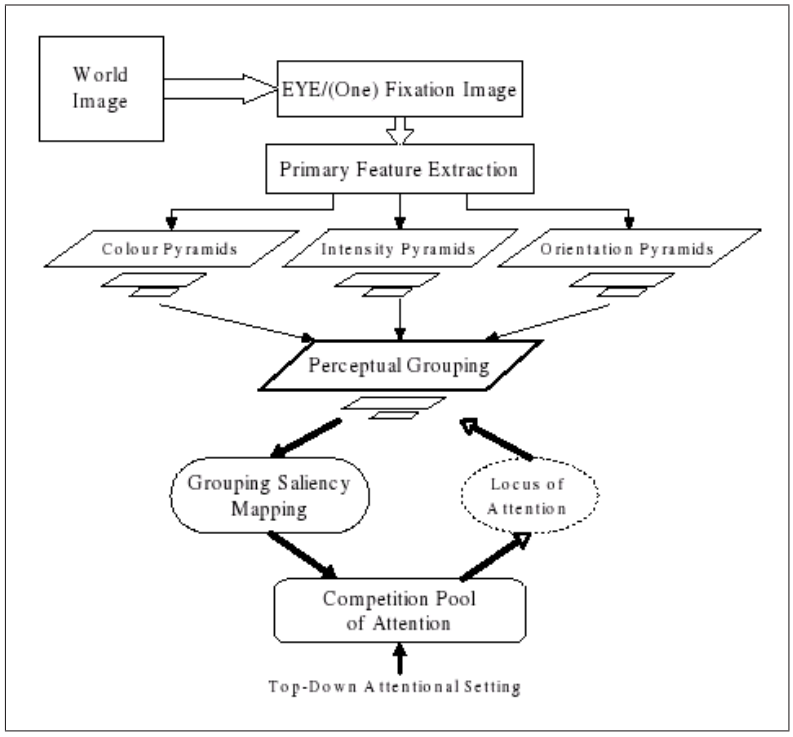



Figura 3.4: Arquitectura del modelo de Sun y Fisher (Sun y Fisher, 2003)

La figura 3.4 muestra la arquitectura del modelo propuesto por Sun y Fisher. El sistema primeramente extrae características básicas (colores, intensidades y orientaciones) de la imagen a varias escalas, formando así una representación piramidal de mapas de características. Tras esta primera fase, se aplica un proceso de agrupación consistente en la segmentación de la imagen en sus diferentes escalas. A continuación, se generan mapas de saliencia *bottom-up* de las agrupaciones extraídas a partir de las características previamente calculadas. La saliencia de una agrupación se obtiene a partir de los valores de saliencia de los componentes incluidos en esa agrupación, como resultado de aplicar un proceso en el que dichos componentes cooperan para competir contra elementos ajenos a la agrupación, a la vez que compiten entre ellos. El efecto de la competición entre dos elementos es el

aumento o reducción de sus valores de saliencia de acuerdo con el resultado de la comparación de sus propiedades. Tras este proceso competitivo, se calcula la saliencia global de cada agrupación mediante una combinación lineal de los valores de saliencia individuales de cada componente en referencia a color, intensidad y orientación.

Una vez determinada la saliencia de las distintas agrupaciones, éstas compiten por la selección atencional mediante la interacción entre saliencia visual e información *top-down* sobre objetivos, desde los niveles de representación más gruesos a los más finos. Los autores afirman que esta idea es coherente con la hipótesis de la competencia integrada en la que se sugiere que la competencia por la selección atencional sucede en múltiples niveles de procesamiento.

Dentro del proceso de selección jerárquica propuesto, la información *top-down* actúa a través de una serie de indicadores que fuerzan a que el mecanismo de competencia dé preferencia o excluya agrupaciones compatibles con determinadas características. Existe, además, un indicador de "*vista de detalles*" que marca si la selección atencional debe o no llevarse a cabo al nivel de las subagrupaciones de una agrupación. A partir de esta información, la competencia por la selección atencional comienza en las agrupaciones del nivel de representación más grueso y se resuelve aplicando un mecanismo WTA. Tras esta selección, en función del indicador de "*vista de detalles*", se decide si el control debe actuar sobre las subagrupaciones del siguiente nivel de representación, a través de un nuevo proceso de competencia, o continuar sobre el nivel actual desplazando la atención hacia la siguiente agrupación de mayor saliencia. Los cambios de atención vienen determinados por un mecanismo de inhibición de retorno que prohíbe la selección sobre una agrupación ya atendida, mientras existan candidatos no seleccionados.

## 3.5. Modelo de orientación contextual de la atención

El modelo de orientación contextual de la atención (Torralba, 2003a)(Torralba, 2003b)(Torralba et al., 2006) es una propuesta de integración de saliencia, información de contexto y mecanismos *top-down* en un control atencional dirigido a la detección y el reconocimiento de objetos. La información de contexto consiste en una descripción global de la escena, obtenida a partir de características de bajo nivel, y se utiliza para predecir la presencia o ausencia del objetivo antes de explorar la imagen y seleccionar las regiones

relevantes.

El análisis de una imagen dentro del modelo se lleva a cabo a través de dos caminos de procesamiento que actúan en paralelo: uno local, encargado de obtener características locales de cada posición, y otro global, mediante el cual se obtiene una representación completa de la imagen a partir de un único conjunto de características que permiten resumir la apariencia de la escena. La representación local es utilizada para obtener un mapa de saliencia *bottom-up*. La representación global mantiene información sobre las regiones de imagen donde se espera encontrar el objetivo.

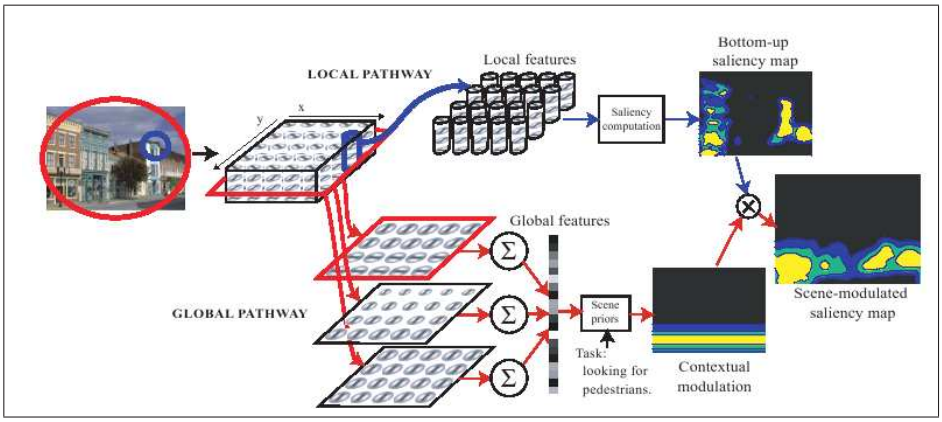



Figura 3.5: Modelo de orientación contextual de la atención (Torralba et al., 2006)

El modelo incluye un marco estadístico de localización que proporciona, para cada posición de la imagen, la probabilidad de presencia del objetivo. Este valor de probabilidad es condicional a las características locales y globales de la imagen y puede descomponerse aplicando la regla de Bayes según la siguiente expresión:

$$p(O|v_l, v_c) = \frac{1}{p(v_l, v_c)} p(v_l|O, v_c)p(O|v_c) \tag{3.1}$$

donde $O$ es un conjunto de variables que describen la apariencia del objetivo, $v_l$ es el conjunto de características locales de cada posición y $v_c$ las características globales que representan la información contextual. La estimación de $p(O|v_l, v_c)$ proporciona la probabilidad de que el objetivo $O$ se encuentre en una posición de la imagen, dado un vector de características de esa posición $(v_l)$ y un contexto descrito por $v_c$. Los términos de la expresión 3.1 pueden interpretarse como diferentes mecanismos que contribuyen al control atencional:

- El término $1/p(v_l, v_c)$ no depende del objetivo y, por lo tanto, constituye un factor puramente *bottom-up*. Proporciona una medida de la improbabilidad de encontrar un conjunto de características $v_l$ en un contexto $v_c$.
- El segundo término, $p(v_l|O, v_c)$, representa un conocimiento *top-down* del aspecto del objetivo. Expresa la probabilidad de que el grupo de características $v_l$ pertenezcan al objetivo $O$ dentro del contexto $v_c$.
- El último término, $p(O|v_c)$, proporciona la probabilidad de presencia del objetivo en la escena.

Los valores de probabilidad obtenidos a partir de la expresión 3.1 para cada posición de la imagen constituyen un mapa de saliencia modulado por el contexto que permite detectar la presencia de un objetivo y localizarlo en la escena.

## 3.6. Integración de atención top-down en el modelo de saliencia de Itti y Koch

Navalpakkam e Itti (Navalpakkam y Itti, 2005)(Navalpakkam y Itti, 2006) proponen una modificación al modelo de saliencia de Itti y Koch que integra información *bottom-up* de la escena y conocimiento *top-down* de los objetivos para llevar a cabo el control atencional (figura 3.6).

Como en el modelo original de Itti y Koch, la escena visual se analiza en primer lugar según diferentes dimensiones de características, obteniendo un mapa de saliencia para cada característica dentro de una dimensión[1] . La influencia *top-down* en el control atencional se representa a través de un conjunto de ganancias que ponderan el valor de saliencia en cada dimensión de características y, también, en cada característica. Las ganancias son determinadas a partir de conocimiento previo sobre el objetivo y los posibles elementos de distracción de la escena. Por cada dimensión $j$, se obtiene un mapa de saliencia $S_j$ (ecuación 3.2) resultante de combinar linealmente los mapas de saliencia de sus distintas características ($s_{ij}$) previamente modulados por los valores de ganancia correspondientes ($g_{ij}$). De manera similar, el mapa global de saliencia se calcula a partir de la influencia *top-down* en cada dimensión ($g_j$) mediante una combinación lineal de los mapas de saliencia

[1]El término dimensión de características es utilizado por los autores para referirse a un grupo de características de un mismo tipo (color, orientación, etc.).

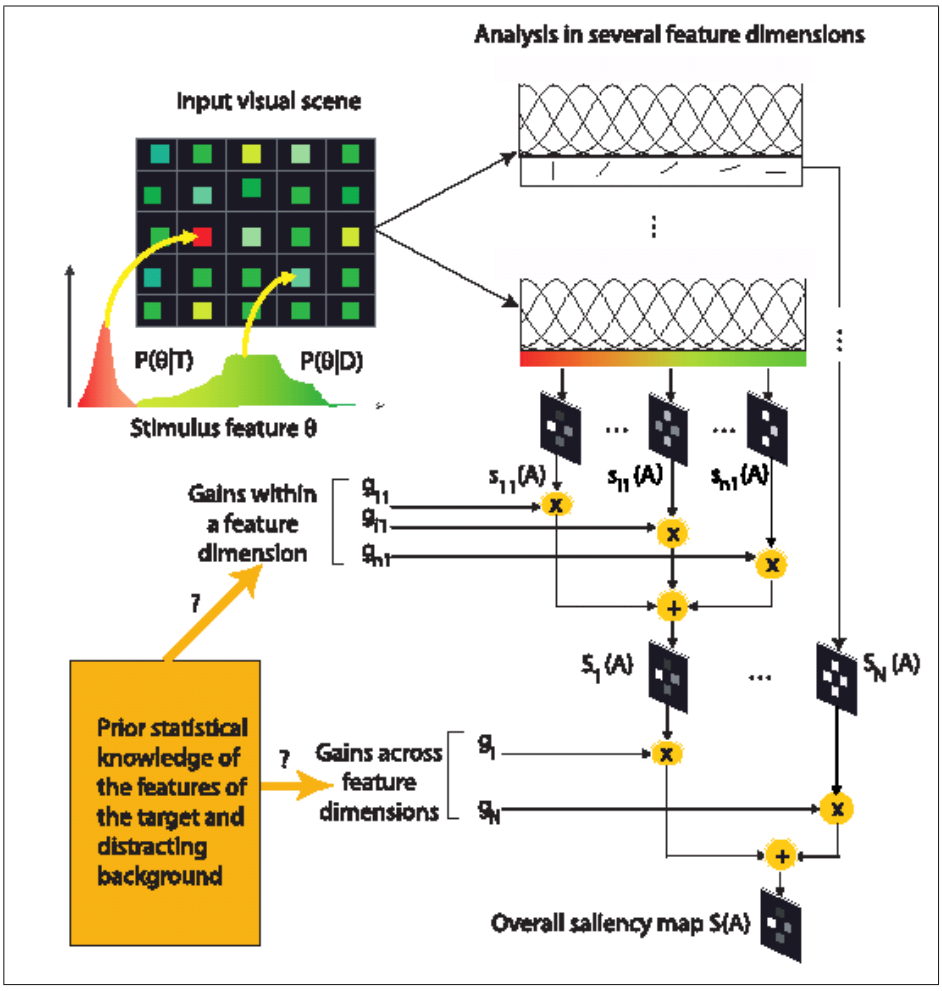



Figura 3.6: Modelo de atención de Navalpakkam e Itti (Navalpakkam y Itti, 2006)

de las diferentes dimensiones (ecuación 3.3).

$$S_j = \sum_{i=1}^{n} g_{ij} s_{ij} \tag{3.2}$$

$$S = \sum_{j=1}^{N} g_j S_j \tag{3.3}$$

Los valores de ganancia son determinados en cada caso de forma que la saliencia global del objetivo se maximice con respecto al resto de la escena. Para ello, el modelo supone conocidas las distribuciones de probabilidad de las características del objetivo ($p(\Theta|T)$)

y de los posibles elementos de distracción ($p(\Theta|D)$). Se considera además que los valores de saliencia, tanto del objetivo como de las zonas de distracción, dependen de otras dos variables aleatorias que son la configuración espacial ($C$) de ambos tipos de elementos en la escena y un posible ruido interno ($\eta$). A partir de todos estos factores, el cálculo de los valores de ganancia se lleva a cabo maximizando la relación entre la saliencia del objetivo ($S_T$) frente a la saliencia de los elementos de distracción ($S_D$). Dicha relación es denominada ratio señal-ruido ($SNR$) y calculada a partir de la esperanza matemática de $S_T$ y $S_D$ en función de las variables aleatorias $\Theta|T$, $\Theta|D$, $C$ y $\eta$:

$$SNR = \frac{E_{\Theta|T,C,\eta}[S_T]}{E_{\Theta|D,C,\eta}[S_D]} \tag{3.4}$$

Expresando la ecuación 3.4 en términos de las saliencias en cada dimensión de características y en cada característica dentro de una dimensión, se obtiene una función de las ganancias $g_j$ y $g_{ij}$. La maximización de dicha expresión en relación a cada término de ganancia proporciona los distintos valores de $g_j$ y $g_{ij}$ que permiten una localización directa del objetivo en la escena de acuerdo con el conocimiento previo sobre dicho objetivo.

## 3.7. Sistema VOCUS de Frintrop

VOCUS (*Visual Object detection with a CompUtational attention System*), propuesto por Frintrop (Frintrop et al., 2005)(Frintrop, 2006), es un sistema de atención dirigido a la detección de objetos que combina componentes de control *bottom-up* y *top-down* en el proceso de selección atencional.

La figura 3.7 muestra un esquema del sistema. La parte *bottom-up* del modelo se encarga de detectar las regiones salientes de la imagen siguiendo un proceso similar al propuesto por Itti y Koch. Para ello utilizan 10 mapas de características, 2 de intensidad, 4 de orientación y 4 de color. El modelo enfatiza las regiones que destacan sobre el resto en alguna característica, para lo cual, cada mapa de características $X$ es transformado en un mapa $W(X)$ resultante de dividir $X$ por la raíz cuadrada del número de máximos locales ($m$) mayor que un cierto umbral ($W(X) = X/\sqrt{m}$). Tras esta transformación, los mapas individuales son sumados para formar el mapa de saliencia *bottom-up*.

El subsistema *top-down* incluye dos modos de funcionamiento: un modo de aprendizaje, que calcula un conjunto de pesos en función de un objetivo específico, y un modo de bús-

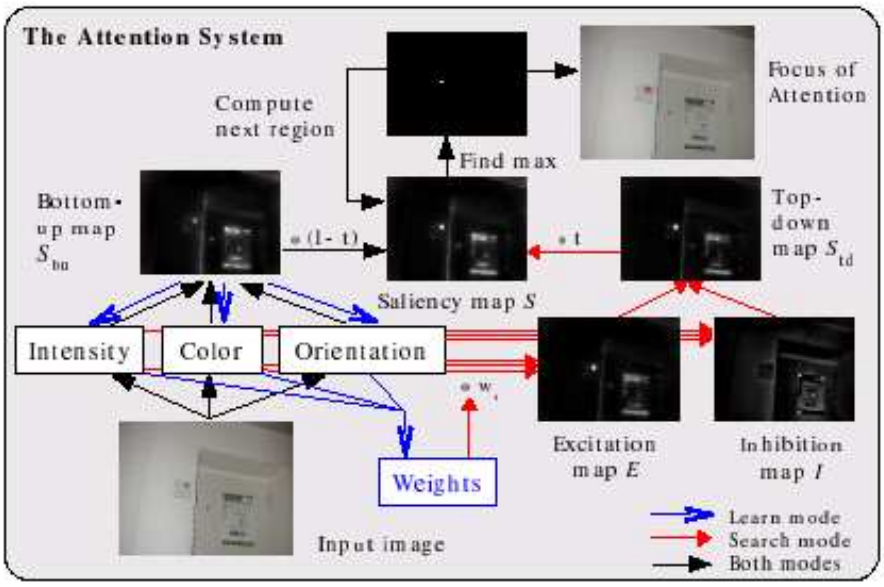



Figura 3.7: Sistema VOCUS de Frintrop (Frintrop et al., 2005)

queda, que utiliza los pesos resultantes del proceso de aprendizaje para ajustar los cálculos de saliencia y obtener un mapa final que permita llevar a cabo la selección atencional.

Durante el modo de aprendizaje, el proceso utiliza una imagen de entrenamiento junto con las coordenadas de la región de interés que contiene al objetivo. A partir de estos datos, el sistema calcula el mapa de saliencia *bottom-up* y la región más saliente (MSR) dentro de la región de interés especificada. A continuación, se determinan los pesos ($w_i$) de cada mapa de características ($X_i$) mediante la razón entre la media de saliencia de la región objetivo ($m_{MSR}$) y la del resto de la imagen ($m_{image-MSR}$): $w_i = m_{MSR}/m_{image-MSR}$. Si se utiliza más de una imagen de entrenamiento, este cálculo es realizado por cada una de ellas y los pesos finales se obtienen como resultado de realizar la media geométrica de los pesos calculados individualmente para cada imagen de muestra.

En el modo de búsqueda, se obtiene un mapa de saliencia *top-down* que es integrado con el mapa *bottom-up* para producir un mapa de saliencia global. El mapa *top-down*, a su vez, está compuesto por un mapa de excitación y otro de inhibición. El mapa de excitación ($E$) codifica las zonas con mayor probabilidad de contener el objetivo y se obtiene mediante la suma ponderada de los mapas de características cuyos pesos asociados son mayores que

1 (ecuación 3.5). El mapa de inhibición ($I$) representa las características que están más presentes en el fondo que en el objetivo, considerando sólo aquellos mapas cuyos pesos asociados son menores que 1 (ecuación 3.6).

$$E = \sum_i w_i * X_i \quad \forall i, w_i > 1 \tag{3.5}$$

$$I = \sum_i (1/w_i) * X_i \quad \forall i, w_i < 1 \tag{3.6}$$

Tras estos cálculos, se obtiene el mapa final de saliencia *top-down* ($S_{td}$) mediante la diferencia entre $E$ e $I$, anulando los valores negativos, y, por último, se normaliza al mismo rango que el mapa de saliencia *bottom-up* ($S_{bu}$).

Para obtener la saliencia final de cada posición en la imagen, integrando los elementos *bottom-up* y *top-down*, los mapas $S_{bu}$ y $S_{td}$ se combinan mediante una suma ponderada formando el mapa global de saliencia $S$ (ecuación 3.7). La contribución de cada mapa se ajusta a partir de un factor $t \in [0..1]$ que define la importancia de cada componente en el resultado final.

$$S = (1 - t) * S_{bu} + t * S_{td} \tag{3.7}$$

Una vez obtenido el mapa final, la región de máxima saliencia es seleccionada como foco de atención y, tras un cierto tiempo, dicha región es inhibida para permitir la selección de otras zonas salientes de la imagen.

## 3.8. Discusión

Esta revisión de los sistemas computacionales de atención visual permite realizar un análisis acerca de las tendencias principales que se han seguido hasta el momento sobre la organización de los mecanismos atencionales y la función de la atención como parte de un sistema más complejo. Los modelos descritos centran la cuestión de la atención en la exploración de una escena, en algunos casos desde una perspectiva puramente ascendente (Koch y Ullman, 1985)(Itti y Koch, 2000) y en otros combinando información contextual (Torralba, 2003a)(Torralba, 2003b) y/o conocimiento sobre propiedades relevantes de los objetivos visuales junto con los datos sensoriales. Dentro de las propuestas que integran

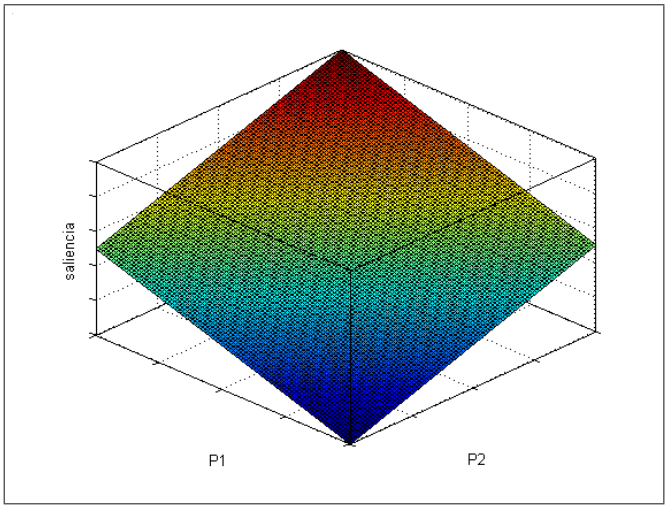
información descendente en el proceso de selección, podemos distinguir dos tipos: las que utilizan dicha información para inhibir las regiones que no concuerdan con las características del objetivo (Tsotsos et al., 1995)(Sun y Fisher, 2003) y las que aplican un mecanismo de ponderación de características para modular la función de saliencia de acuerdo con el objetivo (Navalpakkam y Itti, 2006)(Frintrop, 2006). En el primer caso, la influencia descendente sólo está presente como medio de exclusión. En el segundo grupo de propuestas, la información descendente afecta a todo el proceso atencional, por lo que proporcionan un control más completo y eficaz.

Ya sea a través de un control ascendente o descendente, en los modelos descritos la atención actúa como un filtro para el procesamiento que proporciona un uso eficiente de los recursos computacionales del sistema. Ahora bien, cuando el sistema de atención está integrado en un robot, cabe preguntarse si la selección para el procesamiento es suficiente para que la información visual que alcanza los procesos de alto nivel sea la adecuada en función de las necesidades de cada instante. Durante el transcurso de una tarea, la atención debe ser capaz de mantener una fijación sostenida sobre uno o varios estímulos, evitando las posibles distracciones y seleccionando el objetivo visual más apropiado para cada situación. En este sentido existen dos aspectos clave: ¿cómo modular la atención para seleccionar el mejor candidato en cada momento? y ¿cómo lograr que la atención alterne entre varios objetivos con diferentes propiedades? Desde la perspectiva de la selección para el procesamiento, estas dos cuestiones carecen de importancia puesto que el fin último es la detección de un objetivo desde los procesos de alto nivel a través de una actividad exploratoria del entorno. Sin embargo, cuando la atención está dirigida por y para la acción, este carácter pasivo no permite conseguir los comportamientos adecuados.

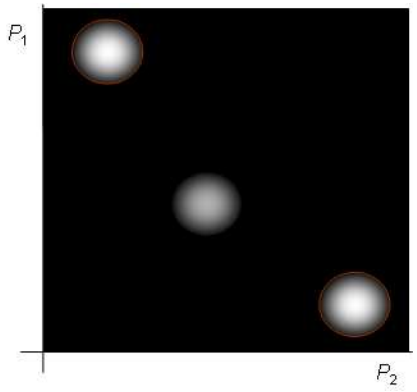
Uno de los principales objetivos de esta tesis es desarrollar un sistema de atención que incluya los aspectos planteados para ampliar la función atencional a la selección para la acción. Las propuestas existentes son difícilmente adaptables a nuestro propósito, en primer lugar, porque no permiten establecer criterios de selección que aseguren la fijación sobre el objetivo más adecuado en cada momento y, en segundo lugar, porque no permiten integrar varios objetivos simultáneamente.

La primera limitación puede observarse en las dos situaciones de las figuras 3.8 y 3.9. En ellas se muestra la aplicación del mecanismo de ponderación de características para la

selección de uno de entre varios estímulos a partir de dos propiedades $P_1$ y $P_2$. La figura 3.8 representa una situación en la que el objetivo debe cumplir simultáneamente ambas propiedades. En la figura 3.9, la selección se lleva a cabo considerando el cumplimiento independiente de una de las dos propiedades. En ambos casos, los pesos que ponderan las dos propiedades deben tener el mismo valor, puesto que las dos tienen la misma importancia en el proceso de selección atencional. Esto da lugar a una misma distribución de saliencia en el espacio de características para las dos situaciones (figuras 3.8a y 3.9a). No existe, por lo tanto, ninguna distinción entre considerar el cumplimiento conjunto o separado de varias propiedades. Además, como se observa en las dos figuras, el mecanismo de ponderación no siempre proporciona el estímulo que mejor se ajusta a los criterios de selección. Así, en la figura 3.8b, tendrían preferencia los dos estímulos que cumplen únicamente una propiedad frente al que cumple las dos. De manera similar, en la figura 3.9b, el estímulo de mayor saliencia cumpliría cada propiedad en menor medida que cualquiera de los otros dos.



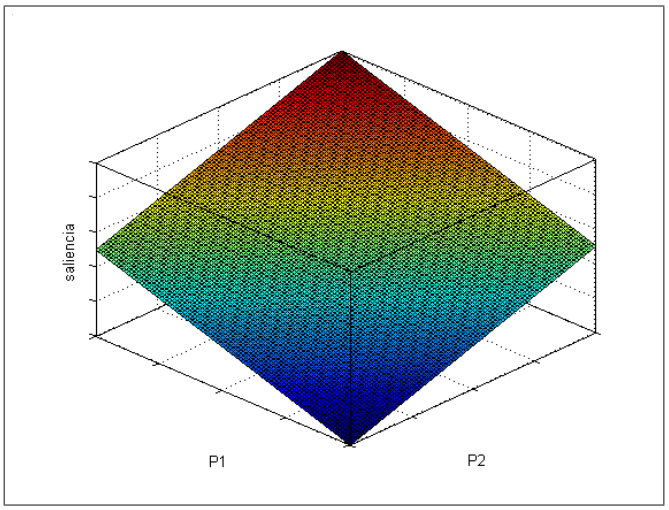

(a) Distribución de saliencia en el espacio de características



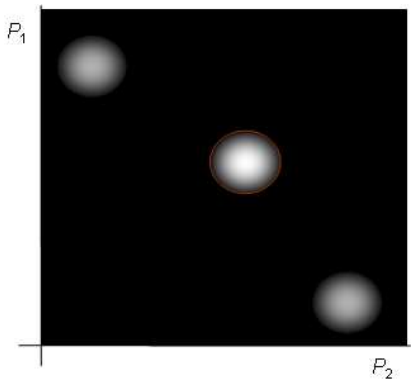

(b) Resultado de la asignación de saliencia a varios estímulos

Figura 3.8: Selección atencional mediante ponderación de características: cumplimiento simultáneo de dos propiedades

Desde el punto de vista de la acción, la fijación sobre varios estímulos es un aspecto clave de los mecanismos atencionales dado que, en la ejecución de una tarea, el robot deberá mantener varios objetivos conductuales que estarán guiados en muchas ocasiones por diferentes objetivos visuales. En dichas circunstancias, la atención deberá alternar entre varios estímulos, manteniendo un control abierto sobre uno de ellos y encubierto sobre el




(a) Distribución de saliencia en el espacio de características




(b) Resultado de la asignación de saliencia a varios estímulos

Figura 3.9: Selección atencional mediante ponderación de características: cumplimiento independiente de dos propiedades

resto. Los modelos descritos carecen de esta capacidad puesto que el control atencional se encuentra centralizado. Esta centralización implica, además, que no siempre será posible modular el sistema para mantener diferentes tipos de objetivos, ya que las propiedades que caracterizan a unos y a otros pueden ser contradictorias.

El modelo atencional propuesto en esta tesis intenta superar las limitaciones descritas anteriormente. En el capítulo 6 se presentarán los detalles concretos de la propuesta.

# Capítulo 4

# Arquitecturas de control en robots

La inteligencia artificial (IA) nació como un campo de investigación cuyo objetivo era construir sistemas inteligentes que modelaran los distintos aspectos de la inteligencia humana. La metodología clásica perseguía como objetivos fundamentales la habilidad de adquirir y utilizar conocimiento simbólico explícito para obtener una representación central del mundo y realizar algún tipo de razonamiento a partir del modelo construido. En su aplicación a la robótica, los sistemas desarrollados bajo esta filosofía presentaban muchos problemas a la hora de mostrar comportamiento autónomo en entornos complejos y sólo ofrecían un funcionamiento estable en entornos estructurados y altamente predecibles. A mediados de los años 80, surgen los sistemas reactivos como alternativa a la inteligencia artificial tradicional. Este nuevo enfoque supone una conexión directa entre la percepción y la acción, sin intervención de ningún sistema de representación, a través de una serie de asociaciones situación-acción. Dentro del paradigma reactivo, nacen los sistemas basados en comportamientos que presentan una capacidad reactiva alta, pero que permiten incorporar, además, conductas más complejas mediante la coordinación de una colección de unidades percepción-acción, denominadas comportamientos. Siguiendo esta línea, se han conseguido construir robots móviles que exhiben un rendimiento robusto en entornos complejos y dinámicos. Sin embargo, presentan dos problemas fundamentales: (1) ¿cómo determinar el conjunto de comportamientos que permiten resolver cada tarea de manera eficiente? y (2) ¿cuál debe ser el mecanismo de coordinación que decida qué acción ejecutar en cada momento? Para integrar las ventajas de las propuestas anteriores y como planteamiento intermedio, algunos investigadores proponen sistemas híbridos que incorporen la capacidad deliberativa de los sistemas concebidos bajo la inteligencia artificial clásica y la inmediatez y flexibilidad de los modelos reactivos.

Estos grandes grupos de propuestas conforman la taxonomía más utilizada para clasificar las arquitecturas de control de robots: sistemas de planificación o deliberativos; sistemas reactivos y basados en comportamientos; y sistemas híbridos. En este capítulo se revisan las principales características y limitaciones de las arquitecturas desarrolladas bajo estos enfoques.

## 4.1. Sistemas de planificación

Los sistemas de planificación o deliberativos, enmarcados dentro de la IA tradicional, se basan en la generación de una secuencia de acciones (plan) que permitan alcanzar un objetivo deseado partiendo de un estado inicial del sistema y de un modelo simbólico del mundo. La arquitectura de control consiste en una secuencia de componentes funcionales que resuelven diferentes fases del proceso. Desde los sensores hasta los actuadores, las fases más comunes del control en este tipo de sistemas son:

1. **Percepción**: se encarga de adquirir la información sensorial y de detectar, a partir de dicha información, características y objetos del entorno.

2. **Modelado**: construye un modelo simbólico del entorno a partir de la información obtenida de la fase anterior.

3. **Planificación**: opera sobre la descripción simbólica del mundo para producir la secuencia de acciones que permiten alcanzar el objetivo.

4. **Acción**: el plan obtenido en la fase anterior se hace efectivo mediante el envío de los comandos apropiados a los actuadores.

Uno de los primeros robots construidos bajo este esquema de control fue el robot Shakey (Nilsson, 1969)(Nilsson, 1984). Las tareas del robot consistían en navegar en diferentes habitaciones y empujar bloques de un lugar a otro, a la vez que se evitaban los posibles obstáculos. La arquitectura de control utiliza el planificador simbólico STRIPS (Fikes y Nilsson, 1971) y está organizada en 3 capas que representan distintos niveles de urgencia de las acciones. La capa de bajo nivel está formada por acciones que se disparan directamente a partir de la información sensorial, sin ningún proceso deliberativo intermedio, o por la ejecución de acciones de mayor nivel para cumplir un plan específico. La capa intermedia

combina las acciones de bajo nivel para dar lugar a comportamientos más complejos. Por último, la capa de alto nivel se encarga de generar y ejecutar un plan mediante una secuencia de acciones del nivel intermedio. Esta arquitectura de tres capas no debe confundirse con una arquitectura de control híbrida ya que, aunque las acciones de bajo nivel pueden ser interpretadas como comportamientos reflejos, el sistema tiene una base fundamentalmente simbólica. Así, aunque estas acciones se ejecuten con independencia del nivel de planificación, el sistema genera y mantiene siempre un modelo simbólico del mundo que es enviado a los niveles superiores para deliberar sobre la situación actual.

Desde Shakey se han desarrollado otras arquitecturas de control deliberativas, entre las cuales pueden considerarse representativas el proyecto Hilare (Giralt et al., 1984), la arquitectura SOAR (Laird et al., 1987) o las arquitecturas NASREM (Albus et al., 1989) y RCS (Albus, 1993). Todas estos modelos mantienen características comunes que permiten definir las bases principales de este paradigma (Arkin, 1998):

- Mantienen una estructura jerárquica con una división claramente funcional.
- La comunicación y el control suceden de manera predecible y predeterminada.
- Los niveles más altos de la jerarquía proporcionan subobjetivos a los niveles inferiores.
- El alcance temporal y espacial de la planificación varía durante el descenso en la jerarquía. En los niveles inferiores, los requerimientos de tiempo son menores y las consideraciones espaciales más locales.
- Su funcionamiento está basado en gran medida en modelos de representación simbólica del mundo.

A pesar de los esfuerzos y de las diferentes propuestas, los sistemas construidos bajo este enfoque presentan serias limitaciones cuando se implantan en entornos reales. Entre los principales problemas detectados destacan los siguientes:

- **Tiempo elevado de procesamiento**: el modelado y la planificación son procesos secuenciales que requieren un tiempo de cómputo excesivo dando lugar a sistemas poco adecuados para escenarios reales, en los que los tiempos de reacción del robot deben ser lo suficientemente rápidos para que cualquier tarea se ejecute correctamente.

- **Necesidad de un modelo preciso del mundo**: la actuación conlleva la generación de un plan que, a su vez, necesita de un modelado preciso del mundo. Esto sólo es posible con sensores de alta precisión que proporcionen medidas reales con una mínima incertidumbre.
- **Problema del marco**: el sistema debe decidir de manera ágil qué información de la representación global del mundo es importante para la actividad actual, teniendo en cuenta que cualquier porción puede afectar a la totalidad del sistema.
- **Problema de la correspondencia símbolos-mundo real**: el modelo del mundo es puramente simbólico, mientras que el robot actúa en un entorno real. El correcto funcionamiento del sistema depende de la asociación adecuada entre símbolos y percepciones reales.

## 4.2. Sistemas reactivos y basados en comportamientos

En el extremo opuesto a los sistemas deliberativos se encuentra el enfoque puramente reactivo en el que se plantea una asociación directa entre la percepción y la acción que dote al robot de la capacidad de producir respuestas inmediatas en su interacción con un mundo dinámico y no estructurado. Con este fin, el control se lleva a cabo a través de una colección de reglas condición-acción que permiten disparar una acción apropiada en la situación definida por la percepción actual del mundo (Agre y Chapman, 1987). Aunque con este nuevo enfoque se consiguieron construir robots que se desenvolvían de manera adecuada en entornos reales, los sistemas puramente reactivos no están exentos de inconvenientes. Sus principales limitaciones residen en la carencia de estados internos, lo que reduce su actuación al momento presente, anulando cualquier tipo de anticipación y eliminando la posibilidad de aprender nuevos comportamientos.

Dentro del enfoque reactivo, surgen los sistemas basados en comportamientos liderados por los trabajos de Rodney Brooks (Brooks, 1986). Aunque estos sistemas mantienen las ideas principales de los sistemas reactivos, existen diferencias fundamentales entre ellos (Matarić, 1992). Los aspectos clave de las arquitecturas basadas en comportamientos, compartidos en su mayoría por el enfoque reactivo, pueden sintetizarse en los siguiente puntos (Brooks, 1991a):

- *El mundo es el mejor modelo de sí mismo* : no existe ningún modelo interno del

mundo. Los sistemas basados en comportamientos, y en general los sistemas reactivos, utilizan la información sensorial en lugar de acceder a una representación interna resultante de un modelado simbólico del mundo.

- Acción situada: el robot está situado en un mundo real. Por este motivo, no debe actuar en función de una representación abstracta de la realidad; debe utilizar la percepción del mundo real.
- Corporeidad: el robot tiene una presencia física (un cuerpo). Esto implica que está sujeto a cualquier influencia de su entorno, por lo que sus dinámicas de interacción con el mundo no pueden ser validadas completamente en una simulación. El sistema debe funcionar en un cuerpo real que opera en un mundo físico.
- Emergencia: el comportamiento complejo emerge de la interacción del robot y su entorno a través de comportamientos básicos.

En esencia, una arquitectura basada en comportamientos consiste en una colección de comportamientos que actúan en paralelo cumpliendo objetivos determinados. Los comportamientos se implementan como leyes de control, que toman su entrada de los sensores y/o de otros comportamientos, y envían su salida a los actuadores y/o a otros comportamientos. A diferencia de las reglas definidas en un sistema reactivo, los comportamientos incluyen estados internos, lo que les dota de una cierta capacidad deliberativa. Puesto que los comportamientos actúan en paralelo, una arquitectura de este tipo debe incluir algún mecanismo de coordinación que asegure el envío de un único comando a los actuadores.

### 4.2.1. Algunas arquitecturas basadas en comportamientos

En 1986, Rodney Brooks diseño la arquitectura de subsunción (Brooks, 1986). Su trabajo constituyo el punto de partida de la robótica basada en comportamientos.

La arquitectura de subsunción consiste en una distribución del control en diferentes capas denominadas niveles de competencia. Cada nivel es una especificación informal de los comportamientos del robot y persigue un objetivo individual. Todos ellos tienen acceso directo a la información sensorial y funcionan de manera concurrente y asíncrona. La organización de los niveles es jerárquica, de manera que los más altos pueden inhibir o suprimir las señales de control de los niveles inferiores. La supresión sustituye la señal de control

del nivel inferior por la procedente del nivel superior. La inhibición anula la transmisión de la señal hacia los actuadores. Estos dos mecanismos constituyen el método de coordinación de la arquitectura, en el que se emplea una estrategia competitiva con prioridades fijas.

Otras de las propuestas de mayor influencia en el enfoque basado en comportamientos son las arquitecturas basadas en esquemas. El concepto de esquema surge dentro de la psicología y la neurología para describir modelos que expliquen la organización de los procesos mentales en la percepción y en la respuesta coordinada a los estímulos del entorno. La primera aplicación de la teoría de esquemas a la robótica nace de los trabajos de Arbib (Arbib, 1981), pero es sin duda la arquitectura de Esquemas Motores de Arkin (Arkin, 1987) la más representativa dentro de este grupo de modelos.

Arkin define esquema como "la unidad básica de comportamiento desde la que se pueden construir acciones complejas; consiste en el conocimiento sobre cómo actuar o percibir así como en el proceso computacional que lo representa" (Arkin, 1993). Existen dos tipos de esquemas: los esquemas motores, que están relacionados con el control de los actuadores, y los esquemas perceptivos, que proporcionan información del entorno. Cada esquema perceptivo está orientado a una tarea, por lo que no toda la información sensorial es procesada, únicamente la implicada en los objetivos del esquema. Cada esquema motor recibe datos sensoriales del esquema perceptivo asociado y genera un vector de acciones (orientación y velocidad) que define la forma en la que el robot debe moverse de acuerdo con el estímulo percibido. El método de coordinación empleado es cooperativo y consiste en la suma vectorial de los vectores generados por los esquemas motores activos. Esto permite generar un único vector de movimiento que será el que finalmente se envíe a los motores.

Las dos arquitecturas descritas pueden considerarse los sistemas pioneros dentro del campo de la robótica basada en comportamientos. Sin embargo, son numerosas las propuestas surgidas a partir de ellas. Todas comparten los principios fundamentales de este enfoque: relación directa percepción-acción, ausencia de conocimiento simbólico, y uso de unidades de comportamiento como elementos básicos con los que se construye la arquitectura. Las principales diferencias entre ellas radican en la forma en la que los comportamientos interactúan y en el método utilizado para coordinarlos.

### 4.2.2. Mecanismos de coordinación: el problema de la selección de la acción

En los sistemas basados en comportamientos, el control del robot está distribuido en múltiples comportamientos con objetivos diferentes y, en algunos casos, incompatibles. Cada comportamiento es responsable de alcanzar o mantener un determinado objetivo. El objetivo de un comportamiento puede estar en conflicto con los objetivos de otros (por ejemplo, alcanzar una posición final y evitar obstáculos), lo que requiere una toma de decisión sobre qué acción concreta se debe llevar a cabo en cada momento. Esta situación es conocida como el problema de la selección de la acción y debe ser solucionada incluyendo en el sistema algún mecanismo de coordinación entre comportamientos.

Existen numerosos métodos de coordinación (o de selección de la acción) dentro de las propuestas basadas en comportamientos. No obstante, todos ellos pueden incluirse en uno de entre dos grandes grupos: métodos de arbitraje y métodos de fusión de comandos (Saffiotti, 1997). El arbitraje consiste en determinar qué comportamiento debe influir en la acción final del robot en cada instante. La fusión de comandos se basa en obtener una única salida motora a través de la combinación de las acciones de los diferentes comportamientos.

Dentro de esta clasificación, pueden hacerse subdivisiones más específicas en función de los mecanismos concretos utilizados para la selección de la acción. Pirjanian (Pirjanian, 1998) distingue entre métodos basados en la prioridad, basados en el estado y mecanismos de "el ganador lo toma todo" como subcategorias de arbitraje. Asimismo, también subdivide las propuestas de fusión de comandos en métodos de votación, borrosos y de superposición. A continuación, se describen brevemente cada uno de estos mecanismos, algunos de ellos propios no sólo del enfoque basado en comportamientos, sino también de las arquitecturas híbridas.

- Métodos de arbitraje
  - Basados en la prioridad (Brooks, 1986): la acción es seleccionada en función de prioridades asignadas a priori a cada comportamiento. En cada situación, el control es otorgado al comportamiento activo de mayor prioridad.
  - Basados en el estado (Košecká y Bajcsy, 1993)(Arkin y MacKenzie, 1994): la selección de comportamientos se realiza mediante transiciones de estado que

se producen tras la detección de determinados eventos y provocan cambios de comportamientos.

- "El ganador lo toma todo" (Maes, 1989): los sistemas que utilizan esta técnica consisten en un conjunto de comportamientos conectados formando una red. Cada comportamiento tiene asociado un nivel de activación que cuantifica la aplicabilidad del comportamiento a una situación. Cuando esta medida sobrepasa un umbral, el comportamiento es seleccionado para que ejecute su acción. Los niveles de activación son modificados por fuentes internas y externas siguiendo determinados criterios.

▪ Métodos de fusión

- Votación (Rosenblatt, 1995)(Riekki y Röning, 1997)(Hoff y Bekey, 1995): cada comportamiento realiza una votación a favor o en contra de las posibles acciones. Estos votos se combinan y la acción con el máximo valor es seleccionada para su ejecución.
- Borrosos (Saffiotti et al., 1995)(Yen y Pfluger, 1995): la filosofía es similar a la de la propuesta anterior, sólo que en este caso las salidas de los distintos comportamientos se combinan utilizando lógica borrosa.
- Superposición (Arkin, 1987): las salidas generadas por los comportamientos son interpretadas como campos de fuerza y el robot como una partícula que se ve afectado por estos campos y busca una configuración estable.

La elección de un método concreto de selección de la acción no es una cuestión sencilla. Los mecanismos anteriores resultan adecuados en ciertas situaciones, pero no responden correctamente ante otras. En general, el arbitraje permite responder de forma segura ante situaciones de peligro, dado que la acción se concentra sobre un único comportamiento. No obstante, cuando se requiere la actuación de varios comportamientos en tareas en las que el robot debe mantener objetivos múltiples, las acciones de los comportamientos que se disparan consecutivamente pueden llegar a contradecirse mutuamente produciendo cambios bruscos de posición y/u orientación o, incluso, situaciones de bloqueo. Por el contrario, la fusión de comandos proporciona respuestas más suaves que el arbitraje al considerar las acciones individuales de los diferentes comportamientos, pero esta misma cualidad provoca que, en situaciones críticas, la acción resultante ignore cuestiones de seguridad.

## 4.3. Sistemas híbridos

Los sistemas de control híbridos nacen como un intento de integrar las ventajas de los enfoques reactivos/basados en comportamientos y de los modelos deliberativos. Aunque las arquitecturas basadas en comportamientos muestran un rendimiento robusto en entornos complejos y dinámicos, no siempre resultan ser la mejor opción. En ciertas ocasiones, la incorporación de alguna forma de conocimiento en el sistema permite que la ejecución de una tarea se lleve a cabo de una manera más eficiente, introduciendo aspectos como la anticipación o la descomposición de un objetivo en subobjetivos de menor complejidad. De la integración de estos aspectos, propios de los sistemas deliberativos, en las arquitecturas basadas en comportamientos surge un nuevo grupo de propuestas que se conocen como arquitecturas híbridas.

La cuestión principal del diseño de una arquitectura de este tipo es cómo combinar las capacidades reactivas y deliberativas de manera adecuada en el control global. En general, el sistema está compuesto al menos por dos capas, una capa deliberativa encargada de modelar el mundo, razonar sobre él y crear planes, y una capa reactiva responsable de la ejecución de los planes y de reaccionar rápidamente ante situaciones inesperadas. No obstante, muchas arquitecturas híbridas incluyen una tercera capa que se encarga de la coordinación entre los componentes reactivos y deliberativos.

Una de las primeras arquitecturas híbridas fue AuRA, propuesta por Arkin (Arkin, 1989). Se trata de una arquitectura de dos capas formada por un componente jerárquico deliberativo y un componente reactivo basado en esquemas. La parte deliberativa incluye: un planificador de misiones, encargado de establecer objetivos de alto nivel; un razonador espacial, que utiliza información cartográfica almacenada en una memoria a largo plazo para construir una secuencia de trayectorias que el robot debe seguir para completar su objetivo; y un secuenciador de planes, responsable de transformar las trayectorias generadas por el razonador espacial en un conjunto de comportamientos (esquemas) motores. En la parte reactiva, el controlador de esquemas se encarga de controlar y monitorizar los comportamientos activados por el secuenciador de planes. Cada esquema motor está asociado con un esquema perceptivo que le envía la información sensorial necesaria para su ejecución. Los esquemas activos producen respuestas vectoriales que son combinadas para producir una única respuesta motora, siguiendo la técnica de fusión de comandos. Una vez que el componente reactivo comienza su ejecución, el componente deliberativo permanece

desactivado hasta que la misión se completa o se detecta un fallo.

Un ejemplo de arquitectura de 3 capas es Atlantis, desarrollado por Gat (Gat, 1992). El sistema está estructurado en: un nivel deliberativo, que se ocupa de la planificación y del modelado del mundo; un secuenciador, encargado de iniciar y finalizar las actividades de bajo nivel, así como de afrontar los fallos del sistema reactivo para poder completar una tarea; y un controlador reactivo, responsable de controlar las actividades de bajo nivel. Los niveles actúan asíncronamente. El nivel deliberativo responde a las peticiones del secuenciador y sus resultados son tratados como recomendaciones, es decir, no son ejecutados necesariamente. El secuenciador monitoriza las tareas en ejecución de manera que puede detectar cuando no es posible completar una tarea. Esto proporciona una vía para reestructurar los planes de acuerdo a cada situación.

Los distintos trabajos desarrollados desde el enfoque híbrido plantean diferentes formas de interacción entre la planificación y el control reactivo. Englobando las propuestas más representativas, Arkin define cuatro estrategias principales (Arkin, 1998):

- Selección: el componente de planificación determina qué comportamientos deben activarse, así como los parámetros utilizados durante su ejecución (Arkin, 1989).
- Recomendación: el planificador sugiere cambios de actuación que pueden ser llevados a cabo o no por el sistema de control reactivo (Gat, 1992).
- Adaptación: el planificador modifica las ejecuciones del componente reactivo para adaptar sus respuestas a las condiciones actuales (Lyons y Hendriks, 1995).
- Aplazamiento: los planes sólo son desarrollados cuando es necesario (Georgeff y Lansky, 1987).

Las arquitecturas híbridas son las más utilizadas hoy en día. Su éxito radica en combinar las ventajas de los sistemas reactivos y deliberativos, ambos con diversas limitaciones cuando son considerados de manera aislada. No obstante, el diseño de una arquitectura de este tipo requiere de un compromiso entre los dos enfoques que no es fácil definir. La cuestión principal es cómo establecer los límites entre la deliberación y la ejecución reactiva. Los diferentes planteamientos sobre este asunto han dado lugar a un amplio abanico de propuestas que no permiten llegar a una solución homogénea.

## 4.4. Conclusiones

En este capítulo se han presentado los principales fundamentos y características de los distintos tipos de arquitecturas de control para robots. De un lado, las arquitecturas deliberativas, que se basan en la utilización de un modelo simbólico del mundo para generar secuencias de acciones que lleven a la consecución de un objetivo. En el extremo opuesto, nos encontramos con los sistemas puramente reactivos, que emplean una estrategia de control basada en la definición de una colección de reglas condición-acción. Este tipo de sistemas no mantienen ningún modelo interno del mundo y básicamente se encargan de producir la respuesta más adecuada a la entrada sensorial recibida en cada instante. Dentro del enfoque reactivo, se encuentran los sistemas basados en comportamientos, en los que el control está distribuido en una serie de unidades que enlazan la percepción con la acción. El comportamiento complejo emerge a través de la interacción entre estas unidades, proporcionando conductas más ricas que los sistemas puramente reactivos. Por último, existe otro grupo de propuestas enmarcadas dentro de las denominadas arquitecturas híbridas, que se plantean como una forma de integración de los enfoques anteriores a partir de la inclusión de componentes reactivos y deliberativos en el control global.

Siguiendo un camino diferente al eje definido entre lo inmediato y lo modelado, el dinamicismo propone modelar el control inteligente utilizando exclusivamente la teoría de los sistemas dinámicos y, normalmente, estructuras conexionistas como soporte (Johnson et al., 2007). Este campo está menos desarrollado en su aplicación a la robótica aunque es previsible que se consigan importantes avances en un futuro próximo.

El sistema de control propuesto en esta tesis se basa en las relaciones entre atención y acción para generar comportamiento autónomo. En este sentido, el tipo de arquitectura que mejor se adapta a nuestro propósito es el de las arquitecturas basadas en comportamientos, aunque con ciertos matices. El más importante es que, en nuestra propuesta, la percepción y la acción están conectadas a través de la atención. Así, un comportamiento dentro del sistema puede definirse como una unidad de enlace entre atención y acción, más que entre percepción y acción. Además, partimos de la base de que los comportamientos pueden tener una naturaleza tanto reactiva como deliberativa. Es decir, se admite la posible coexistencia de comportamientos dedicados a generar acciones motoras que proporcionen respuestas adecuadas a diferentes situaciones, así como comportamientos cuya función implique mantener cierta información del entorno para realizar algún tipo de razonamiento

sobre la situación actual. No se pretende con este enfoque definir una arquitectura híbrida en el sentido expuesto en la sección 4.3, pero sí plantear la posibilidad de hibridación desde las propuestas basadas en comportamientos. La definición de comportamientos de carácter deliberativo dotaría al sistema de conductas más ricas que no se limitarían a dar respuestas inmediatas a cada situación, sino también a mantener una existencia efectiva y duradera del robot en su entorno.

# Parte II

# Propuesta

# Capítulo 5

# Arquitectura hardware y software del sistema propuesto

Antes de describir la propuesta concreta de esta tesis, se presentan en este capítulo los distintos aspectos y decisiones de diseño hardware y software que se han llevado a cabo para poner en marcha el sistema y lograr su funcionamiento en un robot móvil.

## 5.1. La plataforma robótica

"*La mente no puede reducirse al cerebro más de lo que puede hacerlo a la cultura. La mente no es simplemente una computadora, en el sentido tradicional del término; podemos decir que realiza cálculos, pero la sustancia y estructura de estos cálculos son debidas al tipo de cuerpo que tenemos y a los entornos en los que habitamos.*"(Rohrer, 2006)

Durante los últimos años, la robótica ha dado un vuelco en los principios que rigen la construcción de sistemas autónomos. La *nueva robótica* defiende el estudio de las interacciones entre un robot y su entorno como base para la generación de comportamiento inteligente. Desde esta perspectiva, definir un sistema robótico con capacidad de autonomía conlleva enmarcar el problema en un sistema físico concreto, el cuerpo del robot.

El sistema propuesto ha sido diseñado para un robot móvil compuesto por una cabeza de visión estereoscópica situada sobre una base móvil.

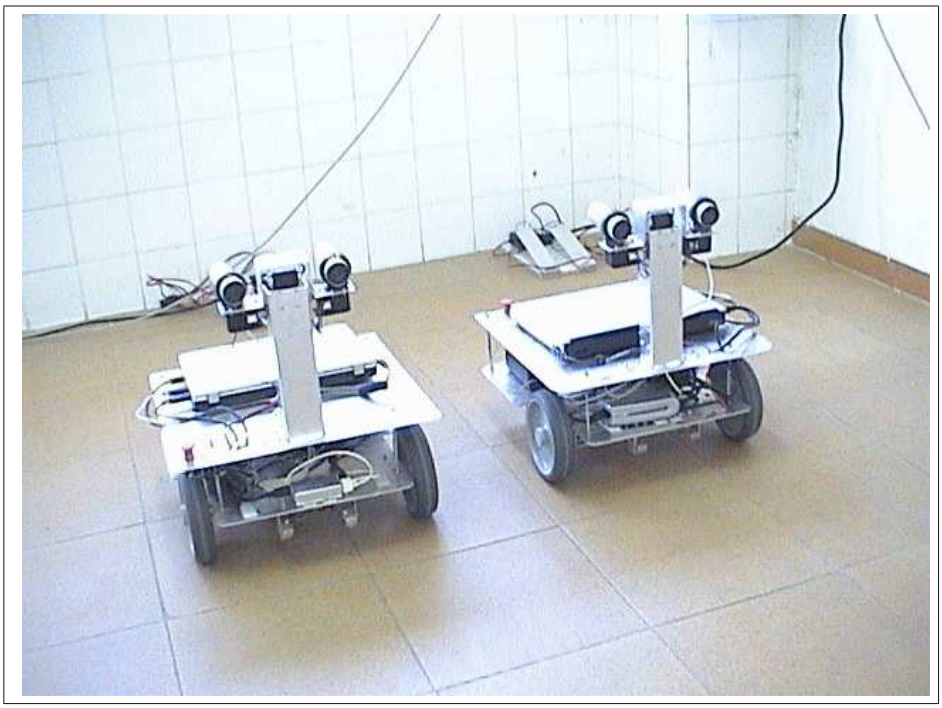


Figura 5.1: Robots utilizados para la puesta en marcha del sistema propuesto

### 5.1.1. Diseño de la torreta estéreo

La torreta estéreo (figura 5.2) está formada por una estructura en forma de U invertida con dos pies sobre los que se sitúan las cámaras. Cada cámara está acoplada a su base a través de una sujeción ensamblada a un motor que proporciona un giro independiente sobre el eje vertical. En el interior de la U, en uno de sus laterales, se sitúa un tercer motor que hace girar toda la estructura, proporcionando un movimiento rotatorio simultáneo de ambas cámaras sobre el eje horizontal. Este diseño admite, por lo tanto, 3 grados de libertad sobre la cabeza robótica. No obstante, el rango de giro de cada cámara está limitado a causa del diseño mecánico para evitar el choque de éstas con la estructura, por lo que no todos los movimientos son posibles. Para evitar esta limitación, el control de la torreta se apoya en la base, haciendo que ésta gire a modo de cuello en la dirección adecuada cuando las cámaras por sí mismas no pueden alcanzar la posición deseada.

Las cámaras utilizadas en la torreta estéreo son de la marca iSight de Apple. Este modelo de cámara digital tiene un sensor CCD de 1/4" con resolución VGA de 640x480 píxels. Es capaz de capturar a través de la interfaz IEEE 1394 a una velocidad de 30 imágenes

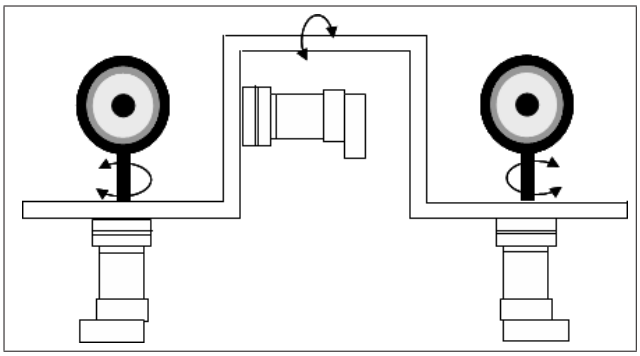


Figura 5.2: Vista frontal de la torreta estéreo

por segundo en 24 bits de color. El sensor está acoplado a un mecanismo motorizado de enfoque que permite un funcionamiento en modo auto-enfoque. También es posible establecer un control automático de exposición. Estos y otros parámetros operacionales pueden ser controlados a través de la interfaz IEEE 1394.

### 5.1.2. Diseño de la base móvil

La base está constituida por una plataforma y tres ruedas, dos de ellas motrices, colocadas en la parte delantera, y una tercera de giro libre, situada en la parte trasera. Las dos ruedas delanteras conforman el sistema de tracción diferencial del vehículo, siendo éstas accionadas de modo independiente. La rueda trasera sirve de apoyo y se orienta en la dirección correcta dependiendo del movimiento de las motrices. Este diseño es uno de los menos complicados de construir y proporciona un control sencillo a partir de la combinación de velocidades de las dos ruedas motrices, permitiendo que el robot se mueva en línea recta, trace curvas o gire sobre sí mismo.

Para dotar al robot de suficiente autonomía, debe realizarse un diseño mecánico de la plataforma que permita transportar los diferentes elementos de control, comunicación y alimentación necesarios para su funcionamiento. Con este fin, se ha optado por una plataforma de dos alturas (figura 5.3) que incluya, en su parte inferior, la electrónica encargada de alimentar al robot y los circuitos para el control de los motores y, en la parte superior, la torreta estéreo, otros sistemas sensoriales y un ordenador portátil que actúe como controlador de alto nivel.

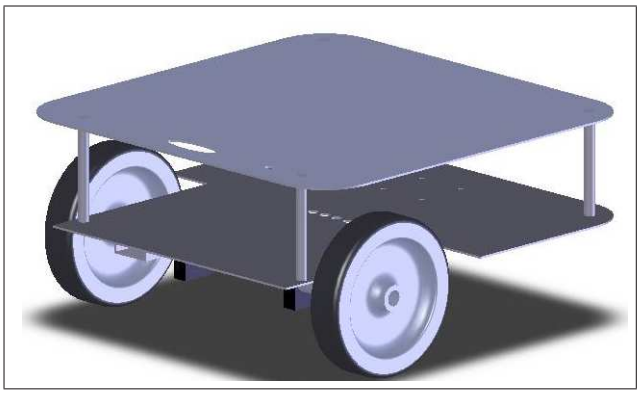


Figura 5.3: Modelo de la base móvil del robot

### 5.1.3. Control motor

El control motor de la torreta estéreo se lleva a cabo a través de 3 actuadores de la serie RX-10 de Dynamixel, cada uno de los cuales integra reductora, motor de precisión y circuito de control. Los tres actuadores se encuentran conectados a un único bus TTL junto con un controlador principal que se encarga de la comunicación con el PC a través del puerto RS232. El controlador principal se comunica con los actuadores enviando y recibiendo paquetes de datos. Los paquetes pueden ser de dos tipos: instrucciones (enviados desde el controlador principal a los actuadores) o paquetes de estado (enviados de los actuadores al controlador principal). Dentro de los campos de una instrucción, se incluye el identificador del actuador al que va dirigida. Esto permite que, aunque los datos sean recibidos por todas las unidades conectadas, sólo una de ellas responda enviando su estado y ejecutando la operación que corresponda.

El movimiento de la plataforma está controlado por dos motores de corriente continua de la casa Maxon que disponen de encoder óptico y de reductora con rodamiento metálico de salida. El sistema utilizado para llevar a cabo el control de ambos motores consiste en un circuito diseñado en el laboratorio, basado en el microcontrolador Atmega32 de Atmel, y de otros circuitos auxiliares. Las funciones principales que realiza son:

- Control de posición mediante PID
- Control de velocidad mediante PID
- Odómetro

- Control de dispositivos I2C

Para procesar las señales de los encoders se ha incorporado al circuito el contador LS7266R1, que proporciona directamente las posiciones de dos encoders mediante las señales en cuadratura que le llegan de éstos. La inclusión de este bloque dentro del circuito evita que el microcontrolador esté continuamente calculando la posición de los motores.

El sistema de control también se encarga de realizar las operaciones y cálculos necesarios para el control PID de los motores. El bucle de control PID de cada motor funciona a una frecuencia de 1 KHz, proporcionando un tiempo de actualización de 1 milisegundo.

## 5.2. Arquitectura software

La complejidad de un sistema software como el desarrollado en esta tesis prácticamente obliga a utilizar técnicas de programación que permitan aumentar la funcionalidad del sistema añadiendo nuevos bloques y manteniendo los ya existentes, así como sustituir ciertos módulos sin que dicha sustitución conlleve la reconstrucción de todo el sistema. Con este objetivo, se ha empleado una metodología de diseño software basada en componentes (CBSE) (Szyperski, 1998) (He et al., 2005) para construir el sistema propuesto. Esta rama de la ingeniería del software enfatiza la descomposición de sistemas en componentes funcionales que son diseñados y construidos independientemente a la vez que mantienen la capacidad de interoperabilidad. Para ello, cada componente debe describir de forma completa las interfaces que ofrece, así como las interfaces que requiere para su operación, y debe operar correctamente con independencia de los mecanismos internos que utilice para soportar la funcionalidad de la interfaz. Los componentes se consideran elementos de un nivel de abstracción superior al de los objetos y deben mantener una serie de propiedades (Szyperski, 1998):

- *Interfaces especificadas contractualmente*: un componente tendrá asociado una interfaz que especifique su comportamiento de cara al exterior y que permita que el componente sea usado como una *caja negra*.
- *Dependencias contextuales completamente explícitas*: la definición de un componente conlleva no sólo la especificación de los servicios que proporciona, sino también de los que requiere.

- *Utilización independiente de la implementación*: la utilización de un componente dentro de un sistema no debe conllevar la modificación de su código fuente.
- *Composición independiente del sistema*: un componente debe poder utilizarse en sistemas que requieran de su funcionalidad y para los que no fue concebido inicialmente.

Todas estas propiedades permiten mantener los principales objetivos de la ingeniería del software basada en componentes (Medvidovic y Taylor, 2000) (Roshandel et al., 2004) (He et al., 2005):

- Construir sistemas rápida y eficazmente incorporando componentes ya existentes.
- Construir sistemas más fiables incorporando componentes que hayan sido probados dentro de múltiples proyectos.
- Obtener sistemas modulares con dependencias explícitas.
- Construir sistemas flexibles en los que los componentes puedan ser reemplazados por otros.

### 5.2.1. Plataforma de desarrollo distribuido

Para que una arquitectura de componentes pueda funcionar adecuadamente es necesario disponer de un entorno normalizado que proporcione soporte a los mecanismos de comunicación entre interfaces. Cada uno de los componentes software que constituyen el sistema desarrollado ha sido implementado como un proceso Unix, escrito en C++, que utiliza una capa software de desarrollo de aplicaciones distribuidas para implementar la comunicación entre interfaces. Entre las distintas plataformas de desarrollo distribuido, se ha optado por utilizar Ice (ZeroC, 2007) frente a otras como DCOM (Grimes, 1997) o CORBA (Henning y Vinoski, 1999), por las distintas ventajas que presenta frente a ellas :

- DCOM se planteó como una solución de Microsoft que no podía ser utilizada en redes heterogéneas que contuvieran máquinas con otros sistemas operativos. Sin embargo, Ice proporciona una plataforma de desarrollo orientado a objetos adecuado para su uso en entornos heterogéneos. Actualmente existe una versión de DCOM en entorno Linux/Unix, aunque su uso no está demasiado extendido y el proyecto puede considerarse inmaduro por el momento.

- Tanto DCOM como CORBA mantienen una complejidad excesiva, convirtiendo el desarrollo de aplicaciones sobre cualquiera de las dos plataformas en una tarea demasiado complicada. En cambio, en Ice, el desarrollo de aplicaciones resulta mucho más sencillo. Por ejemplo, para C++, Ice utiliza el estándar STL, lo que facilita su uso en programas escritos en este lenguaje.

- Ice soporta una gran variedad de lenguajes de programación como C++, Java, Python, Ruby, PHP, C# y Visual Basic. No ocurre así con CORBA. De hecho, la mayoría de versiones comerciales sólo ofrecen soporte para C++ y Java.

- Aunque existen diferentes implementaciones de CORBA disponibles, es difícil encontrar una que proporcione un rendimiento satisfactorio. Sin embargo, Ice proporciona una implementación eficiente en cuanto a ancho de banda de red, uso de memoria y carga de CPU.

- Ice ofrece un lenguaje sencillo para descripción de interfaces (*Slice*), que puede ser compilado en los distintos lenguajes de programación para los que da soporte. Slice ofrece una serie de ventajas frente a IDL, su equivalente en CORBA, resultando menos complicado en cuanto a las construcciones del lenguaje, pero a la vez más flexible.

### 5.2.2. Sistemas reusables en robótica

Existen varios proyectos abiertos de robótica que mantienen repositorios de implementaciones reusables de algoritmos para robots móviles. Entre las más destacables están CARMEN (Montemerlo et al., 2003), Player (Gerkey et al., 2001) (Gerkey et al., 2003) y Orca (Brooks et al., 2005) (Brooks et al., 2007). CARMEN (Carnegie Mellon Navigation Toolkit) es una recopilación de software de control de robots que proporciona una interfaz consistente y un conjunto de primitivas de aplicación a la investigación sobre robots móviles en una amplia variedad de plataformas comerciales. Su aplicación, sin embargo, está limitada a vehículos de navegación en interiores. Player es un servidor de dispositivos que proporciona control para una extensa variedad de sensores y actuadores. Aunque no fue diseñado explícitamente utilizando los principio de CBSE, utiliza muchas de las ideas de esta metodología. La principal limitación de este proyecto es la separación estricta entre los espacios de servidores y clientes, lo que dificulta el diseño de sistemas en los que se requiera que los componentes actúen a la vez como clientes y como servidores. ORCA es un proyecto

software de código abierto, diseñado siguiendo técnicas basadas en componentes, en el que se propone un repositorio de componentes software para aplicaciones a la robótica. En sus primeras implementaciones, utilizaban CORBA para la comunicación entre componentes, pero, posteriormente, consideraron Ice como una mejor alternativa y es ésta la plataforma de comunicación utilizada en sus versiones más recientes.

El diseño software que se plantea en este trabajo es similar al propuesto en ORCA. No obstante, no existen componentes relacionados con procesamiento visual complejo dentro de este repositorio, por lo que el uso de ORCA requería el diseño de nuevos componentes desde el comienzo. Otro de los motivos por los que se planteó el uso de un diseño propio de componentes es que en ORCA se emplean diferentes mecanismos de comunicación, algunos de los cuales no resultan adecuados para nuestras aplicación y obligan a complicar la estructura de componentes. Este es el caso de los métodos de comunicación basados en la subscripción de clientes a servidores, en los que los clientes reciben los datos de los servidores asíncronamente. Cuando el envío es de datos masivos, como ocurre en aplicaciones de visión, esta comunicación asíncrona puede provocar un tráfico excesivo e innecesario en la red, resultando más conveniente que el envío de datos se realice a la frecuencia de procesamiento de los clientes y no a la de los servidores. Dada la escasa aplicabilidad de estos mecanismos en nuestro sistema, en el diseño propuesto se suprime la comunicación cliente-servidor por subscripción, simplificando así la estructura de componentes.

### 5.2.3. Estructura de un componente software

Las aplicaciones de clientes y servidores en Ice presentan una estructura lógica como la que se muestra en la figura 5.4. Cada una de ellas está formada por código de la aplicación, librerías y código generado de las definiciones de interfaces:

- El núcleo de Ice contiene el soporte en ejecución de la comunicación remota. Dicho núcleo está ligado a la aplicación a través de librerías con las que ésta enlaza.
- El código proxy es el código generado a partir de la definición de interfaces. Un proxy representa un objeto del servidor en la parte del cliente. Cuando el cliente invoca a una función del proxy, la llamada se traduce en un mensaje RPC al servidor, que permite ejecutar la función correspondiente. Además de ser el responsable del extremo cliente de la comunicación, el proxy contiene código para serializar y recuperar estructuras

complejas de datos transmitidas y recibidas durante la comunicación, proporcionando una abstracción mayor para el programador.

- El código esqueleto es código generado a partir de la definición de interfaces en la parte del servidor. Proporciona una interfaz de llamada que permite al núcleo de Ice transferir el hilo de control al código de aplicación asociado con los servicios.
- El adaptador de objeto es una parte de la API de Ice específica de los servidores. Se encarga de asociar peticiones entrantes de clientes a métodos específicos del servidor.

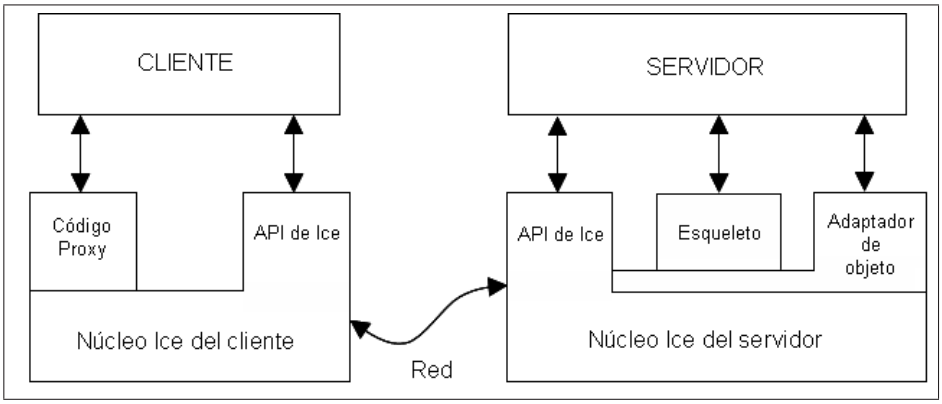



Figura 5.4: Estructura de clientes y servidores en Ice

Todo el soporte de comunicación en tiempo de ejecución se obtiene a partir de las librerías de Ice y del código generado a partir de definiciones Slice. La creación de clientes y servidores se reduce, por lo tanto, a escribir el código propio de cada aplicación y las definiciones de las interfaces que serán utilizadas por ambos para comunicarse. La definición de Slice proporcionará el esqueleto de una clase software, que contiene, a partir de sus métodos, los distintos servicios que podrán ser utilizados por los clientes de la aplicación. El código de la parte cliente deberá contener una instancia de dicha clase para poder invocar los métodos del servidor. El código del servidor reimplementará los métodos de la clase, añadiéndole las funcionalidades propias de cada servicio, y creará una instancia de ésta, que constituirá el extremo final de las llamadas RPC.

La figura 5.5 muestra la estructura genérica de un componente software del sistema propuesto. Se distinguen tres partes principales: la interfaz con los clientes, el manejador del componente y las interfaces con los servidores. El manejador implementa la funcionalidad

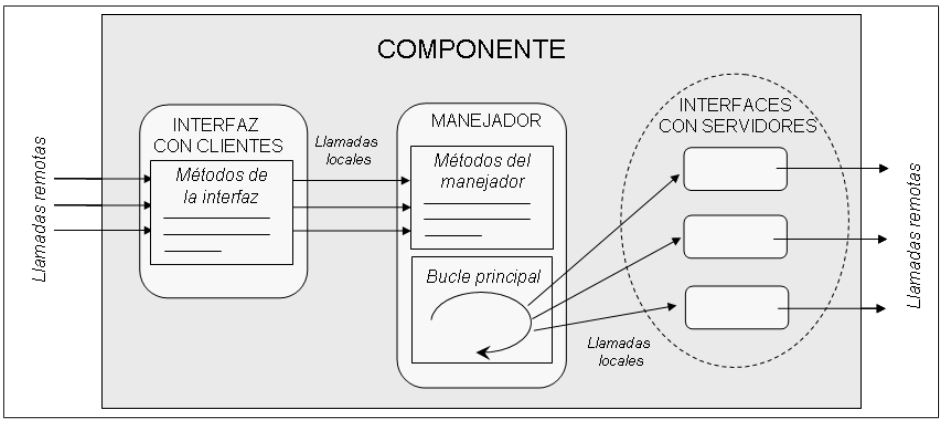



Figura 5.5: Estructura de un componente software del sistema

propia del componente. Mantiene un hilo de ejecución que realiza el bucle de procesamiento correspondiente y atiende posibles peticiones de otros componentes a través de la interfaz con los clientes. Esta última es una clase derivada de la clase generada a partir de la definición Slice. Contiene un método por cada servicio definido en la interfaz del componente e incluye una referencia al manejador del componente. Los métodos derivados de la definición Slice están reimplementados para responder a las peticiones de los clientes mediante llamadas locales a métodos del manejador. La última parte de la estructura de un componente está formada por las distintas interfaces con aquellos componentes que proporcionan servicios necesarios para el proceso realizado por el manejador. Cada interfaz no es más que una clase generada a partir de la definición Slice propia de cada componente servidor. El manejador contiene una instancia de cada interfaz de servidor de manera que, cuando tiene necesidad de utilizar alguno de los servicios proporcionados por éstos, realiza una llamada local al método correspondiente de la interfaz asociada y ésta se transforma en un mensaje RPC que es enviado al componente en cuestión.

### 5.2.4. Componentes de bajo nivel del sistema

Cada uno de los elementos que intervienen en el sistema propuesto ha sido implementado mediante un componente software siguiendo la metodología descrita anteriormente. La arquitectura global se encuentra apoyada sobre tres componentes de bajo nivel que se encargan de la comunicación con el hardware de captura y de control motor, pero que también permiten conjuntamente sincronizar a los restantes componentes del sistema. Es-

tos tres componentes, denominados *cameraComp*, *headComp* y *baseComp*, mantienen la siguiente funcionalidad:

- **cameraComp**: realiza la captura de imágenes a la frecuencia soportada por el hardware y atiende las peticiones de otros componentes devolviendo las imágenes más recientes obtenidas de dicha captura.
- **headComp**: se encarga de la comunicación con los motores de la torreta estéreo. Mantiene una copia del estado de cada motor para responder de manera inmediata a una petición de lectura por parte de otro componente. Recibe además peticiones de movimiento que lleva a cabo mediante el envío de los comandos correspondientes al hardware de control.
- **baseComp**: controla el movimiento de la plataforma. Recibe comandos de movimiento por parte de otros componentes a través de distintos indicadores: velocidad independiente de cada rueda, velocidad de giro y de traslación, posición destino del movimiento. Mantiene además la odometría de la base, pudiendo responder en todo momento de manera precisa a peticiones de localización del robot.

Los diferentes ritmos de procesamiento de cada componente del sistema provocan que sus resultados puedan proceder de imágenes capturadas en distintos instantes de tiempo y, en definitiva, para distintos estados del robot. Esto da lugar a situaciones de inconsistencia cuando varios componentes hacen referencia a un mismo dato asociado con imágenes obtenidas en distintos instantes de captura; en particular, cuando dichos instantes se corresponden con cambios de posición de las cámaras y/o desplazamientos del robot. Para poder resolver estas situaciones es necesario imponer algún mecanismo de sincronización entre los componentes.

Una posibilidad a la hora de sincronizar los componentes del sistema es incluir, en cada uno de ellos, un atributo de tiempo que almacene el instante de captura asociado con los datos que están siendo procesados actualmente. A través de este atributo, cada componente puede validar, y desechar cuando sea necesario, las referencias a datos obtenidos a partir de distintas capturas. Aunque este método soluciona el problema de la inconsistencia, provoca una pérdida continua de posibles resultados, que se hace más patente cuando el número de componentes es alto. Para evitar este problema, se ha desarrollado una solución alternativa consistente en asociar a la imagen obtenida en cada captura el estado de los motores en

ese instante. Este estado es propagado a los componentes de alto nivel de manera que las referencias a los datos obtenidos por distintos componentes son consideradas como referencias relativas al estado de cada componente. Cuando exista un intercambio de datos entre componentes, éstos enviarán también información sobre su estado para que cada referencia relativa pueda transformarse en una referencia absoluta. Por ejemplo, la posición de un mismo píxel puede ser diferente en dos componentes distintos, pero, a partir del estado de los motores almacenado en ambos componentes, es posible realizar una transformación que proporcione una correspondencia entre las dos posiciones.

Para que este método de sincronización sea posible, los tres componentes de bajo nivel deben mantener una comunicación continua que mantenga la relación correcta entre la captura y el estado de los motores. Con este fin, en cada captura, el componente *cameraComp* realiza una petición de lectura de estado a los componentes *headComp* y *baseComp*. Por cada petición de imagen, *cameraComp* envía, junto con los datos reclamados, el estado actual de la torreta y de la base, lo que permite actualizar los atributos de estado del componente que realiza la petición. De la misma forma, por cada petición de datos a un componente, éste envía el estado almacenado, proporcionando la información necesaria para la correcta sincronización entre los elementos que componen el sistema.

# Capítulo 6

# Sistema de control basado en la atención

El diseño de una arquitectura de control para un robot móvil conlleva buscar soluciones a cuestiones como el análisis y la intepretación de la información sensorial proporcionada sobre el entorno, el alcance limitado de los sensores, la limitación del tiempo de procesamiento y el mantenimiento simultáneo de múltiples objetivos conductuales. Desde el punto de vista de la percepción, parece lógico pensar que sólo se considere un subconjunto de información que pueda ser relevante para la tarea en curso. Teorías como la "ceguera al cambio" (Simons y Levin, 1998) (O'Regan et al., 1999) demuestran que no vemos todo a nuestro alrededor, sino que sólo se utiliza una fracción de la información recibida que pueda influir directamente en el comportamiento. La mayoría de los sistemas biológicos de visión emplean una estrategia de serialización del flujo de información visual a través de atención abierta y encubierta en la exploración de una escena. Este aspecto de la atención dota a la percepción de una capacidad activa que permite orientar los sensores hacia zonas de interés del espacio perceptivo, dedicando los recursos a una parte de la información potencialmente visible y reduciendo de esta forma los costes de procesamiento. Algunos investigadores plantean además la necesidad de la atención como medio de selección para la acción. Desde este punto de vista, la atención permitiría centrar la acción en un objetivo concreto, obviando elementos de distracción para la tarea y guiando su ejecución.

Apoyado en las ideas anteriores, se propone un sistema de control basado en la atención. Se trata de una arquitectura de comportamientos que utiliza la atención como intermediario entre percepción y acción. En este sentido, la función de la atención es doble: orientar el proceso perceptivo en función de las acciones y modular las acciones de acuerdo con la percepción. El modelo de control será utilizado en un robot móvil dotado de sensores de

visión para generar en él distintas formas de interacción con su entorno.

## 6.1. Descripción general

La tesis fundamental de este trabajo se centra en dos aspectos de la atención visual:

- *La atención como mecanismo de serialización del flujo de información visual*: ésta es, sin duda, la propiedad menos cuestionada de la atención, mediante la cual es posible realizar un análisis selectivo de las zonas del campo visual que resultan de interés en una situación determinada.
- *La atención como elemento modulador de la acción*: este aspecto de la atención constituye una ampliación de su función selectora. Desde este punto de vista, la atención sobre un estímulo podría provocar la ejecución de una acción coherente con dicho estímulo y con la situación actual. En este sentido, la atención actuaría como una vía de serialización de la acción producida por el disparo ordenado de las acciones asociadas con la secuencia de estímulos visuales proporcionada por la fijación atencional.

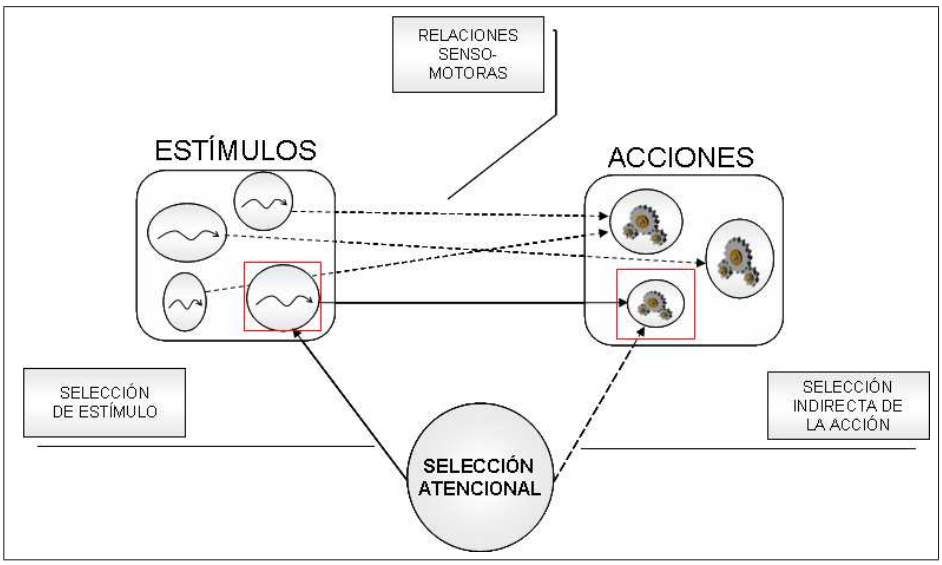



Figura 6.1: Selección de la acción a través de la selección atencional

Estas dos propiedades hacen referencia a la función selectora de la atención desde dos puntos de vista distintos. Desde el punto de vista de la percepción, la atención actúa como

un mecanismo de serialización que permite acceder en secuencia a zonas de interés, obviando aquellas que resultan elementos de distracción en una situación determinada. Desde la perspectiva de las acciones, la atención sobre un estímulo permite realizar una toma de decisión sobre qué acción llevar a cabo dentro de un conjunto de posibles acciones. Este último enfoque dota a la atención de una capacidad generadora de actividad dirigida a mantener una relación dinámica entre el agente y su entorno. Partiendo de una serie de relaciones contextualizadas entre estímulos visuales y acciones motoras, la selección atencional de un estímulo activaría un proceso de actuación en el robot, dando lugar a una forma de selección indirecta de la acción (figura 6.1).

El estudio de estas propiedades y su influencia en la dinámica de interacción entre un robot y su entorno constituyen el principal objetivo de este trabajo. Con este fin, se presenta un modelo de control que incluye la atención como centro neurálgico del sistema.

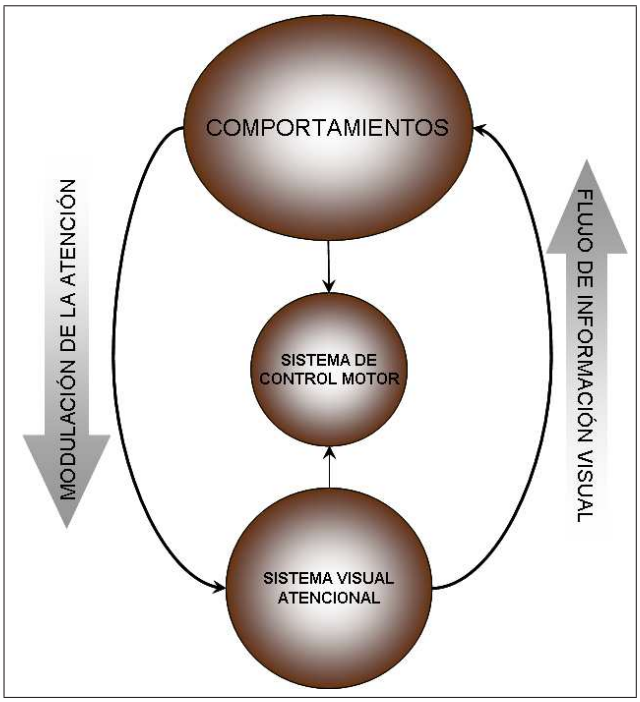



Figura 6.2: Modelo de control basado en la atención

La figura 6.2 muestra un esquema general de este sistema de control basado en la atención. Las capacidades senso-motoras del sistema se dividen en dos grupos que, a su vez, forman dos subsistemas: el sistema visual atencional, en el que se modelan los mecanismos

de fijación que dan lugar a la selección de información visual, y el conjunto de comportamientos de alto nivel que utilizan información visual para mantener sus objetivos dentro del sistema. Ambos sistemas están conectados al sistema de control motor encargado de hacer efectiva la respuesta motora generada por los dos subsistemas anteriores.

Cada comportamiento de alto nivel modula el sistema visual de manera específica para que éste responda con la secuencia de información visual más adecuada para el cumplimiento de sus objetivos. El flujo de información recibido afecta al comportamiento influyendo en la ejecución de sus acciones, por lo que podemos afirmar que el sistema visual también modula en cierta manera al subsistema de comportamientos. Esto se traduce dentro del sistema global en dos lazos de control que actúan en paralelo, perturbándose mutuamente y cooperando para lograr un objetivo común. Uno de ellos, el asociado con el sistema visual, controla los movimientos oculares que proporcionan la fijación de un objetivo visual. El otro, lleva a cabo el control senso-motor propio de cada comportamiento.

La filosofía atencional del sistema favorece la ejecución simultánea de varios comportamientos. La selección de la acción en este caso se resuelve mediante la selección de un objetivo visual, consiguiendo serializar las acciones de los distintos comportamientos activos. En definitiva, la cuestión de la selección motora se transforma en una selección sensorial, asegurando que únicamente se ejecuten las acciones compatibles con el foco de atención. Las acciones pueden ser contradictorias lo que, en determinadas circunstancias, complica excesivamente el proceso de selección. Sin embargo, aunque los estímulos compiten por tomar el control de la atención, nunca entran en conflicto con los restantes, puesto que son sus propiedades las que les convierten en ganadores de esa selección. Siempre habrá algún estímulo cuyas propiedades sean más adecuadas para la situación actual que las del resto, por lo que la selección siempre tendrá lugar, proporcionando la información sensorial necesaria para ejecutar las acciones más apropiadas en dicha situación.

En resumen, las principales propiedades del modelo propuesto podrían sintetizarse en los siguientes puntos:

- Dos niveles senso-motores: las capacidades del sistema se encuentran separadas en dos subsistemas senso-motores. Por un lado, el sistema atencional, encargado de seleccionar información sensorial, y, por otro lado, el grupo de comportamientos de alto nivel que utilizan dicha información para llevar a cabo un control específico.

- Control distribuido: el control global se distribuye entre los dos subsistemas anteriores de manera que existen en todo momento dos bucles de control en ejecución actuando en paralelo y cooperando para cumplir un objetivo determinado.
- Modulación descendente: el sistema atencional es modulado por los distintos comportamientos de alto nivel para que la selección del objetivo visual se lleve a cabo de acuerdo con propiedades específicas coherentes con la situación actual y con las necesidades conductuales.
- Modulación ascendente: el sistema visual responde en todo momento enviando información sobre el foco de atención actual. Las propiedades de la región atendida afectan a los comportamientos de alto nivel de manera que las acciones que finalmente tienen lugar son únicamente aquellas compatibles con dicho foco de atención.

## 6.2. El sistema de atención visual

El sistema visual propuesto es modelado como una colección de componentes que colaboran para fijar un objetivo visual y elegir el siguiente. Cada objetivo visual es una región de interés de la imagen que se corresponde con una zona separable del resto por sus propiedades visuales. Tal y como se muestra en la figura 6.3, tras las detección de regiones de interés, el flujo de procesamiento se separa en dos ramas que convergen en la parte superior del grafo. Las dos ramas dividen la función visual en una analogía del "qué" y el "cómo" propuesto en la neurociencia. Esta división permite una especialización de las diferentes funciones, dedicando recursos específicos a cada rama y compartiendo información común de los procesos de bajo nivel. La parte del proceso dedicada al "qué" se encarga de obtener propiedades de aspecto de las regiones de interés relacionadas con color, textura, etc. La rama dedicada al "cómo" extrae características espaciales sobre las distintas regiones de interés, tales como posición, movimiento u orientación. La información procedente de ambas ramas es integrada a través de un conjunto de componentes de alto nivel, a los que denominaremos selectores de objetivo, y que son los encargados de guiar el proceso de atención del sistema.

Cada selector de objetivo se encarga de construir un mapa de control de atención que le permita seleccionar una zona a la que dirigir la mirada. Esta selección se lleva a cabo a partir de la información procedente de las dos ramas de procesamiento visual, de acuer-

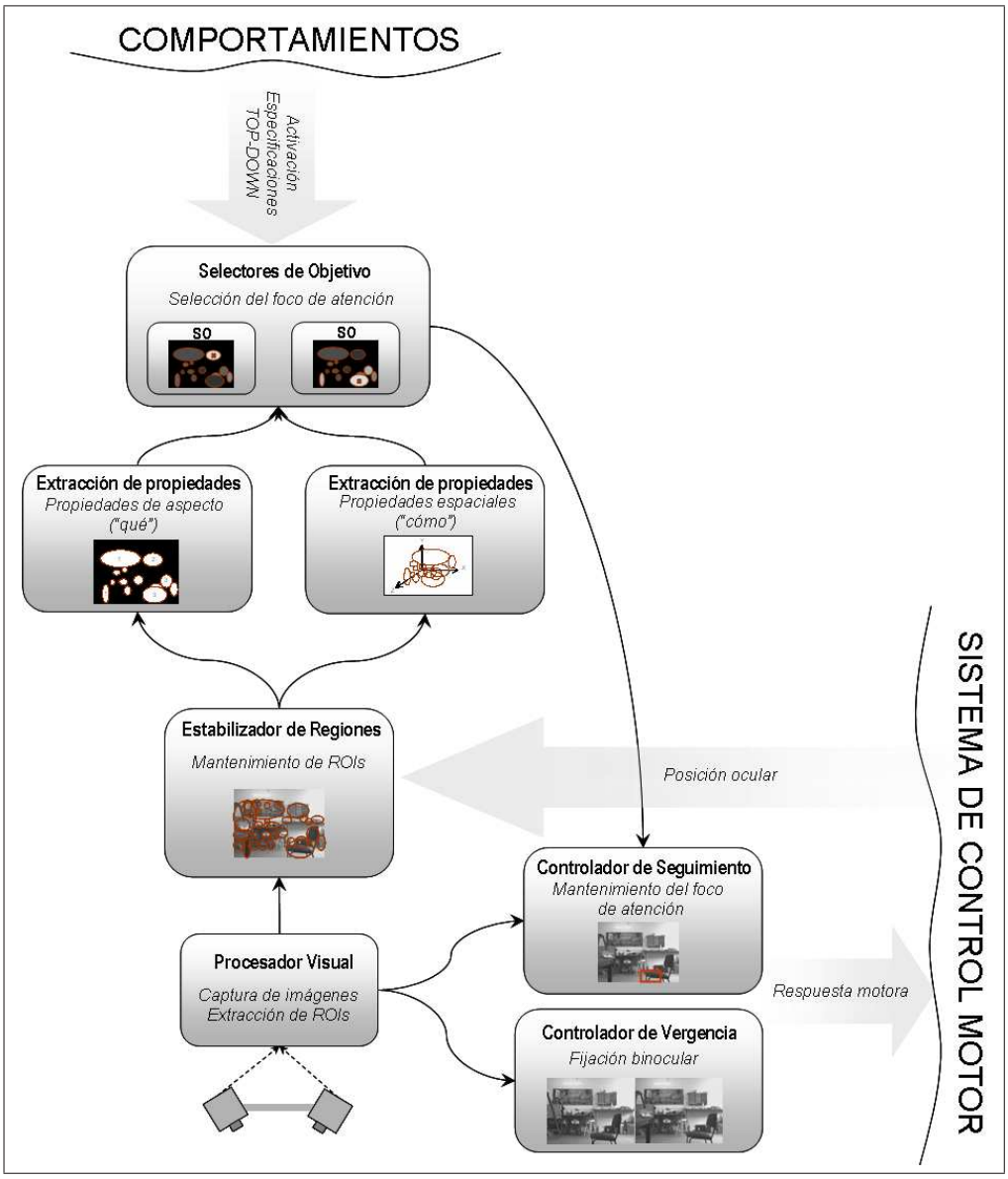



Figura 6.3: Arquitectura del sistema de atención visual

do a determinados criterios de selección. Cada componente de este tipo se encuentra en comunicación con un comportamiento de alto nivel que se encarga de activarlo y modularlo para que mantenga la atención sobre las zonas de la escena de mayor relevancia en base a sus objetivos conductuales. En un instante determinado, el control atencional puede estar distribuido entre varios selectores, cada uno dirigido a un tipo de objetivo concreto. La atención abierta en este caso sólo es controlada por uno de los selectores, mientras que los restantes mantienen atención encubierta a la espera de hacer efectiva su acción oculomotora. La frecuencia de adquisición del control abierto de la atención en cada selector es modulada por los comportamientos de alto nivel de acuerdo con sus necesidades de información.

La selección de un foco de atención se transforma en una respuesta motora cuyo objetivo es mantener una fijación binocular de la región seleccionada. Para ello, la fijación 3D se descompone en dos movimientos, uno sacádico y de seguimiento en una de las cámaras, y otro asimétrico de vergencia en la otra cámara. Esta separación permite que la selección del foco de atención se haga efectiva a través de un sacádico programado para una única cámara, a la vez que se mantiene una fijación estable en ambas cámaras.

En definitiva, en este esquema, cada componente está dedicado a tareas específicas de procesamiento que conjuntamente proporcionan el control de atención. El proceso global constituye un ciclo percepción-acción que comienza por la captura de imágenes y la extracción de regiones de interés y culmina en una respuesta motora a través de la selección de un foco de atención, proporcionando una relación dinámica entre los sistemas sensoriales y motores. En el resto de este apartado, se detallan las distintas fases de procesamiento incluidas dentro de este ciclo.

### 6.2.1. Extracción de regiones de interés

En esta primera fase, el sistema realiza una preselección de las zonas de imagen que tienen entidad suficiente para constituir un foco de atención. Serán regiones de la imagen que destaquen significativamente del resto y en las que se observe un alto contenido de información. Esta labor es realizada por el procesador visual tras la captura de imágenes. El principal criterio utilizado en esta primera tarea de selección de información es la estabilidad ante diferentes condiciones visuales.

Durante los últimos años, se han propuesto varios métodos de detección de puntos de interés que muestran un funcionamiento estable ante transformaciones significativas de una escena visual. En tareas de búsqueda de correspondencias entre imágenes, destaca el detector de esquinas de Moravec (Moravec, 1981), posteriormente mejorado por Harris y Stephens (Harris y Stephens, 1988). Las aplicaciones iniciales de este último fueron la correspondencia estéreo y el seguimiento del movimiento, pero más tarde se extendió a problemas más complejos (Zhang et al., 1995) (Torr, 1995). El detector de Harris es invariante a rotaciones y proporciona resultados geométricamente estables, demostrando un alto rendimiento en tareas de correspondencia y seguimiento.

En cuanto a la detección de regiones, se han ideado métodos que utilizan estructuras multi-escala para extraer zonas de interés de la imagen ante diferentes condiciones como cambios de escala o transformaciones afines. La idea de búsqueda en el espacio de escala fue introducida por Crowley (Crowley, 1981) y Crowley y Parker (Crowley y Parker, 1984). En esta propuesta, se construye una representación piramidal utilizando diferencias de Gaussianas y se detectan puntos característicos como máximos locales de toda la estructura en los que el valor absoluto supera un cierto umbral. En 1998, Linderberg (Lindeberg, 1998) propone una variante del método anterior utilizando el operador Laplaciano de gaussiana (LoG) y otros operadores basados en derivadas. En este caso, el espacio de escala se construye a partir de suavizados sucesivos sobre la imagen de alta resolución utilizando varios filtros Gaussianos de diferentes tamaños. Lowe, en 1999, (Lowe, 1999) propuso un algoritmo de detección de objetos basado en máximos locales dentro del espacio de escala piramidal construido mediante filtros de tipo diferencia de Gaussianas (DoG). Esta estructura se construye muestreando sucesivamente la imagen de alta resolución previamente suavizada con filtros Gaussianos y calculando la diferencia entre dos imágenes sucesivas. Los máximos locales en esta representación piramidal determinan la posición y la escala de los puntos de interés. Este método permite acelerar el tiempo de procesamiento mediante una aproximación de la función LoG al operador DoG. No obstante, la base de ambos operadores es la detección de máximos en el entorno de contornos y bordes, en los que la señal cambia en una única dirección. Estos máximos no son muy estables porque su posición es más sensible al ruido o a cambios de textura en sus vecinos. Para resolver este problema se han propuesto métodos más restrictivos en la selección de máximos locales dentro del espacio de escala. Mikolajczyk propone un método de selección en el que los puntos de interés deben cumplir simultáneamente la propiedad de máximos locales en la traza y el

determinante de la matriz Hessiana (Mikolajczyk, 2002). Esto permite penalizar aquellos puntos en los que el cambio de la señal se produce en una única dirección. Esta idea es similar a la utilizada en el detector de Harris, aunque este último proporciona resultados más fiables ante variaciones como rotaciones, cambios de iluminación o transformaciones de perspectiva. En 2001, Mikolajczyk y Schmid proponen un método de detección de puntos de interés invariante a la escala aplicando el detector de Harris (Mikolajczyk y Schmid, 2001). El método, conocido como Harris-Laplace, consiste en calcular puntos de interés en los distintos niveles de escala aplicando el detector de Harris y seleccionar sólo aquellos que maximizan el valor del Laplaciano en todos los niveles. Esto permite obtener los puntos más distintivos de la imagen, así como su escala característica, con independencia de la escala original, rotaciones, traslaciones y cambios de iluminación. Esta idea fue adaptada posteriormente (Mikolajczyk y Schmid, 2004) para extender su propiedad de invariabilidad a transformaciones afines, mediante un método iterativo que modifica la posición, escala y vecindad de cada punto proporcionando un conjunto de puntos estables ante diferentes condiciones visuales.

#### 6.2.1.1. Propuesta concreta: Harris-Laplace sobre el prisma multi-escala

Los métodos anteriores surgen mayoritariamente como mecanismos de segmentación aplicados a la detección de objetos. Sin embargo, nuestro interés no es tanto la segmentación de una escena como la localización dentro de la misma de información relevante. Se trata de encontrar zonas de la imagen en las que sea posible fijar y mantener un foco estable de atención ante situaciones cambiantes. Con este objetivo, y buscando un compromiso entre la validez de los resultados y el tiempo de procesamiento, se ha desarrollado un método, basado en Harris-Laplace, que realiza una distinción entre fóvea y periferia en la selección de escala de las regiones de interés. Para ello, se aplica el método original a una estructura multi-escala de tipo prisma (figura 6.4), correspondiente a una sección centrada del espacio de escala. Esta estructura presenta ciertas similitudes con una superficie retiniana en la que la máxima agudeza visual se concentra en la parte central y disminuye radialmente en función de la excentricidad (Bandera y Scott, 1989).

El método desarrollado tiene como objetivo extraer información detallada en la fóvea y más general en la periferia. La aplicación de Harris-Laplace sobre el prisma multi-escala permite que sólo las características de la zona de fóvea estén presentes en todos los nive-

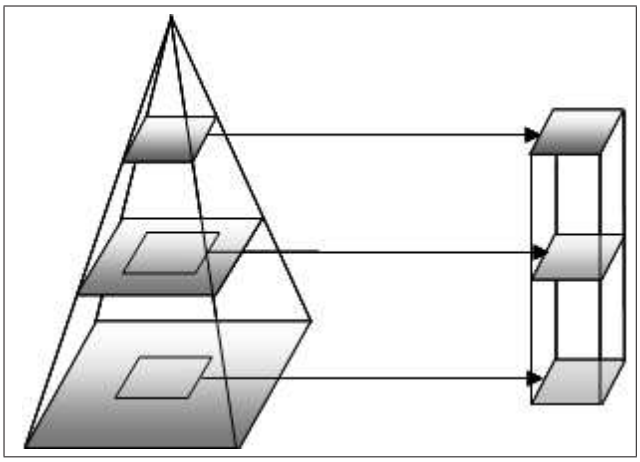


Figura 6.4: Estructura multi-escala de tipo prisma

les de la estructura y que las de periferia se sitúen únicamente en los niveles superiores, decreciendo el número de niveles en los que la zona en cuestión está presente, a medida que aumenta su excentricidad. Esto implica que, en la zona de fóvea, sea posible detectar cualquier región con independencia de su tamaño, mientras que, en la de periferia, sólo sea posible localizar regiones de una cierta extensión, obviando posibles detalles de su estructura. La figura 6.5 muestra esta idea.

Cada nivel $n$ del prisma está asociado con una ventana central de la imagen de máxima resolución de tamaño $W_P/s^n$ x $H_P/s^n$, siendo $W_P$ y $H_P$ el ancho y alto de las imágenes del prisma y $s$ el factor de escala ($s < 1$). Asimismo, tomando como coordenadas $(0, 0)$ el centro de imagen, cada punto de interés $(x_i, y_i)$ extraído en un nivel $n$ se corresponde con un región de la imagen original centrada en la posición $(x, y) = (x_i/s^n, y_i/s^n)$. Puesto que el proceso de selección de escala de cada punto de interés consiste en un recorrido vertical en el prisma multi-escala, dicho recorrido sólo podrá realizarse en aquellos niveles en los que el punto en cuestión esté presente. Teniendo en cuenta la transformación de escala, un punto con coordenadas $(x, y)$ en la imagen de máxima resolución pertenecerá a cada nivel $n$ del prisma para el que se verifiquen las dos expresiones siguientes:

$$\frac{-W_P}{2} \leq x * s^n \leq \frac{W_P}{2} \tag{6.1}$$

$$\frac{-H_P}{2} \leq y * s^n \leq \frac{H_P}{2} \tag{6.2}$$

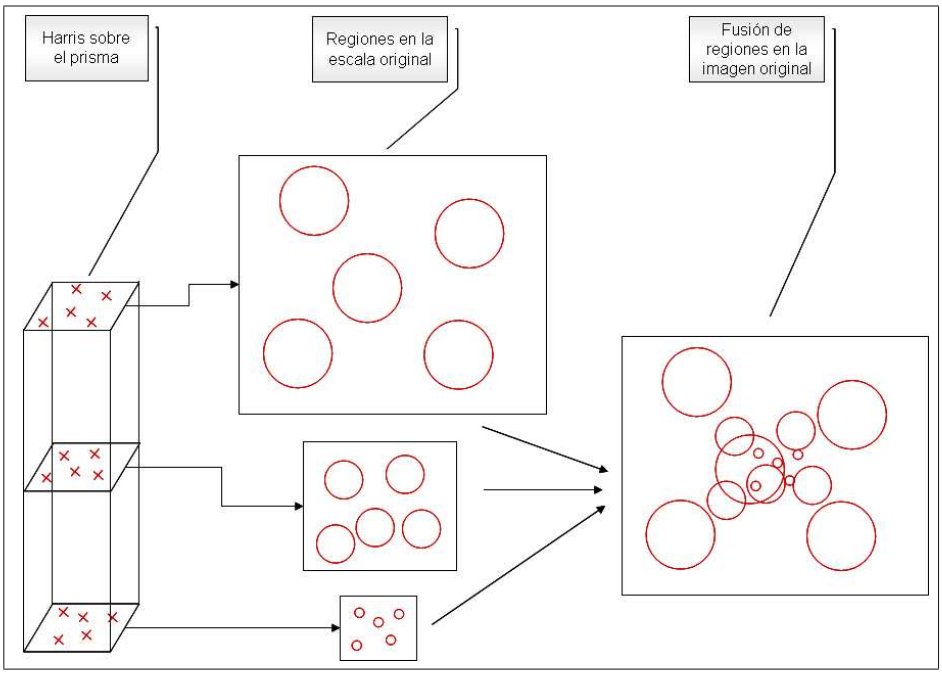
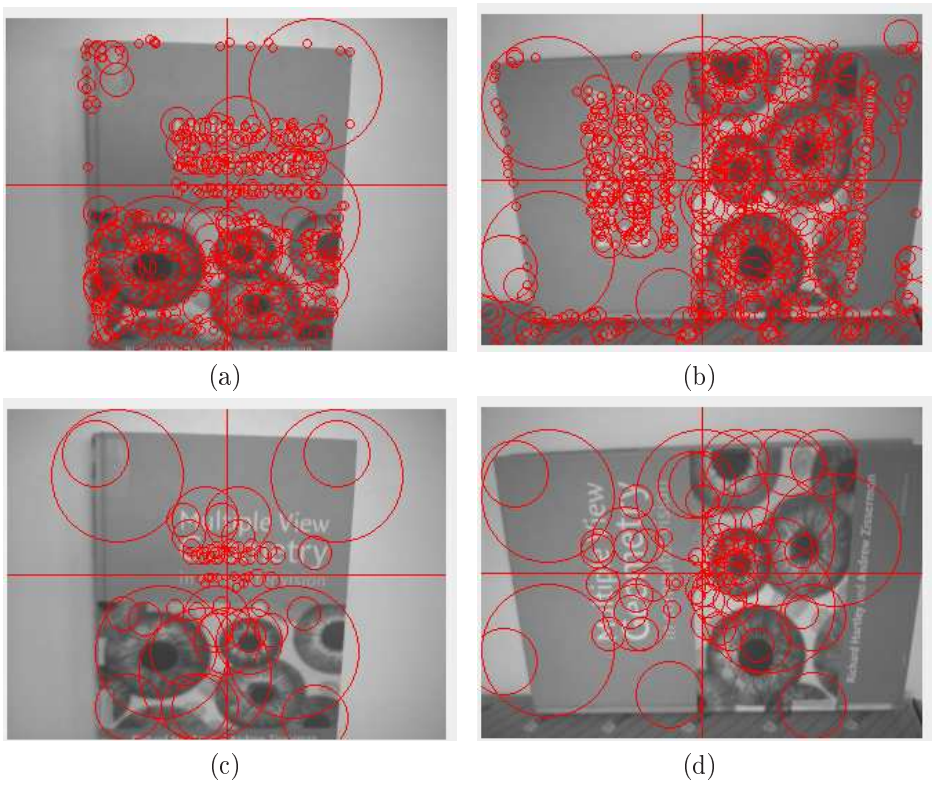



Figura 6.5: Harris-Laplace sobre el prisma multi-escala

Las expresiones anteriores obligan a que cuanto mayor sea la distancia del punto de interés al centro de imagen, mayor sea también el mínimo nivel al que dicho punto puede pertenecer, consiguiendo que la búsqueda sea menos costosa a medida que nos alejemos de la zona de fóvea. Además, este proceso de búsqueda permite limitar las posibles extensiones de cada región en función de su posición en la imagen. Concretamente, la selección de escala asigna a cada región detectada una determinada extensión que viene dada por las dimensiones consideradas para las regiones del nivel inferior ponderadas por el factor de escala correspondiente $(1/s^n)$. Así, cuanto mayor sea el nivel seleccionado para un determinado punto, mayor será también la extensión de la región asociada en la imagen original. Puesto que la excentricidad de un punto determina su mínimo nivel de pertenencia, también límita sus dimensiones mínimas. Esta limitación forzará a que las regiones situadas en la periferia deban tener una mayor extensión que las de fóvea para poder ser detectadas, logrando así un efecto de detalle en la parte central de la imagen y de información más general en el resto.

Las imágenes de la figura 6.6 muestran el resultado de aplicar el método Harris-Laplace original ((a) y (b)) y la variante propuesta ((c) y (d)) a una misma imagen. Como se observa en la figura, la aplicación sobre la pirámide detecta los distintos detalles de la imagen

con independencia de su posición en la misma. Sin embargo, la detección de regiones a través del prisma incluye únicamente las regiones de mayor extensión de la periferia mientras que mantiene el mismo nivel de detalle que el método original en la parte central de la imagen. Esto proporciona una reducción significativa de los tiempos de procesamiento sin que se produzca una pérdida completa de información visual de la escena. Dicha pérdida sólo tiene lugar a ciertos niveles de detalle sobre las zonas de periferia, lo que puede subsanarse mediante movimientos de cámara que permitan focalizar las diferentes zonas.



(a) (b) (c) (d)

Figura 6.6: Resultados de Harris-Laplace sobre la pirámide ((a) y (b)) y el prisma multiescala ((c) y (d))

### 6.2.2. Mantenimiento de regiones

En numerosos trabajos se ha mostrado que es posible construir robots móviles muy simples capaces de desplazarse de forma robusta en entornos naturales sin colisionar. Estos logros se han debido, fundamentalmente, a un cambio conceptual en la forma de diseñar las arquitecturas de control inteligente (Brooks, 1991b), en la que se cuestiona la necesidad

de una representación explícita del espacio y sugieren una interpretación directa más en la línea de los *affordances* de Gibson (Gibson, 1979). Sin embargo, en los enfoques estrictamente reactivos aparecen rápidamente limitaciones cuando se pretende ampliar las capacidades de navegación e interacción del robot. Uno de los problemas clásicos es la permanencia de la percepción. Las regiones del mundo no nos aparecen como nuevas cada vez que cambiamos el punto de vista y nuestra percepción del entorno es más rápida cuando ya ha sido visto con anterioridad (Bridgeman et al., 1994). Además, a pesar de que se puedan establecer multitud de interacciones útiles entre el entorno y el robot sin hacer uso de ninguna forma de representación, uno de los aspectos donde más consenso hay sobre la inteligencia es la capacidad predictiva y, en general, esta capacidad lleva asociada algún tipo de memoria que permita mantener internamente ciertas propiedades del mundo externo.

Siguiendo esta línea, las regiones detectadas en la fase anterior son integradas en un modelo interno que hace la función de memoria a corto plazo. La estructura utilizada para mantener esta representación es un mapa construido sobre el espacio de movimientos de la cámara encargada de dirigir la atención (figura 6.7). Cada región es indexada por su posición angular, lo que permite mantener información sobre regiones situadas fuera del campo de visión y, de esta forma, volver a fijar la atención rápidamente sobre zonas ya vistas.

En cada ciclo de procesamiento, las nuevas regiones son incluidas en la estructura de memoria y las almacenadas previamente son actualizadas en base a la información obtenida por el procesador visual. Esta actualización consiste en la relocalización de las distintas regiones en la nueva escena. Para ello, se realiza una búsqueda de cada región en un entorno próximo de la posición correspondiente dentro del espacio de escala. Esta búsqueda tridimensional permite localizar la región aunque su extensión cambie, por ejemplo, por un movimiento de translación del robot.

Dentro de este proceso de actualización, cada región modifica el valor de un atributo asociado de permanencia que aumenta o disminuye dependiendo de si la operación de búsqueda tuvo éxito o no. Aquellas regiones que no superan un determinado valor de permanencia se consideran "olvidadas" y son eliminadas, por tanto, de la estructura de memoria. Las restantes actualizan otros atributos utilizados en fases posteriores del procesamiento, destacando los siguientes:

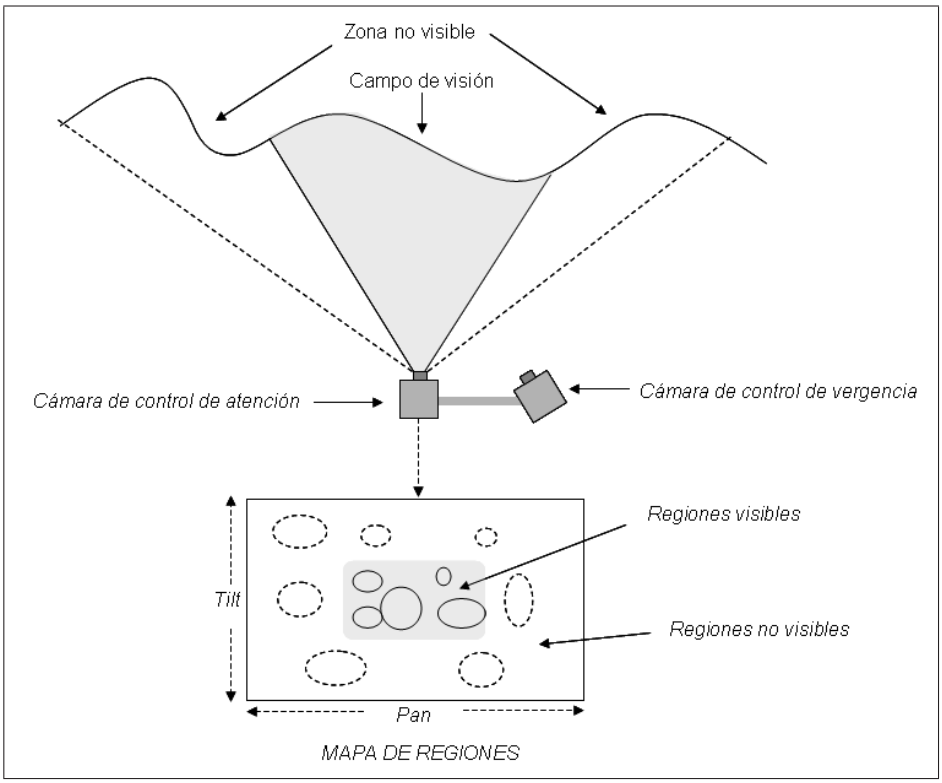



Figura 6.7: Mapa de regiones

- Escala actual y cambio de escala: como se ha comentado anteriormente, en su actualización, una región puede cambiar su escala original puesto que el proceso de búsqueda es realizado en toda la estructura multi-resolución. El cambio de escala proporcionará información sobre movimiento, ya sea propio o externo, y el valor de escala actual permitirá conocer la extensión de la zona de imagen asociada con la región.

- Posición angular: el acceso y fijación de una región se lleva a cabo a través de la posición de dicha región en el espacio de movimientos de la cámara, por lo que los atributos posicionales que tienen mayor interés son los que indican posición angular (*pan* y *tilt* de cámara). Esto permitirá que la actualización de cada región pueda realizarse en entornos próximos con independencia de los movimientos en la cámara, además de mantener en la estructura de memoria regiones que no estén actualmente en el campo de visión.

- Ventanas de imagen: se trata de las zonas de imagen asociadas con la región en

la escala actual y en la resolución original. Se utilizarán en fases posteriores del procesamiento para extraer propiedades de la región que sean independientes de la escala, en un caso, o en las que deba tenerse en cuenta una porción de mayor resolución de la imagen.

- Velocidad 2D: al igual que el cambio de escala, el cambio de posición proporciona información de movimiento, que será de utilidad para extraer ciertas propiedades de la región, así como para incluir mecanismos predictivos en la actualización de posición de una región.

- Tiempo de atención: cuantifica la proximidad de la región a las zonas seleccionadas recientemente como focos de atención. La actualización de este valor se lleva a cabo a través de una función gaussiana desplazada que permite amplificar el valor de atención de las zonas cercanas a la fóvea y disminuir el de las regiones situadas en la periferia. Dentro de una estrategia de control de atención en la que sea necesario dedicar tiempos limitados a la fijación de cada objetivo para poder visitar zonas no atendidas, este atributo resulta fundamental.

Los atributos obtenidos durante esta fase forman un primer grupo de propiedades que serán ampliadas en etapas posteriores del sistema. Para facilitar la existencia de diferentes "caminos" de procesamiento que realicen una extracción de propiedades de manera separada e independiente es necesario proporcionar un mecanismo de integración que asigne los distintos subconjuntos de atributos obtenidos a la misma región de procedencia. Con este fin, cada región mantiene un identificador único que es asignado a las nuevas regiones y conservado para las ya existentes durante esta fase del sistema. Este identificador actúa como referencia inequívoca de cada región, proporcionando un medio de reunificación de subconjuntos de propiedades obtenidos a partir de diferentes vías.

### 6.2.3. Extracción de propiedades de alto nivel

La evolución ha dotado a los primates de un conjunto de áreas visuales que ocupa sobre el 50 % del córtex visual posterior (Zeki, 1993). En investigaciones con monos, se han identificado dos flujos de proyección en el cerebro originados en el área visual primaria (V1): un flujo ventral proyectado hacia el córtex temporal inferior (IT) y un flujo dorsal proyectado hacia el córtex parietal posterior (PP) (figura 6.8). Ungerleider y Mishkin (Ungerleider y Mishkin, 1982) propusieron que estos dos flujos de procesamiento visual desempeñan diferentes papeles en la percepción de la información visual entrante. Según esa propuesta,

el flujo ventral juega un papel crítico en la identificación y el reconocimiento de objetos, mientras que el flujo dorsal media en la localización de los mismos. A estas dos vías de procesamiento visual se las conoce como el "qué" y el "dónde". Esta idea está basada y es coherente con el estudio de monos con diferentes lesiones cerebrales. Sin embargo, los resultados obtenidos a partir de estudios más recientes (Milner y Goodale, 1995) llevan a una distinción no entre subdominios de percepción, sino entre la percepción, referida al reconocimiento, por un lado y la guía de la acción por otro. A estos dos subsistemas de procesamiento se los conoce como el "qué" (*the what system*) y el "cómo" (*the how system*). Los argumentos señalados para esta nueva distinción son varios. La mayor fuente de evidencia proviene del estudio de las propiedades visuales de las neuronas de los flujos ventral y dorsal:

- Las neuronas del flujo ventral se activan con ciertas características de los objetos y muchas muestran una especificación categorizadora notable. Se ven poco afectadas por el comportamiento motor.
- Las neuronas del flujo dorsal muestran diferentes propiedades. Diferentes subconjuntos de neuronas en el córtex PP se activan por estímulos visuales como resultado de las diferentes clases de respuestas que se realizan sobre esos estímulos.

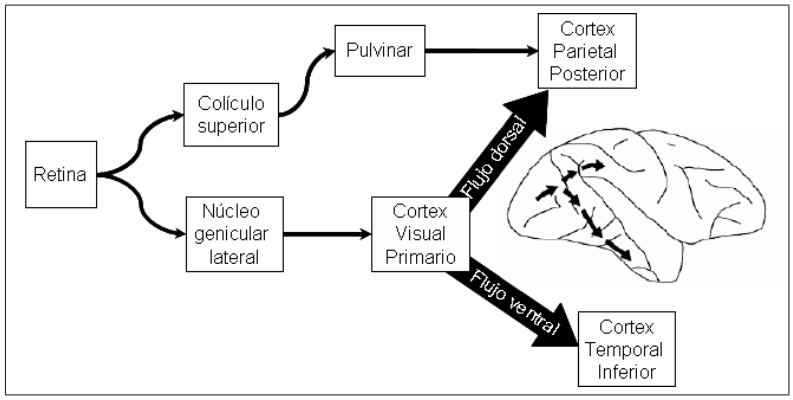



Figura 6.8: Flujos de la entrada visual

Aunque aún quedan muchas cuestiones por resolver sobre la interacción de los dos flujos de procesamiento, los resultados obtenidos a partir de experimentos realizados con animales y el efecto de lesiones cerebrales en estas áreas apuntan a la validez de esta última

distinción funcional.

Inspirado en estos resultados, el flujo de procesamiento del sistema visual propuesto se divide en esta fase en dos subsistemas dedicados a extraer propiedades de las regiones relacionadas con el "qué" y el "cómo".

#### 6.2.3.1. Extracción de propiedades de aspecto: el *qué*

El primer subsistema de procesamiento se encarga de extraer propiedades de las regiones que permitan identificarlas por su apariencia visual. Para que este proceso de identificación se lleve a cabo de manera eficiente en diferentes situaciones, las propiedades extraídas deben mantener cierta capacidad de invariabilidad ante transformaciones del punto de vista, la iluminación de la escena, etc.

En los últimos años, se han propuesto una serie de descriptores de regiones de imagen que presentan invariabilidad ante ciertas deformaciones como cambios de escala, rotaciones y variaciones de iluminación. La construcción de estos descriptores se basa en la subdivisión de la región de soporte en anillos concéntricos y la obtención de un histograma de atributos de apariencia para cada uno de ellos. La representación de la imagen a través de histogramas proporciona estabilidad ante diversas deformaciones. Los resultados experimentales obtenidos a partir de estudios comparativos (Mikolajczyk y Schmid, 2005) muestran una mayor tasa de reconocimiento para este tipo de descriptores que para otros descriptores tradicionales basados en el cálculo de funciones de la región de soporte al completo (bancos de filtros, invariantes diferenciales).

Dentro de los descriptores basados en histogramas, destacan especialmente las representaciones SIFT (Lowe, 2004), RIFT y Spin (Lazebnik et al., 2005). El sistema propuesto utiliza los dos últimos para extraer propiedades de apariencia de las regiones del entorno. Ambos son descriptores invariables a la escala y a rotaciones y extraen, respectivamente, información de gradientes e intensidades en la imagen.

El descriptor RIFT es una generalización del descriptor SIFT, propuesto por Lowe. Construye, por cada anillo de subdivisión de la región, un histograma que representa la distribución de orientaciones de gradientes de esa zona de imagen. Para conseguir la invariabilidad a rotación, cada orientación es calculada de manera relativa a la dirección de la

recta que pasa por el punto y el centro de la región. Es decir, el intervalo del histograma al que contribuye un determinado punto se calcula a partir del ángulo que forman el gradiente en ese punto y el vector de dirección del rayo que sale del centro y pasa por él.

La estructura de un descriptor RIFT puede tratarse como un histograma bidimensional indexado por intervalos de la distancia al centro y de la orientación relativa del gradiente. La contribución de cada punto a la entrada del histograma que corresponda se obtiene del módulo del gradiente en dicho punto. Así, cada punto $x$ de la región con distancia al centro $d$ y gradiente $\vec{g}$, contribuye al intervalo del histograma indexado por $(d_i, \theta_j)$ en $|\vec{g}|$, donde $d_i$ y $\theta_j$ cumplen:

$$d_i \leq d < d_{i+1} \tag{6.3}$$

$$\theta_j \leq \theta < \theta_{j+1} \tag{6.4}$$

Sea $x_0$ el punto central de la región y $\vec{d}$ el vector $\vec{x_0x}$, el ángulo $\theta$ de la expresión anterior se obtendría como la diferencia entre $\theta(\vec{g})$ y $\theta(\vec{d})$, ángulos absolutos de los vectores $\vec{g}$ y $\vec{d}$, respectivamente.

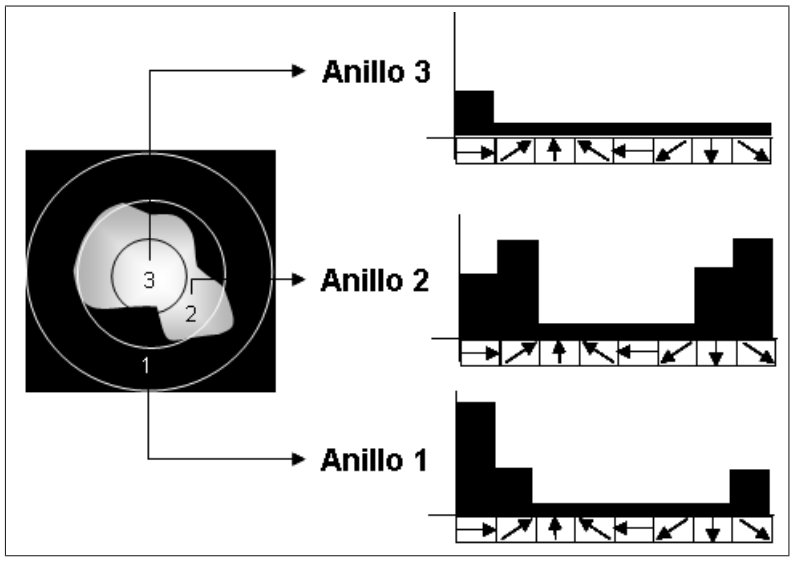



Figura 6.9: Descriptor RIFT de 3 anillos

La figura 6.9 muestra el descriptor RIFT de una imagen dividida en 3 anillos. La orientación está representada por 8 intervalos distintos, por lo que el descriptor final tiene un

total de 24 entradas. La distribución de gradientes que se observa en cada histograma se obtiene del contenido de cada anillo de manera independiente.

El segundo descriptor utilizado en el sistema propuesto es el descriptor Spin. Está basado en las imágenes Spin utilizadas en la representación de superficies (Johnson y Hebert, 1999). El resultado de calcular este descriptor es un histograma bidimensional que codifica la distribución de intensidades en la región de imagen correspondiente. Las dos dimensiones del histograma son la distancia al centro de la región y el valor de intensidad. Para obtener una representación suave, cada píxel contribuye a más de un intervalo. En concreto, la contribución de un píxel situado en una posición $x$ al intervalo indexado por $(d, i)$ viene dada por:

$$exp\left(-\frac{(|x - x_0| - d)^2}{2\alpha^2} - \frac{(I(x) - i)^2}{2\beta^2}\right) \tag{6.5}$$

siendo $x_0$ la posición del centro de la región.

Los parámetros $\alpha$ y $\beta$ representan la anchura suave de cada intervalo bidimensional del histograma. El uso de este tipo de histogramas alivia los efectos del aliasing, proporcionando una representación más fiable que un histograma de intensidades convencional.

La figura 6.10 muestra un ejemplo de descriptor Spin tomando 3 intervalos de distancia (3 anillos) y 10 de intensidad. En este caso, la información de cada anillo está representada en los distintos histogramas por el efecto de suavizado de los parámetros $\alpha$ y $\beta$.

Con el fin de aumentar la capacidad discriminatoria del descriptor Spin, éste es calculado en nuestro sistema sobre la imagen RGB de cada región. Para preservar la constancia del color ante cambios de iluminación, la construcción del descriptor se lleva a cabo a partir de los planos RGB normalizados. Los histogramas obtenidos por cada plano de color forman el descriptor final, que puede tratarse como un histograma tridimensional en el que se representa la distribución de color de la imagen asociada.

#### 6.2.3.2. Extracción de propiedades espaciales: el *cómo*

El grupo de propiedades extraídas por el segundo subsistema de procesamiento está formado por características de las regiones que permiten responder, de una u otra forma, a "cómo" realizar una acción sobre la superficie del espacio asociada con dicha región vi-

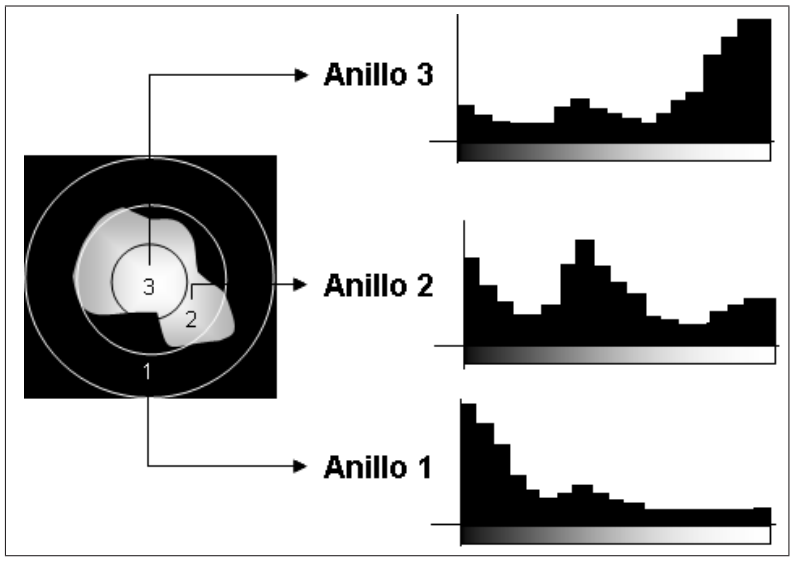



Figura 6.10: Descriptor Spin de 3 anillos

sual. Se trata, por lo tanto, de propiedades relacionadas con la posición, la orientación y el movimiento de cada región en el espacio.

**Posición**

Para obtener información espacial, tratándose de un sistema estéreo, es posible utilizar los datos procedentes de las imágenes del par de cámaras y resolver el problema mediante técnicas de visión binocular. Partiendo de que el sistema estéreo está calibrado, la cuestión inicial es la localización de la región en el espacio suponiendo un sistema de referencia centrado en el robot. Dado que el tamaño y la orientación reales de la región son aún propiedades desconocidas, esta localización se limita a un punto determinado, el centro de la región.

El proceso de detección de regiones permite tratar cada región como una esquina en su nivel de escala característico. Así, el problema de localización de regiones puede plantearse como una búsqueda de pares de puntos homólogos en los distintos niveles de la estructura multi-resolución. La geometría epipolar del sistema estéreo permite imponer restricciones en la búsqueda de pares de puntos homólogos que garantizan, en cierto modo, la fiabilidad del resultado. La restricción epipolar proporciona, para cada punto de una imagen, un conjunto de posibles homólogos en la otra imagen. Imponiendo el criterio de máxima

similitud en el entorno de pares de homólogos, medida como la correlación normalizada entre dichos entornos, obtenemos una correspondencia entre las esquinas de las dos estructuras multi-resolución. Por cada par de esquinas entre las que se ha establecido una correspondencia, se pueden calcular las coordenadas del punto 3D que se proyecta sobre ellas. Esta reconstrucción debe realizarse teniendo en cuenta que el cambio de resolución en cada nivel, implica un cambio en la distancia focal.

Los puntos 3D obtenidos por cada par de esquinas se corresponden con la posición en el espacio del centro de cada región visual. Esta identificación permite conocer la traslación entre el robot y cada zona detectada durante el proceso de extracción de regiones. Sin embargo, para lograr una correcta interacción entre el robot y su entorno, es necesario estimar propiedades adicionales de las superficies 3D que dan lugar a cada región visual. En particular, nos centraremos en la obtención de la orientación y el movimiento de las superficies localizadas.

**Orientación**

Puesto que los modelos geométricos de las superficies del entorno son desconocidos, partiremos de ciertas suposiciones a la hora de obtener características de interés sobre las regiones detectadas. Con esta idea, la estimación de la orientación se realizará bajo la hipótesis de planaridad de cada superficie situada en el espacio y obtenida por la reconstrucción de regiones homólogas del paso anterior. Aunque esta suposición, en algunos casos, no se ajuste a la realidad, sí permitirá, en ciertas situaciones, extraer información útil de acuerdo a las capacidades de interactuación del robot. Así, por ejemplo, será posible distinguir entre un objeto situado sobre el suelo o una "mancha" perteneciente al propio suelo, permitiendo ser más específico en el concepto de obstáculo.

Bajo la hipótesis de planaridad, calcular la orientación de cada superficie es equivalente a estimar el vector normal que caracteriza al plano. Todos los puntos que pertenecen al plano comparten una transformación homográfica en las dos vistas proporcionadas por el par estéreo. Así, los pares de puntos de las dos cámaras, resultantes de la proyección de los puntos del plano, cumplen la relación (Hartley y Zisserman, 2004):

$$x_d = H x_i \tag{6.6}$$

siendo $x_i$ y $x_d$ puntos en correspondencia de las imágenes izquierda y derecha, respectivamente, expresados en coordenadas homogéneas, y $H$ una matriz 3x3. Para un plano $\pi$ definido a través de su vector normal $n$ y de la distancia al plano $d$ ($\pi = (n^T, d)^T$), suponiendo un sistema de referencia centrado en la cámara izquierda, la homografía $H$ puede obtenerse como (Hartley y Zisserman, 2004):

$$H = K_d(R - tn^T/d)K_i^{-1} \tag{6.7}$$

donde $R$ y $t$ son la rotación y traslación de la cámara derecha con respecto a la izquierda y $K_i$ y $K_d$ las matrices de intrínsecos que caracterizan a ambas cámaras.

Con el fin de mantener invariable el sistema de referencia ante cambios de posición en las cámaras, la orientación es estimada bajo un sistema de coordenadas estático situado en el punto medio entre cámaras. En este nuevo sistema de referencia, el plano $\pi = (n^T, d)^T$ se transforma en $\pi_c = (n_c^T, d_c)^T$. Para incluir esta nueva definición del plano en la ecuación de la homografía (6.7), será necesario expresar $n$ y $d$ en función de $n_c$ y $d_c$. Tengamos en cuenta además que uno de los puntos del plano en el sistema de referencia central es conocido. Dicho punto es el centro de la región ($p_r$), cuya posición espacial habrá sido determinada por la reconstrucción 3D del paso anterior. A partir de este nuevo dato y de la rotación $R_i$ y traslación $t_i$ de la cámara izquierda con respecto al sistema central, podremos obtener las expresiones buscadas para $n$ y $d$. Para el caso de $n$, basta con aplicar la rotación entre los dos sistemas de referencia al vector $n_c$, obteniendo así la expresión:

$$n = R_i n_c \tag{6.8}$$

La obtención de una expresión para $d$ no es tan inmediata. Para ello, partiremos de la transformación del centro de la región $p_r$ a un punto $p_{ri}$ en el sistema de referencia de la cámara izquierda:

$$p_{ri} = R_i(p_r - t_i) \tag{6.9}$$

Puesto que $p_{ri}$ es un punto del plano $\pi$, podemos afirmar que la siguiente equivalencia es cierta:

$$p_{ri}^T n + d = 0 \tag{6.10}$$

Sustituyendo $p_{ri}$ por su equivalente de la expresión 6.9, obtenemos la ecuación:

$$(p_r - t_i)^T R_i^T n + d = 0 \tag{6.11}$$

Dado que, además, conocemos la correspondencia entre $n$ y $n_c$ por la ecuación 6.8, la expresión anterior puede escribirse como:

$$(p_r - t_i)^T n_c + d = 0 \tag{6.12}$$

de lo cual obtenemos una equivalencia final para $d$:

$$d = (t_i - p_r)^T n_c \tag{6.13}$$

Partiendo de las expresiones obtenidas para $n$ (ecuación 6.8) y $d$ (ecuación 6.13) y realizando las sustituciones correspondientes en la expresión de la homografía (6.7), se obtiene una ecuación final (6.14) donde el único dato que se debe estimar es la orientación del plano con respecto al sistema de referencia central del robot ($n_c$).

$$H = K_d \left( R - \frac{t n_c^T R_i^T}{(t_i - p_r)^T n_c} \right) K_i^{-1} \tag{6.14}$$

Mediante la definición anterior de $H$ es posible obtener $n_c$ una vez estimada la homografía a partir de las imágenes proporcionadas por las cámaras del par estéreo. Sin embargo, para que esta estimación tenga éxito, es necesario detectar al menos cuatro correspondencias entre puntos del interior de la región en ambas imágenes, lo que, en muchos casos, no es posible por falta de textura. En esta situación, no queda más remedio que plantear el problema al contrario, esto es, dados varios vectores normales al plano, cuál es el que mejor explica las imágenes resultantes de su proyección. Para ello, será necesario calcular varias homografías a partir de distintas orientaciones del plano. Debido a la complejidad computacional del proceso, esta propiedad sólo es extraída para las regiones próximas al foco de atención actual, por lo que, la orientación de regiones no podrá ser incluida como característica de selección de un objetivo visual. Será, por lo tanto, utilizada en una fase posterior del proceso, como posible elemento de influencia en la ejecución de la acción.

Por simplificación, representaremos la orientación mediante los ángulos de giro que definen la rotación del plano con respecto al sistema de referencia del robot. Esto permite definir $n_c$ como:

$$n_c = R_\alpha R_\beta (0,0,1)^T \tag{6.15}$$

siendo $R_\alpha$ y $R_\beta$ matrices de rotación calculadas, respectivamente, para un ángulo $\alpha$ de giro en el eje horizontal y otro $\beta$ de giro en el vertical. Partiendo de una cierta discretización de estos dos ángulos de giro, se obtienen varias posibles orientaciones del plano que dan lugar a distintas homografías. A través de estas homografías, la estrategia empleada consiste en calcular las proyecciones de la imagen de la región en la otra cámara y comprobar cuál es la que mantiene la máxima semejanza con la imagen real. La figura 6.11 muestra esta idea para dos posibles ángulos de giro en los ejes horizontal y vertical.

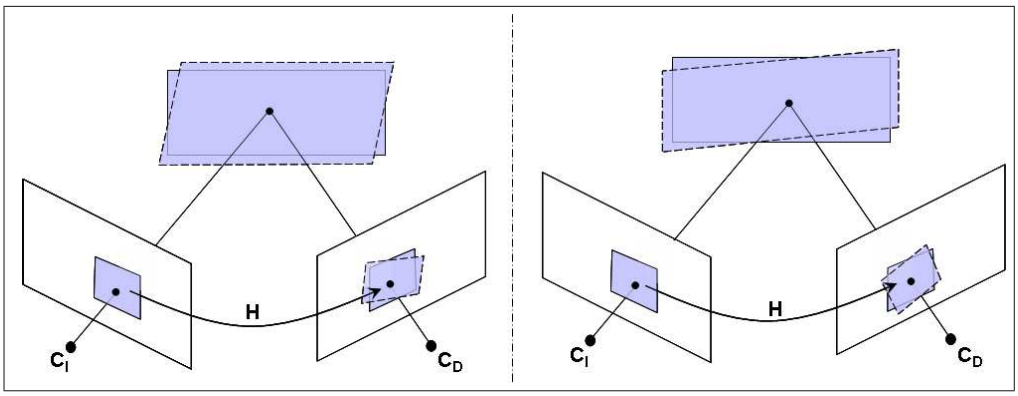



Figura 6.11: Cálculo de orientación mediante homografías

Mediante este procedimiento, el cálculo de la orientación se convierte en un proceso de aproximación a la orientación real, cuyo margen de error dependerá de la discretización realizada sobre los ángulos $\alpha$ y $\beta$. Si el número de intervalos es alto, el error de aproximación será menor que si se utiliza una discretización de menos particiones. No obstante, a medida que crezca el número de intervalos, aumentará también el tiempo de proceso afectando a fases posteriores del sistema. Buscando un compromiso entre ambas cuestiones, se ha optado por realizar un ajuste iterativo de la orientación que se mantiene durante un tiempo preestablecido, correspondiente a una fracción del periodo de procesamiento de este componente del sistema. Este ajuste consiste en delimitar de manera jerárquica el intervalo de los ángulos $\alpha$ y $\beta$ en el que se encuentra la orientación real del plano. Para ello, en cada iteración se realiza una discretización con un número fijo de intervalos y se selecciona el que mejor se ajusta a la realidad, según la homografía correspondiente. El intervalo seleccionado constituye el punto de partida de la siguiente iteración, en la que únicamente se

tiene en cuenta dicho rango de ángulos para realizar la nueva discretización. De esta forma, cada iteración constituye un paso de refinamiento en el cálculo final, llegando a un resultado en el que el error disminuye lo máximo posible dentro del tiempo asignado a este cálculo.

**Movimiento**

En un entorno estático, la posición y la orientación permiten una localización completa de las superficies del espacio. Sin embargo, en entornos reales, la capacidad de autonomía de un robot se ve limitada si no se considera el posible movimiento de los objetos presentes en la escena. Con el fin de completar las propiedades espaciales de las regiones del entorno, este segundo subsistema de procesamiento extrae información de movimiento de las superficies detectadas a partir de las características 3D ya calculadas y de propiedades obtenidas durante la fase de mantenimiento de regiones. Por simplificación y dado que cada región puede ser tratada como una esquina de imagen, se anulará la componente rotacional del movimiento. Así, las componentes de velocidad $(v_x, v_y)$ de un punto de la imagen, con coordenadas espaciales $(X^{(C)}, Y^{(C)}, Z^{(C)})$ en el sistema de referencia de la cámara, pueden obtenerse como (Trucco y Verri, 1998):

$$v_x = \frac{T_Z^{(C)} x - T_X^{(C)} f}{Z^{(C)}} \tag{6.16}$$

$$v_y = \frac{T_Z^{(C)} y - T_Y^{(C)} f}{Z^{(C)}} \tag{6.17}$$

donde $x$ e $y$ son las coordenadas del punto en la imagen, $f$ la distancia focal y $(T_X^{(C)}, T_Y^{(C)}, T_Z^{(C)})$ la traslación del punto en el espacio durante el incremento de tiempo utilizado para la toma de medidas, bajo el sistema de referencia de la cámara.

Para independizar la estimación del movimiento de los cambios de posición de la cámara, las componentes de velocidad son calculadas desde el sistema de referencia del robot, a partir de la rectificación de las coordenadas de imagen en cada captura. Así, suponiendo $R_1$ y $R_2$ las matrices de rotación de la cámara en dos capturas consecutivas, para dos puntos $p_1$ y $p_2$ en correspondencia en ambas capturas, calcularemos las componentes de velocidad en la imagen a partir de sus coordenadas rectificadas ($p_1^{(r)}$ y $p_2^{(r)}$):

$$p_1^{(r)} = f \frac{R_1^T p_1}{r_{13}^T p_1} \tag{6.18}$$

$$p_2^{(r)} = f\frac{R_2^T p_2}{r_{23}^T p_2} \tag{6.19}$$

siendo $r_{13}$ y $r_{23}$ la tercera columna de $R_1$ y $R_2$, respectivamente. A partir de las posiciones anteriores, definiremos las componentes de velocidad horizontal y vertical en la imagen como $v_x^{(r)}$ y $v_y^{(r)}$. Dado que los sistemas de referencia de la cámara y del robot se encuentran alineados en el eje horizontal, las ecuaciones que relacionan el movimiento 3D de un punto, de coordenadas espaciales $(X, Y, Z)$ en el sistema de referencia del robot, con la velocidad en el plano de imagen pueden expresarse como:

$$v_x^{(r)} = \frac{T_Z x^{(r)} - T_X f}{Z} \tag{6.20}$$

$$v_y^{(r)} = \frac{T_Z y^{(r)} - T_Y f}{Z} \tag{6.21}$$

donde $(x^{(r)}, y^{(r)})$ son las coordenadas rectificadas del punto en la imagen y $(T_X, T_Y, T_Z)$ la componente traslacional del movimiento desde el sistema de referencia del robot.

Las ecuaciones 6.20 y 6.21 no permiten estimar por sí solas la traslación de una región en el espacio, pero sí proporcionarán cierta información de movimiento de aquellas regiones para las que no sea posible realizar un cálculo de velocidad a partir de su posición 3D medida a lo largo del tiempo. Esta situación se puede dar con frecuencia por varios motivos. El más habitual es que la posición actual de vergencia provoque disparidades altas en determinadas regiones, impidiendo la localización de pares de regiones homólogas y, consecuentemente, la obtención de una posición espacial para dichas regiones. No obstante, las expresiones anteriores proporcionan un medio adicional para mantener información espacial de las regiones del campo visual. A partir de ellas será posible obtener, para cada región, información relativa a sus cambios de posición espacial, lo que permitirá, por ejemplo, tomar decisiones acerca de la fijación de atención en una determinada zona para realizar un análisis más exhaustivo sobre ella.

### 6.2.4. Selección del foco de atención

La selección del foco de atención se realiza a través de varios componentes de control que deciden, individualmente, la zona de fijación mediante la evaluación de las propiedades de las regiones en base a criterios específicos. La diversificación de los componentes de control de atención proporciona serias ventajas con respecto a un control global. En pri-

mer lugar, permite un diseño más claro y simplificado del proceso de selección individual. En segundo lugar, admite la coexistencia de distintos tipos de objetivos visuales proporcionando una base fundamental para un control de atención guiado por tarea. En este sentido, la suposición de la que partimos es que es necesario separar la forma de integrar las propiedades de las regiones del entorno en función de los objetivos (Bachiller et al., 2007). Por ejemplo, en un problema básico de navegación en el que el robot debe moverse por su entorno siguiendo un grupo de balizas, encontramos dos tipos de objetivos visuales: las balizas que guían la navegación y los posibles obstáculos. Las propiedades que ambos tipos de regiones mantienen como elementos de atracción en una de las tareas podrían transformarles en zonas de distracción en la otra, lo que complicaría la integración de información relevante en un selector único que permitiera guiar la atención. Extendiendo esta idea y relacionándola con la propia supervivencia del robot, surge otra cuestión de gran importancia: ¿cómo incluir la reacción ante lo inesperado a la vez que se mantienen otros objetivos? Si el control atencional se encuentra centralizado, este asunto es complicado, puesto que, al igual que ocurría en el ejemplo de la baliza y el obstáculo, lo inesperado y lo buscado pueden estar definidos por propiedades contradictorias que provocarán un estado de alerta excesivo en algunos casos, o bien, una respuesta pasiva ante situaciones de peligro.

La naturaleza distribuida del control atencional que planteamos en nuestro sistema está en cierto modo relacionada con la teoría premotora de la atención (Rizzolatti et al., 1987). De acuerdo con esta teoría, la atención deriva de la activación de diversos circuitos neuronales, denominados mapas pragmáticos, que codifican el espacio de diferentes formas con finalidad ejecutiva. De manera similar, la atención en nuestro sistema emerge de la activación de varios selectores desde los que se realizan diferentes interpretaciones de la escena en función de los objetivos conductuales. Asimismo, en la línea que proclama la teoría premotora, la atención encubierta en el sistema propuesto surge de la activación de un programa motor, es decir, del control de un selector de objetivo, cuyo resultado no llega a ejecutarse.

#### 6.2.4.1. Proceso de selección individual

Cada selector de objetivo se encarga de obtener un foco de atención cuyas propiedades se ajusten lo máximo posible a los criterios de selección acordes con él. Para ello, debe construir un mapa de saliencia que represente la importancia de cada región en función de los objetivos. Dicho mapa actúa como una superficie de control cuyos máximos se corres-

ponden con las zonas del campo visual candidatas a la fijación de atención.

Existen varias alternativas para la construcción de los mapas de saliencia. En la mayoría de sistemas que enfatizan la integración de información *top-down* y características *bottom-up*, la obtención del mapa se lleva a cabo a través de un conjunto de pesos que ponderan la relevancia de determinadas características en el proceso de selección (Navalpakkam y Itti, 2006) (Frintrop, 2006). Este método de integración presenta una serie de inconvenientes. En primer lugar, la manera de combinar la información *top-down* con las características de la imagen es fija, por lo que no existe distinción en la forma de componer diferentes criterios que lleven a la selección de un objetivo visual. Además, el ajuste de pesos no siempre permite asegurar la selección del mejor candidato, dado que el carácter lineal del método restringe las posibles ordenaciones que pueden realizarse sobre el espacio de características. Estas limitaciones dificultan la realización práctica de los mecanismos que proporcionan el control de atención, más aún cuando dicho control está orientado a la selección para la acción, donde no basta con seleccionar un posible candidato, sino que es necesario fijar la atención sobre el que mejor se ajusta a cada situación. Por otro lado, la lógica borrosa (Zadeh, 1965) proporciona un marco de desarrollo que resulta muy adecuado para esta aplicación, principalmente por 3 motivos:

- Las propiedades de las regiones utilizadas en la selección mantienen un cierto grado de incertidumbre e imprecisión. La lógica borrosa permite manejarlas de manera adecuada a través de variables lingüísticas.
- Los criterios de selección pueden definirse de manera sencilla a través de reglas borrosas que asocien el cumplimiento de determinadas características con un nivel de saliencia.
- Las salidas de las distintas reglas pueden fusionarse mediante métodos de inferencia que proporcionan un mecanismo de integración de criterios de selección que llevan a la obtención de un único mapa de saliencia.

En varios trabajos se han aplicado con éxito técnicas borrosas para el diseño de sistemas de atención que proporcionan un control dirigido por estímulo (Brown et al., 2000)(Brown et al., 2001). La propuesta que aquí se plantea permite un control de atención dirigido por objetivo integrando información *top-down*, a través de la definición de criterios de selección,

y propiedades *bottom-up* de los estímulos. Además, la metodología de diseño empleada difiere sustancialmente de las utilizadas en trabajos similares.

Los siguientes apartados describen el proceso de diseño del sistema propuesto. En los tres primeros, se detallan las distintas fases que constituyen este proceso. A continuación, se resumen los aspectos más relevantes del método de diseño empleado y, por ultimo, se muestra un ejemplo de aplicación para un selector de objetivo concreto.

**Diseño del sistema: consideraciones generales**

Siguiendo los principios del razonamiento borroso, el mapa de saliencia se obtiene a partir de la definición de un conjunto de reglas borrosas que asocian el cumplimiento de determinados criterios (antecedente) a diferentes niveles de saliencia (consecuente). Las premisas que componen el antecedente de cada regla pueden expresarse a través de etiquetas lingüísticas sobre las propiedades de las regiones visuales de la escena. Estas etiquetas dan lugar a conjuntos borrosos definidos en el dominio de las propiedades relevantes para la selección del foco de atención. Para especificar los consecuentes de las reglas existen varias alternativas en función del modelo utilizado.

Los dos sistemas borrosos de mayor relevancia son los propuestos por Mamdani y Assilian (Mamdani y Assilian, 1975) y por Takagi y Sugeno (Takagi y Sugeno, 1985). En el primero, la salida, al igual que cada entrada del sistema, es tratada como una variable lingüística a la que hay que asociar una partición borrosa. Tras la evaluación de reglas, para obtener un valor concreto de salida, es necesario llevar a cabo una fase de desborrosificación. En el enfoque Takagi-Sugeno se mantiene la misma especificación de las particiones borrosas de los dominios de las entradas que en el modelo de Mamdani, pero no se requiere una partición borrosa del dominio de salida. En su lugar, la salida de cada regla es expresada a través de una función de las entradas del sistema. Este último enfoque resulta más cómodo y sencillo de aplicar y es adecuado para nuestro propósito, ya que la salida de cada regla debe cuantificar la importancia de una región en el proceso de selección del foco de atención, por lo que puede ser tratada como una función de las entradas. Así, basándonos en el enfoque Takagi-Sugeno, cada una de las reglas ($R_r$) incluidas en el proceso de selección tendría la siguiente forma:

$R_r$: SI $P_1$ es $A_i$ Y ... Y $P_n$ es $X_k$ ENTONCES *saliencia* $= F_r(P_1, ..., P_n)$

donde $P_1, .., P_n$ son propiedades de las regiones, $A_i, ..., X_k$ conjuntos borrosos definidos sobre dichas propiedades y $F_r(P_1, ..., P_n)$ es una función que proporciona un nivel de saliencia de acuerdo a las entradas del sistema.

Mediante esta metodología, el diseño del sistema consiste en la determinación de las particiones borrosas de las variables de entrada, la definición de reglas y funciones de salidas y la aplicación de un método de inferencia que proporcione la salida final del sistema. En sistemas borrosos que utilizan el modelo de Takagi-Sugeno, las siguientes consideraciones de diseño resultan especialmente eficientes (Hanss, 1999) y serán las que utilicemos en nuestros sistemas de selección:

1. Los conjuntos borrosos ($A_i$) de las variables de entrada del sistema ($x$) están definidos mediante funciones de pertenencia triangulares ($\mu_{Ai}(x)$), que satisfacen la condición:

$$\sum_{i=1}^{L} \mu_{Ai}(x)^1 = 1 \; \forall \text{ instancia de } x \tag{6.22}$$

   siendo $L$ el número total de conjuntos. Cada conjunto, por lo tanto, debe solaparse con sus adyacentes con un nivel de cruce de 0.5.

2. Las reglas del sistema deben cubrir completamente el espacio de las entradas. Es decir, para cualquier combinación de valores de entrada, existe un grupo de reglas que permiten determinar el valor de salida.

3. Las funciones de salida de cada regla son lineales: $F_r(x_1, ..., x_n) = c_0^{(r)} + c_1^{(r)} * x_1 + \cdots + c_n^{(r)} * x_n$, siendo $c_0^{(r)}, c_1^{(r)} \cdots c_n^{(r)}$ los coeficientes de la función y $x_1 \cdots x_n$ las variables de entrada del sistema.

4. La inferencia o disparo de cada regla, consistente en la combinación de los valores de pertenencia en cada premisa del antecedente para obtener el grado de cumplimiento del consecuente ($\alpha_r$), se llevará a cabo mediante las operaciones de producto algebraico (conjunción de premisas, ecuación 6.23) y suma algebraica (disyunción de

[1] Por simplificación, a partir de ahora nos referiremos a $\mu_{Ai}(x)$ como $\mu_{Ai}$

premisas, ecuación 6.24) para evitar no linealidades indeseadas en el modelo.

$$R_r: \text{ SI } x_1 \text{ es } A_i \text{ Y } x_2 \text{ es } B_j \text{ ENTONCES...:} \quad \alpha_r = \mu_{Ai} * \mu_{Bj} \tag{6.23}$$

$$R_r: \text{ SI } x_1 \text{ es } A_i \text{ O } x_2 \text{ es } B_j \text{ ENTONCES...:} \quad \alpha_r = \mu_{Ai} + \mu_{Bj} - \mu_{Ai} * \mu_{Bj} \tag{6.24}$$

5. La salida final, para un sistema de $R$ reglas, se obtendrá mediante el centro de masas de las salidas de cada regla a partir de sus correspondientes grados de cumplimiento:

$$\frac{\sum_{r=1}^{R} \alpha_r * F_r(x_1, ..., x_n)}{\sum_{r=1}^{R} \alpha_r} \tag{6.25}$$

   Si las consideraciones 1, 2 y 4 se cumplen, la suma de los grados de cumplimiento del total de reglas es la unidad. Así, la expresión 6.25 puede obtenerse como:

$$\sum_{r=1}^{R} \alpha_r * F_r(x_1, ..., x_n) \tag{6.26}$$

Una vez establecidas las particiones borrosas de las variables de entrada, se puede especificar la base de reglas que compondrá el bloque de razonamiento del sistema. El punto crucial de esta fase consiste en la determinación de los coeficientes de las funciones de salida que proporcionan el funcionamiento deseado. En nuestro caso, cada regla permite evaluar el cumplimiento de un criterio de selección/exclusión que proporciona un valor de importancia de cada región con respecto al resto. Esto es, las distintas reglas permiten ordenar las regiones mediante la evaluación de propiedades relevantes para su selección en la fijación de atención, asignándoles un valor de saliencia que debe ser alto para regiones que cumplan los criterios de selección y bajo para las coherentes con los criterios de exclusión. Puesto que la prioridad en la selección de la región depende del grado de cumplimiento de las reglas y no directamente del valor de sus propiedades, las funciones de salida pueden expresarse a través de valores constantes, que permiten discretizar los posibles niveles de saliencia.

La cuestión que se plantea ahora es qué valor de saliencia debemos asignar a cada regla para que la ordenación de regiones sea la adecuada. Para ello distinguiremos dos tipos

de reglas en el sistema: reglas excluyentes y reglas de selección. Los valores de saliencia estrictamente negativos (considerando el 0 como valor negativo) estarán asociados con reglas excluyentes, es decir, reglas cuyas hipótesis definen regiones que no tienen ningún interés en el proceso de selección y que, por lo tanto, no deben ser atendidas. Los valores positivos impondrán un orden en la toma del control de atención de cada región, dando prioridad a aquella región cuyo valor de saliencia sea más alto. Así, la distribución de valores positivos entre las reglas no excluyentes se realizará de manera que las reglas que definan un nivel más alto de cumplimiento de los criterios de selección tomarán un valor de saliencia más alto que aquellas que expresen un menor grado de cumplimiento de dichos criterios. Por ejemplo, partiendo de una evaluación de regiones a través de dos propiedades $P_1$ y $P_2$ particionadas en 3 ($A_1$, $A_2$ y $A_3$) y 2 ($B_1$ y $B_2$) conjuntos borrosos, respectivamente, definiríamos el siguiente grupo de reglas:

$R_{11}$: SI $P_1$ es $A_1$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{11}$
$R_{12}$: SI $P_1$ es $A_1$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{12}$
$R_{21}$: SI $P_1$ es $A_2$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{21}$
$R_{22}$: SI $P_1$ es $A_2$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{22}$
$R_{31}$: SI $P_1$ es $A_3$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{31}$
$R_{32}$: SI $P_1$ es $A_3$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{32}$

Suponiendo que las 3 primeras son reglas excluyentes y que las restantes están ordenadas por grado de cumplimiento de los criterios de selección, los valores de saliencia deben cumplir:

$$s_{11} \leq 0 \; s_{12} \leq 0 \; s_{21} \leq 0$$

$$0 < s_{22} \leq s_{31} \leq s_{32}$$

Esto nos permite establecer el orden de saliencia en función de cada regla y asignar valores que permiten descartar regiones durante el proceso de selección. Sin embargo, la separación entre reglas excluyentes y reglas de selección no es suficiente para realizar una asignación de valores concretos a los consecuentes del sistema de reglas. Una posibilidad sería establecer dichos valores de manera arbitraria, pero, en tal caso, el comportamiento del sistema sería indeterminado ante situaciones no expresadas directamente por el grupo de reglas originales. Para evitar un comportamiento indeseado, los consecuentes de las reglas serán fijados de acuerdo a dos tipos de consideraciones:

- Imposición de condiciones adicionales: será necesario especificar cómo debe comportarse el sistema ante situaciones no descritas directamente por ninguna regla. Estas condiciones permitirán relacionar matemáticamente los consecuentes del sistema de reglas, proporcionando un medio para determinar una posible combinación de valores de saliencia del total de reglas que den lugar al funcionamiento deseado.
- Agrupamiento de reglas: cuando el sistema deba mantener un comportamiento similar ante situaciones expresadas por distintas reglas, deberá imponerse un mismo valor de saliencia a todas ellas. Esto permitirá que, en algunos casos, sea posible reducir el número de reglas del sistema sustituyendo grupos de reglas de consecuente común por reglas únicas que engloben sus distintas premisas.

En los dos apartados siguientes, se desarrollan estos dos mecanismos para la determinación de los consecuentes de la base de reglas que constituirán la última fase de diseño del sistema.

**Determinación de los consecuentes del sistema: condiciones adicionales**

La salida del sistema propuesto representa un valor de prioridad en el proceso de fijación de atención. A la hora de establecer los consecuentes concretos de cada regla, debe asegurarse que los cambios de prioridad se produzcan en el momento en que se cumplan las circunstancias adecuadas. Para expresar estas variaciones de prioridad será necesario imponer condiciones adicionales que describan cuál debe ser el comportamiento del sistema ante ciertas situaciones, es decir, qué prioridad deberá asignarse en cada caso. La inclusión de nuevas condiciones de funcionamiento fuerza a que la salida del sistema (ecuación 6.26) tome un valor concreto cuando se producen los valores de las entradas que identifican tales condiciones. Esto permite obtener un grupo de expresiones en las que se fijan las relaciones entre los consecuentes del sistema de reglas que proporcionan el comportamiento deseado. A partir de dichas relaciones será posible concretar valores para los consecuentes de las reglas, obteniendo un sistema final que mantiene un funcionamiento coherente con todas las condiciones impuestas.

La situación descrita por cada condición adicional puede ser expresada a través de nuevas premisas sobre las entradas del sistema en las que se impondrá un determinado grado de pertenencia de cada entrada a los conjuntos de su partición. Esta nueva forma de

expresión dará lugar a lo que denominaremos reglas derivadas.

Las reglas derivadas mantienen la estructura de las reglas originales, pero representan un grado de cumplimiento determinado de cada una de sus premisas. A través de ellas es posible expresar cuál debe ser la salida del sistema si una hipótesis sobre una entrada se cumple más o menos, si lo hace completamente o si cumple por igual dos hipótesis. Para incluir el grado de cumplimiento de una premisa en la especificación de cada regla derivada, anotaremos entre paréntesis dicho valor al lado de la premisa correspondiente. Puesto que el grado de cumplimiento puede ser menor que 1, una regla puede incluir varias hipótesis sobre la misma variable. Por simplificación, haremos explícita una única premisa sobre cada variable y supondremos el cumplimiento complementario de la premisa asociada con el conjunto adyacente de mayor orden. Por ejemplo, en un sistema con dos entradas, el antecedente de una regla derivada tendría la forma "SI $P_1$ es $A_i$ ($\hat{\mu}_{Ai}$) Y $P_2$ es $B_j$ ($\hat{\mu}_{Bj}$)", siendo $\hat{\mu}_{Ai}$ y $\hat{\mu}_{Bj}$ valores concretos de pertenencia. La primera premisa expresa un grado de cumplimiento $\hat{\mu}_{Ai}$ de la hipótesis "$P_1$ es $A_i$" y el cumplimiento complementario $(1-\hat{\mu}_{Ai})$ de la hipótesis "$P_1$ es $A_{i+1}$". De igual forma, la segunda parte del antecedente representaría una situación en la que $P_2$ pertenece a $B_j$ con un grado de pertenencia de $\hat{\mu}_{Bj}$ y a $B_{j+1}$ con pertenencia $1-\hat{\mu}_{Bj}$.

El consecuente de una regla derivada debe atribuir, a través de su saliencia, un nivel de prioridad coherente con la situación expresada por su antecedente. Los posibles niveles de prioridad vienen determinados por el orden de los consecuentes de las reglas de selección originales. Así, para que una regla derivada asigne un nivel de prioridad determinado, su consecuente deberá coincidir con el de la regla de selección que mantenga ese mismo nivel de prioridad. Será posible, además, expresar una prioridad nula, asignando un valor 0 al consecuente de la regla, cuando se desee obtener un resultado excluyente en el proceso de selección. En este último caso, la regla se utilizará para indicar las condiciones iniciales de exclusión de una región.

Supongamos, por ejemplo, el siguiente sistema de reglas en el que las 3 primeras son reglas de exclusión y las restantes reglas de selección, ordenadas de menor a mayor prioridad ($s_{22} < s_{31} < s_{32}$):

$R_{11}$: SI $P_1$ es $A_1$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{11}$
$R_{12}$: SI $P_1$ es $A_1$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{12}$

$R_{21}$: SI $P_1$ es $A_2$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{21}$
$R_{22}$: SI $P_1$ es $A_2$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{22}$
$R_{31}$: SI $P_1$ es $A_3$ Y $P_2$ es $B_1$ ENTONCES $saliencia = s_{31}$
$R_{32}$: SI $P_1$ es $A_3$ Y $P_2$ es $B_2$ ENTONCES $saliencia = s_{32}$

Podríamos definir las siguientes reglas derivadas para imponer condiciones adicionales en el funcionamiento del sistema:

$RD_1$: SI $P_1$ es $A_2$ (1) Y $P_2$ es $B_1$ (0.5) ENTONCES $saliencia = 0$
$RD_2$: SI $P_1$ es $A_2$ (0.5) Y $P_2$ es $B_1$ (1) ENTONCES $saliencia = s_{22}$
$RD_3$: SI $P_1$ es $A_2$ (0.5) Y $P_2$ es $B_1$ (0.25) ENTONCES $saliencia = s_{31}$

La situación definida en el antecedente de cada regla está caracterizada por un grado de cumplimiento concreto en cada una de sus premisas. Además, cada premisa es doble, en el sentido de que determina la pertenencia de una entrada al conjunto indicado y la pertenencia complementaria a su conjunto adyacente de mayor orden. Así, la premisa de $RD_1$ "$P_2$ es $B_1$ (0.5)" expresa un cumplimiento equilibrado de las premisas "$P_2$ es $B_1$" y "$P_2$ es $B_2$". De igual manera, la premisa "$P_1$ es $A_2$ (0.5)" de $RD_2$ y $RD_3$ representa la pertenencia intermedia de $P_1$ a $A_2$ y a $A_3$. Para expresar el mayor o menor cumplimiento de una premisa frente a otra, basta con indicar el grado de pertenencia de la variable al conjunto de menor orden. Este es el caso de la premisa "$P_2$ es $B_1$ (0.25)" de $RD_3$, que representa el cumplimiento en menor grado de "$P_2$ es $B_1$" frente a "$P_2$ es $B_2$".

Los consecuentes de las reglas derivadas asignan una prioridad determinada a las situaciones descritas por sus antecedentes. El sistema de reglas del ejemplo define tres niveles de prioridad, sin considerar el nivel nulo, correspondientes a los tres consecuentes de las reglas de selección ($s_{22}$ -prioridad baja-, $s_{31}$ -prioridad media- y $s_{32}$ -prioridad alta-). El diseño de reglas derivadas debe tener en cuenta estos niveles para asignar una prioridad concreta a cada situación. Tal y como se observa en los consecuentes de las reglas derivadas definidas en el ejemplo, la asignación de niveles se lleva a cabo atribuyendo a cada salida el consecuente de la regla de selección que identifique el nivel correspondiente, o el valor 0 para indicar la prioridad nula. Concretamente, la regla $RD_1$ atribuye una saliencia nula, lo que implica que su antecedente define las condiciones iniciales de exclusión de regiones. Asimismo, las reglas $RD_2$ y $RD_3$ identifican condiciones de selección con niveles de prioridad bajo y medio, respectivamente.

Tras imponer las condiciones adicionales de funcionamiento del sistema, a partir de la definición de reglas derivadas, es posible formar un grupo de ecuaciones que asocian el consecuente de cada regla derivada con la expresión de la salida real que se obtiene por la situación especificada en su antecedente. Para obtener estas ecuaciones, será necesario sustituir, en la expresión general de la salida del sistema, el grado de pertenencia a los conjuntos de cada partición por los grados de pertenencia especificados en cada regla derivada. Por ejemplo, para un sistema de selección que utilice dos propiedades $P_1$ y $P_2$ como variables de entrada, particionadas en $LA$ y $LB$ conjuntos, partiríamos de la salida global del sistema obtenida a partir de la expresión 6.27,

$$saliencia = \sum_{i=1}^{LA} \sum_{j=1}^{LB} \alpha_{ij} * s_{ij} \tag{6.27}$$

siendo $\alpha_{ij}$ y $s_{ij}$, respectivamente, el grado de cumplimiento del antecedente y el valor del consecuente de cada regla $R_{ij}$ del sistema:

$$R_{ij}\text{: SI } P_1 \text{ es } A_i \text{ Y } P_2 \text{ es } B_j \text{ ENTONCES } saliencia = s_{ij}$$

Puesto que las premisas de una regla están unidas a través del operador de conjunción (Y), $\alpha_{ij}$ viene dado por $\mu_{Ai} * \mu_{Bj}$, siendo $\mu_{Ai}$ el grado de pertenencia de $P_1$ al conjunto $A_i$ y $\mu_{Bj}$ el grado de pertenencia de $P_2$ a $B_j$. De esta forma, la expresión general de la salida del sistema puede obtenerse como:

$$saliencia = \sum_{i=1}^{LA} \sum_{j=1}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} \tag{6.28}$$

Utilizando la ecuación 6.28, es posible obtener la expresión de la salida de una regla derivada del sistema, sustituyendo cada $\mu_{Ai}$ y $\mu_{Bj}$ por los valores de pertenencia indicados en su definición. Así, para una regla derivada como la siguiente,

$$RD\text{: SI } P_1 \text{ es } A_i \ (\hat{\mu}_{Ai}) \text{ Y } P_2 \text{ es } B_j \ (\hat{\mu}_{Bj}) \text{ ENTONCES } saliencia = s_{kl}$$

que define grados de pertenencia no nulos de la entrada $P_1$ a los conjuntos $A_i$ y $A_{i+1}$, y de la entrada $P_2$ a $B_j$ y $B_{j+1}$, la salida real del sistema vendría dada por:

$$\begin{aligned} saliencia &= \hat{\mu}_{Ai} * \hat{\mu}_{Bj} * s_{ij} + (1 - \hat{\mu}_{Ai}) * \hat{\mu}_{Bj} * s_{i+1j} + \\ &+ \hat{\mu}_{Ai} * (1 - \hat{\mu}_{Bj}) * s_{ij+1} + (1 - \hat{\mu}_{Ai}) * (1 - \hat{\mu}_{Bj}) * s_{i+1j+1} \end{aligned} \tag{6.29}$$

Igualando la expresión anterior al consecuente ($s_{kl}$) de la nueva regla, llegamos finalmente a una representación matemática de la condición impuesta en el sistema que proporciona una relación entre los valores de saliencia asignados por la base de reglas.

Para el caso concreto del ejemplo expuesto anteriormente, las reglas derivadas $RD_1$, $RD_2$ y $RD_3$ establecen los grados de pertenencia no nulos indicados en 6.30, 6.31 y 6.32.

$$RD_1 \;:\; \mu_{A2} = 1 \quad \mu_{B1} = 0.5 \quad \mu_{B2} = 0.5 \tag{6.30}$$

$$RD_2 \;:\; \mu_{A2} = 0.5 \quad \mu_{A3} = 0.5 \quad \mu_{B1} = 1 \tag{6.31}$$

$$RD_3 \;:\; \mu_{A2} = 0.5 \quad \mu_{A3} = 0.5 \quad \mu_{B1} = 0.25 \quad \mu_{B2} = 0.75 \tag{6.32}$$

Aplicando estos grados de pertenencia a la expresión de la salida del sistema (6.29), por cada regla derivada se obtiene una expresión concreta de salida que, igualada al consecuente correspondiente, permite establecer las relaciones 6.33, 6.34 y 6.35. A partir de ellas, se fijarán los valores concretos de los consecuentes del grupo de reglas originales, obteniendo un sistema final que proporciona un comportamiento acorde con todas las condiciones impuestas durante el diseño.

$$RD_1 \;:\; 0.5 * s_{21} + 0.5 * s_{22} = 0 \tag{6.33}$$

$$RD_2 \;:\; 0.5 * s_{21} + 0.5 * s_{31} = s_{22} \tag{6.34}$$

$$RD_3 \;:\; 0.125 * s_{21} + 0.375 s_{22} + 0.125 * s_{31} + 0.375 * s_{32} = s_{31} \tag{6.35}$$

**Determinación de los consecuentes del sistema: agrupamiento de reglas**

Otra consideración a la hora de concretar las salidas de las reglas es el agrupamiento de reglas adyacentes, es decir, la asignación de un consecuente común, siempre que sea posible, a grupos de reglas de este tipo. Como se verá a continuación, este paso de diseño permitirá reducir el número total de reglas que componen el sistema.

Consideraremos grupo de reglas adyacentes al conjunto de reglas definidas para todas las posibles hipótesis de un grupo de propiedades y con hipótesis comunes en las restantes. Por ejemplo, suponiendo un sistema de reglas con 2 variables de entrada $P_1$ y $P_2$, cuyas particiones correspondientes $A$ y $B$ están formadas por $L_A$ y $L_B$ conjuntos borrosos, se consideran grupos de reglas adyacentes a los conjuntos de reglas que mantienen cualquiera

de las siguientes estructuras:

$$\forall i/1 \leq i \leq L_A : \text{SI } P_1 \text{ es } A_i \text{ Y } P_2 \text{ es } B_m \text{ ENTONCES } saliencia = s_{im} \tag{6.36}$$

$$\forall j/1 \leq j \leq L_B : \text{SI } P_1 \text{ es } A_n \text{ Y } P_2 \text{ es } B_j \text{ ENTONCES } saliencia = s_{nj} \tag{6.37}$$

Para el primer caso, imponiendo un valor común $c$ a los consecuentes ($s_{im}$) de todas las reglas, el valor final de salida puede expresarse como:

$$\begin{aligned} saliencia &= \sum_{i=1}^{LA} \sum_{j=1,j\neq m}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} + \sum_{i=1}^{LA} \mu_{Ai} * \mu_{Bm} * c \\ &= \sum_{i=1}^{LA} \sum_{j=1,j\neq m}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} + \mu_{Bm} * c \end{aligned} \tag{6.38}$$

La equivalencia anterior permite reducir el grupo de reglas de la estructura 6.36 a una única regla formada por la hipótesis común:

$$\text{SI } P_2 \text{ es } B_m \text{ ENTONCES } saliencia = c$$

Si, además, quisiéramos minimizar el grupo de reglas de la estructura 6.37, asignando el mismo valor de salida $c$ a todas ellas, se obtiene la siguiente salida global:

$$\begin{aligned} saliencia &= \sum_{i=1,i\neq n}^{LA} \sum_{j=1,j\neq m}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} + \mu_{Bm} * c + \mu_{An} * \sum_{j=1,j\neq m}^{LB} \mu_{Bj} * c \\ &= \sum_{i=1,i\neq n}^{LA} \sum_{j=1,j\neq m}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} + \mu_{Bm} * c + \mu_{An} * (1 - \mu_{Bm}) * c \\ &= \sum_{i=1,i\neq n}^{LA} \sum_{j=1,j\neq m}^{LB} \mu_{Ai} * \mu_{Bj} * s_{ij} + (\mu_{An} + \mu_{Bm} - \mu_{An} * \mu_{Bm}) * c \end{aligned} \tag{6.39}$$

La expresión $\mu_{An} + \mu_{Bm} - \mu_{An} * \mu_{Bm}$ de la ecuación 6.39 es la suma algebraica entre $\mu_{An}$ y $\mu_{Bm}$ que cuantifica el grado de cumplimiento de la disyunción entre dos hipótesis. De esta forma, el conjunto de reglas englobado por las dos estructuras 6.36 y 6.37 podría representarse a través de la regla:

$$\text{SI } P_1 \text{ es } A_n \text{ O } P_2 \text{ es } B_m \text{ ENTONCES } saliencia = c$$

Siguiendo esta idea, siempre que sea posible, asignaremos salidas comunes a grupos de reglas adyacentes, lo que, en nuestro caso, tendrá sentido principalmente cuando dichos grupos estén formados por reglas excluyentes.

**Resumen del proceso de diseño**

A partir de todas las consideraciones de diseño expuestas, la construcción de un selector de objetivo implicaría:

1. Definir las particiones borrosas de las propiedades que intervienen en el sistema mediante conjuntos triangulares con nivel de cruce de 0.5.
2. Definir el sistema de reglas de manera que los antecedentes engloben las posibles combinaciones de hipótesis que puedan realizarse con las particiones de las variables de entrada.
3. Identificar las reglas excluyentes y ordenar las restantes de la más a la menos adecuada para la fijación de atención.
4. Asignar salidas constantes a cada regla teniendo en cuenta que las reglas excluyentes deben tener salida estrictamente negativa y la de mayor orden, el mayor valor de salida. Para fijar estos valores, se impondrán condiciones adicionales de funcionamiento, mediante la definición de reglas derivadas, y se favorecerá el agrupamiento de reglas.
5. Utilizar como método de inferencia para la obtención de la salida del sistema el centro de masas, aplicando las operaciones de producto y suma algebraica en el cálculo del grado de cumplimiento de cada regla.

Una vez fijado el conjunto de reglas, el sistema estaría operativo para evaluar las regiones a partir de ciertas propiedades, asignándoles un valor de saliencia que permite descartar del proceso de selección las que presenten una saliencia negativa y ordenar las restantes de mayor a menor importancia en la fijación de atención. La selección final se realiza escogiendo aquellas regiones cuya saliencia difiera en menos de un determinado porcentaje del valor máximo, pudiendo así elegir entre obtener varios candidatos a focos de atención o fijar la atención en el que maximice los criterios de selección.

**Ejemplo de diseño de un selector de objetivo**

Para ilustrar todo el proceso de diseño, se muestra a continuación un selector de obstáculos construido mediante esta metodología. El robot deberá detectar los posibles obstáculos que se encuentran situados en la trayectoria hacia una posición objetivo. Consideraremos 3 propiedades para determinar la cualidad de obstáculo de las regiones visuales del entorno:

- *Profundidad relativa (Pr)*: es la relación (expresión 6.40) entre la profundidad de la región ($Z_R^{(S_T)}$) y la profundidad del objetivo ($Z_O^{(S_T)}$). Ambas profundidades se calculan sobre un sistema de referencia ($S_T$) situado en la posición en el suelo del robot cuyo eje $Z$ se corresponde con la trayectoria en línea recta hacia el objetivo. Esta relación cuantifica el grado en el que una región se encuentra situada a una profundidad más cercana al robot que al objetivo, más al objetivo que al robot, o detrás del objetivo. Un valor mayor que 1 de esta relación puede interpretarse como zona del entorno situada fuera de una posible trayectoria hacia el objetivo.

$$Pr = \frac{Z_R^{(S_T)}}{Z_O^{(S_T)}} \tag{6.40}$$

- *Desviación (Dv)*: cuantifica el grado de cercanía de la región a la trayectoria en línea recta hacia el objetivo. Es medida como la distancia de la región a dicha recta. Tomando como sistema de referencia $S_T$, es equivalente al valor absoluto de la coordenada $X$ de la región expresada en dicho sistema de referencia (expresión 6.41).

$$Dv = |X_R^{(S_T)}| \tag{6.41}$$

- *Altura (Al)*: Las dos propiedades anteriores tienen en cuenta la distancia, paralela al suelo, de una región al robot. Sin embargo, aquellas regiones que se alejen en altura del robot no deben ser incluidas en el proceso de selección de obstáculos, ya que no interferirán en la trayectoria hacia el objetivo. Con esta idea, se incluye la altura de las regiones como la última propiedad que permite formar los criterios de selección del sistema.

Una vez definidas las propiedades que intervendrán en el sistema, hay que diseñar los conjuntos borrosos de cada una de ellas. Para la propiedad *profundidad relativa* (figura 6.12) se definen 3 conjuntos, etiquetados como *CercaR*, *CercaO* y *Lejos*. El primero de

ellos considera profundidades más cercanas al robot que al objetivo, el segundo más alejadas del robot que del objetivo y el tercero valores de profundidad asociados con regiones situadas detrás del objetivo.

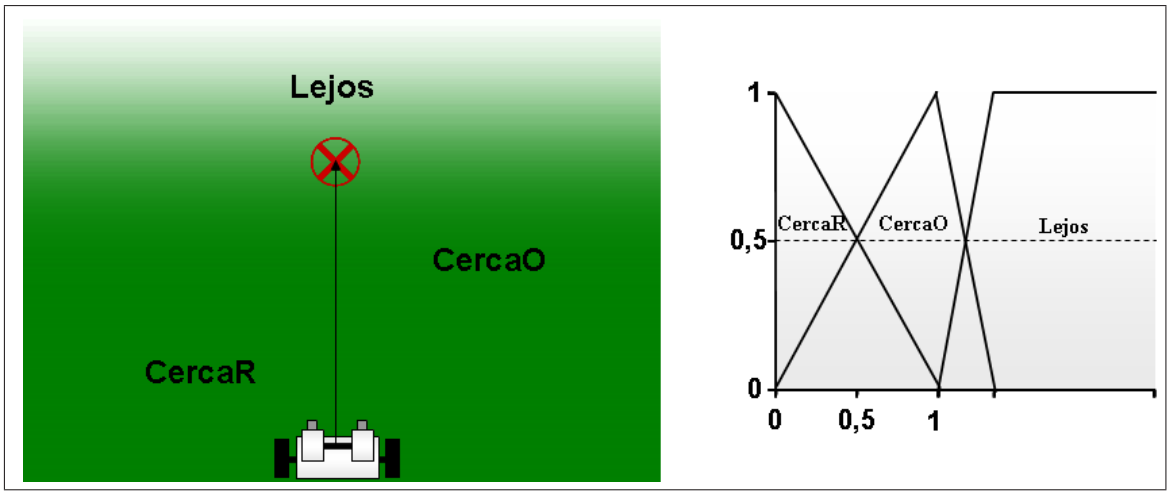



Figura 6.12: Conjuntos borrosos de la propiedad *profundidad relativa*

La propiedad *desviación* esta particionada en 3 conjuntos (figura 6.13): *Baja*, *Media* y *Alta*. Los puntos de cruce de estos conjuntos se seleccionan en función de la anchura del robot. El primer conjunto estará asociado con una anchura suficiente para que el robot pase sin peligro de colisión. El siguiente, con posiciones que podrían obstaculizar en ligeros cambios de trayectoria y el último con zonas de paso poco probables en el camino hacia el objetivo.

La propiedad *altura* de regiones permitirá descartar o incluir regiones durante el proceso de selección, pero no constituirá un elemento de ordenación. Por este motivo, dicha propiedad es particionada a través de dos únicos conjuntos cuyo punto de cruce debe estar relacionado con la altura del robot. Para evitar que regiones grandes, de altura similar a la del robot, sean descartadas como obstáculos, el punto de cruce de los conjuntos superará la altura del robot en un determinado valor.

Una vez que hemos establecido las particiones, debemos definir el sistema de reglas, teniendo en cuenta que debe existir una regla por cada combinación de hipótesis que puede formarse a partir de los conjuntos borrosos de las entradas. Para nuestro ejemplo, ob-

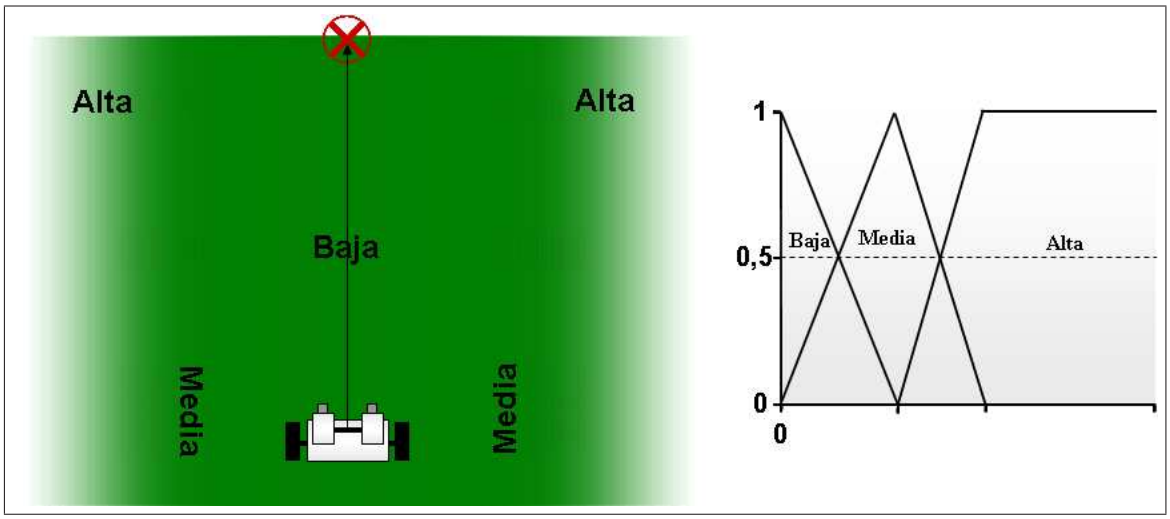



Figura 6.13: Conjuntos borrosos de la propiedad *desviación*

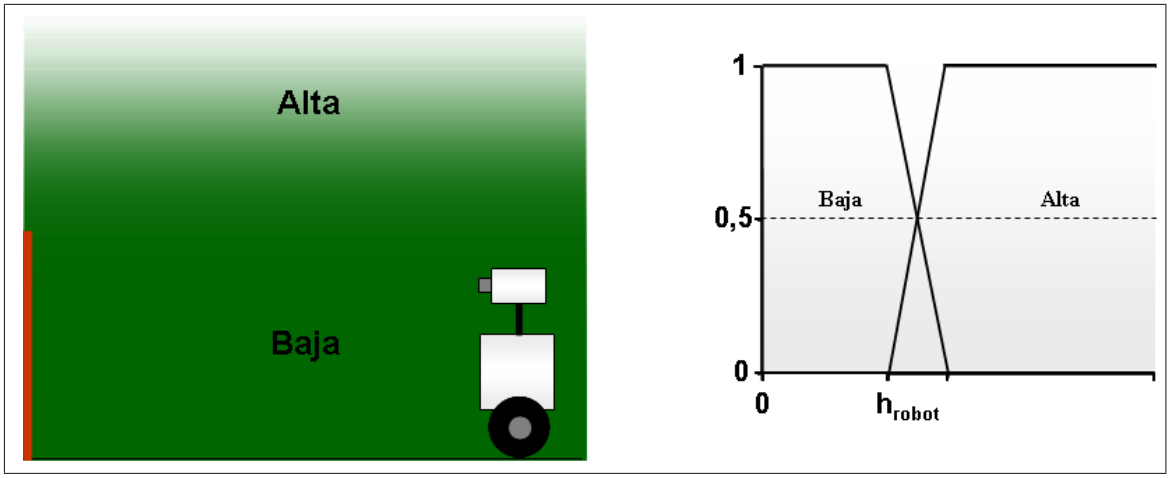



Figura 6.14: Conjuntos borrosos de la propiedad *altura*

tendríamos un sistema con 18 reglas, una por cada combinación de hipótesis que puede formarse con los conjuntos de $Al$, $Dv$ y $Pr$. La tabla 6.1 muestra una representación simplificada de este sistema de reglas al completo. Cada celda interior se corresponde con una regla cuyo antecedente se obtiene de la composición de la hipótesis sobre $Al$ indicada en la columna superior, la hipótesis sobre $Dv$ indicada en la columna inferior y la hipótesis sobre $Pr$ indicada en la fila correspondiente. En el interior de cada celda se anota el valor que toma la salida de dicha regla. Por ejemplo, la primera columna de celdas interiores se corresponde con las reglas:

SI $Al$ es $Baja$ Y $Dv$ es $Baja$ Y $Pr$ es $CercaR$ ENTONCES $saliencia = s_{111}$
SI $Al$ es $Baja$ Y $Dv$ es $Baja$ Y $Pr$ es $CercaO$ ENTONCES $saliencia = s_{112}$
SI $Al$ es $Baja$ Y $Dv$ es $Baja$ Y $Pr$ es $Lejos$ ENTONCES $saliencia = s_{113}$

| | | ***Al*** | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Baja* | | | *Alta* | | |
| ***Pr*** | *CercaR* | $s_{111}$ | $s_{121}$ | $s_{131}$ | $s_{211}$ | $s_{221}$ | $s_{231}$ |
| | *CercaO* | $s_{112}$ | $s_{122}$ | $s_{132}$ | $s_{212}$ | $s_{222}$ | $s_{232}$ |
| | *Lejos* | $s_{113}$ | $s_{123}$ | $s_{133}$ | $s_{213}$ | $s_{223}$ | $s_{233}$ |
| | | *Baja* | *Media* | *Alta* | *Baja* | *Media* | *Alta* |
| | | ***Dv*** | | | | | |

Tabla 6.1: Representación tabular del sistema de reglas de un selector de obstáculos

Una vez definido el sistema de reglas, hay que identificar qué reglas son de exclusión y cuáles de selección. Para ello, podemos hacer las siguientes consideraciones:

- Si la *altura* de una región es *Alta*, dicha región no es un obstáculo, puesto que no interfiere en la trayectoria hacia el objetivo.
- Si la *desviación* de una región es *Alta*, no es necesario incluir la región en la selección del foco de atención, puesto que, a menos que se produzca un cambio brusco de trayectoria, el robot no necesitará pasar cerca de la región para alcanzar el objetivo. Si dicho cambio de trayectoria se produjera, la desviación de la región disminuiría, por lo que, en caso de cercanía, la región no sería excluida del proceso de selección.
- Si la *profundidad relativa* de una región es *Lejos*, la región se encuentra situada detrás del objetivo, por lo que dicha región puede ser excluida.

Teniendo en cuenta los puntos anteriores, consideraremos de exclusión cualquier regla que contenga una hipótesis de la forma "$Al$ es $Alta$", "$Dv$ es $Alta$" o "$Pr$ es $Lejos$". La tabla 6.2 muestra el resultado de esta identificación. Las celdas con salidas en rojo se corresponden con reglas excluyentes y, por lo tanto, con valores negativos. Las restantes son reglas de selección cuyos valores de salida positivos deben imponer un orden de saliencia.

| | | ***Al*** | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Baja* | | | *Alta* | | |
| ***Pr*** | *CercaR* | $s_{111}$ | $s_{121}$ | $s_{131}$ | $s_{211}$ | $s_{221}$ | $s_{231}$ |
| | *CercaO* | $s_{112}$ | $s_{122}$ | $s_{132}$ | $s_{212}$ | $s_{222}$ | $s_{232}$ |
| | *Lejos* | $s_{113}$ | $s_{123}$ | $s_{133}$ | $s_{213}$ | $s_{223}$ | $s_{233}$ |
| | | *Baja* | *Media* | *Alta* | *Baja* | *Media* | *Alta* |
| | | ***Dv*** | | | | | |

Tabla 6.2: Identificación de reglas de exclusión y reglas de selección

Para concretar los valores de salida de cada regla, es necesario definir cómo afecta el cumplimiento de cada hipótesis a la prioridad en la selección. Las regiones de mayor importancia en la selección de obstáculos serán aquellas que se encuentren más cerca del robot. Esto implica que el valor de saliencia máximo debe ser asignado a las regiones que cumplan la hipótesis "*Al* es *Baja* Y *Dv* es *Baja* Y *Pr* es *CercaR*". Por otro lado, todas las regiones que se encuentren cerca de la trayectoria hacia el objetivo tendrán más importancia que aquellas que se desvíen de la línea de avance. Así, tendrán prioridad las que cumplan la hipótesis "*Dv* es *Baja*" frente a las que no la cumplan. Todas estás consideraciones permiten imponer un orden en las salidas de las reglas de selección, de manera que deberá cumplirse la relación $s_{111} > s_{112} > s_{121} > s_{122}$.

La relación anterior asegura el orden impuesto de saliencia de aquellas regiones del entorno que cumplan completamente las hipótesis del conjunto de reglas. Sin embargo, el comportamiento del sistema ante situaciones no explícitas por las reglas de selección es indeterminado si se realiza una asignación de salidas arbitrarias. Para obtener un rendimiento adecuado, es necesario realizar una serie de consideraciones adicionales sobre el funcionamiento del sistema, que expresaremos a través de reglas derivadas. En primer lugar, suponiendo el máximo cumplimiento de la hipótesis "*Al* es *Baja*", impondremos las siguientes condiciones:

$RD_1$: SI *Dv* es *Baja* (0.5) Y *Pr* es *CercaR* (1) ENTONCES $saliencia = s_{112}$
$RD_2$: SI *Dv* es *Baja* (0.5) Y *Pr* es *CercaO* (1) ENTONCES $saliencia = s_{121}$
$RD_3$: SI *Dv* es *Media* (1) Y *Pr* es *CercaO* (0.9) ENTONCES $saliencia = 0$
$RD_4$: SI *Dv* es *Media* (0.5) Y *Pr* es *CercaO* (1) ENTONCES $saliencia = s_{113} = 0$

Las reglas anteriores determinan el siguiente comportamiento:

- Las regiones entre el robot y el objetivo con la mínima desviación deben ser seleccionadas antes que las que presenten una desviación más *Media* que *Baja* (regla $RD_1$). Por el orden de salidas establecido anteriormente, esto se asegura imponiendo la misma saliencia a las regiones situadas en la posición del objetivo ($s_{112}$) que a aquellas que estén en línea con el robot con una desviación entre *Baja* y *Media*.
- Las regiones con desviación más *Baja* que *Media* tendrán prioridad frente a las que comiencen a presentar una desviación *Alta* (regla $RD_2$). Al igual que antes, el orden de salidas impuesto permite expresar esta condición asignando la misma prioridad a las regiones de desviación *Media* situadas a la profundidad del robot ($s_{121}$) que a las que se encuentren en línea con el objetivo con desviación entre *Baja* y *Media*.
- Las regiones, con desviación *Media*, que se alejen ligeramente en profundidad del objetivo y del robot, deberán ser excluidas del proceso (regla $RD_3$).
- A medida que las regiones se encuentren a una profundidad más cercana al objetivo que al robot, la desviación permitida debe ser menor. Las regiones que se encuentren en línea con el objetivo y que presenten una desviación más *Alta* que *Media* serán descartadas. Aquellas que estén en línea con el robot no serán excluidas a menos que su desviación sea completamente *Alta* (regla $RD_4$).

A partir de estas consideraciones, podemos obtener una serie de relaciones entre los consecuentes del sistema de reglas. Dichas relaciones vendrán dadas por las equivalencias entre el consecuente de cada regla derivada y la expresión de la salida del sistema obtenida a partir de los grados de pertenencia especificados en su antecedente:

$$\begin{aligned} RD_1 &: 0.5 * s_{111} + 0.5 * s_{121} = s_{112} \\ RD_2 &: 0.5 * s_{112} + 0.5 * s_{122} = s_{121} \\ RD_3 &: 0.9 * s_{122} + 0.1 * s_{123} = 0 \\ RD_4 &: 0.5 * s_{122} + 0.5 * s_{132} = s_{113} = 0 \end{aligned}$$

Para favorecer el agrupamiento de reglas, debemos intentar asignar el mismo valor de salida al mayor grupo de reglas excluyentes. Esto sera posible para aquellas reglas cuyos consecuentes no estén relacionados entre sí por las ecuaciones anteriores, que son las que cumplen las hipótesis "*Pr* es *Lejos*" ó "*Al* es *Alta*". Todas estás reglas tendrán como consecuente la salida $s_{123}$, determinada por el sistema de ecuaciones anterior.

Con todas estas consideraciones, suponiendo un rango de saliencia [M, m] para las reglas de selección, se obtienen los consecuentes de la tabla 6.3.

| | | ***Al*** | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Baja* | | | *Alta* | | |
| ***Pr*** | *CercaR* | $M$ | $(M+2m)/3$ | $0$ | $-9m$ | $-9m$ | $-9m$ |
| | *CercaO* | $(2M+m)/3$ | $m$ | $-m$ | $-9m$ | $-9m$ | $-9m$ |
| | *Lejos* | $-9m$ | $-9m$ | $-9m$ | $-9m$ | $-9m$ | $-9m$ |
| | | *Baja* | *Media* | *Alta* | *Baja* | *Media* | *Alta* |
| | | ***Dv*** | | | | | |

Tabla 6.3: Determinación de las salidas de reglas

La tabla 6.4 muestra una asignación de salidas coherente con las relaciones anteriores. Para obtener las salidas, se ha supuesto un valor de $M = 30$ y de $m = 6$.

| | | ***Al*** | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Baja* | | | *Alta* | | |
| ***Pr*** | *CercaR* | 30 | 14 | 0 | $-54$ | $-54$ | $-54$ |
| | *CercaO* | 22 | 6 | $-6$ | $-54$ | $-54$ | $-54$ |
| | *Lejos* | $-54$ | $-54$ | $-54$ | $-54$ | $-54$ | $-54$ |
| | | *Baja* | *Media* | *Alta* | *Baja* | *Media* | *Alta* |
| | | ***Dv*** | | | | | |

Tabla 6.4: Ejemplo de asignación de salidas

Con la asignación de salidas de la tabla 6.4, el sistema de reglas final del selector de obstáculos diseñado se expresaría como:

SI *Al* es *Baja* Y *Dv* es *Baja* Y *Pr* es *CercaR* ENTONCES *saliencia* = 30
SI *Al* es *Baja* Y *Dv* es *Baja* Y *Pr* es *CercaO* ENTONCES *saliencia* = 22
SI *Al* es *Baja* Y *Dv* es *Media* Y *Pr* es *CercaR* ENTONCES *saliencia* = 14
SI *Al* es *Baja* Y *Dv* es *Media* Y *Pr* es *CercaO* ENTONCES *saliencia* = 6
SI *Al* es *Baja* Y *Dv* es *Alta* Y *Pr* es *CercaR* ENTONCES *saliencia* = 0
SI *Al* es *Baja* Y *Dv* es *Alta* Y *Pr* es *CercaO* ENTONCES *saliencia* = −6
SI *Al* es *Alta* O *Pr* es *Lejos* ENTONCES *saliencia* = −54

#### 6.2.4.2. Inhibición de retorno

Cuando un selector de objetivo obtiene varios candidatos a focos de atención, debe incluirse un mecanismo que permita repartir el tiempo de atención entre todos ellos. Este mecanismo es la inhibición de retorno que se incluye, dentro de cada selector individual, como la última fase del proceso.

La elección de un foco de atención en cada instante depende de varios factores. En primer lugar, la saliencia de las regiones debe imponer un orden en la fijación, por lo que, las zonas de mayor saliencia deben ser las primeras en tomar el control de la atención. Asimismo, para que la atención se distribuya sobre las distintas zonas candidatas con independencia de su grado de saliencia, es necesario incluir el tiempo de fijación como factor de inhibición dentro del proceso de elección de un foco único, evitando así la fijación continuada sobre las regiones de máxima saliencia. Para tener en cuenta este segundo factor, se utiliza un mapa de inhibición que representa el grado de fijación de cada zona del campo visual.

Cuando una región es foveatizada, el mapa de inhibición se actualiza mediante una distribución de valores con la estructura espacial de una gaussiana centrada en el punto de fijación y con una desviación típica correspondiente al radio de la región (ecuación 6.42).

$$IMap_t(x,y) = IMap_{t-1}(x,y) + \frac{1}{2\pi\sigma_f} e^{-\frac{(x-x_f)^2+(y-y_f)^2}{2\sigma_f^2}} \tag{6.42}$$

Con esta actualización, en cada instante de fijación, aumenta la diferencia de inhibición entre la zona foveatizada y las zonas no atendidas, creciendo así la necesidad de fijar la atención sobre otra región del campo visual. Para evitar un cambio inmediato de atención, el mapa de inhibición sólo es examinado una vez transcurrido un tiempo de fijación que asegure la conclusión de un posible análisis sobre la zona. Tras dicho tiempo, se selecciona como nuevo foco de atención la región de máxima saliencia de entre las que tienen un valor de inhibición cercano al mínimo según un cierto umbral.

La figura 6.15 muestra varios instantes del proceso de selección mediante inhibición de retorno de un grupo de regiones (columna de la derecha). La saliencia de las regiones candidatas en esta situación decrece a medida que éstas se alejan de la posición central. Inicialmente, el mapa de inhibición (columna de la izquierda) contiene valores nulos para

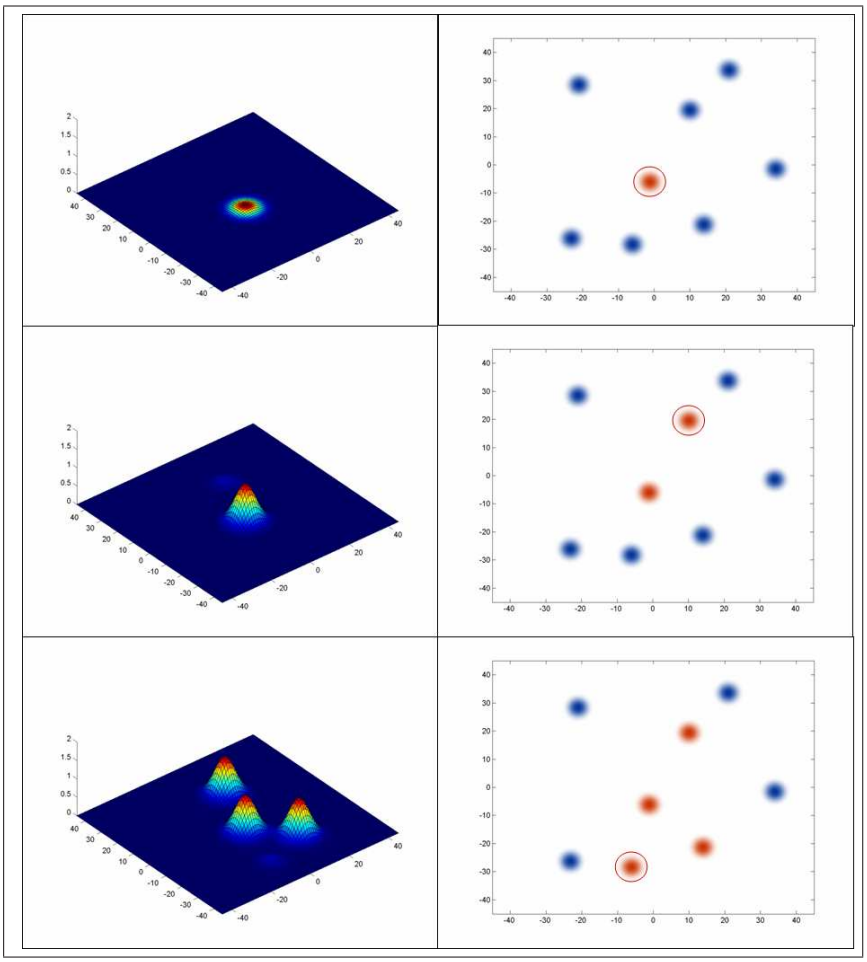


Figura 6.15: Selección de regiones mediante inhibición de retorno

todas las zonas del campo visual, por lo que la región seleccionada es la de máxima saliencia. Transcurridos varios instantes de fijación sobre la región seleccionada, el mapa de inhibición es examinado y la atención cambia a la siguiente región en saliencia. Este proceso se repite permitiendo la distribución de la atención entre las distintas regiones candidatas siguiendo su orden de saliencia.

A medida que las regiones candidatas acumulen más tiempo de fijación, la diferencia en inhibición con las zonas no atendidas se hará mayor. Esto puede provocar que, en instantes futuros, se requiera mayor tiempo de atención sobre una región no atendida previamente para igualar su inhibición con la del resto de regiones. Para evitar esta situación, cuando el valor de inhibición mínimo de todas las regiones sea mayor que un umbral $\mu_{min} < 1$, se

realizará la siguiente transformación sobre cada posición del mapa:

$$IMap(x,y) = \frac{IMap(x,y)}{minR} * \mu_{min} \tag{6.43}$$

siendo $minR$ el mínimo valor de inhibición de todas las regiones.

La expresión 6.43 es equivalente a una puesta a cero progresiva del mapa de inhibición. En una situación estática en la que el grupo de regiones candidatas se mantenga estable en cuanto a número y posición, sería posible utilizar una puesta a cero inmediata tomando como criterio de inicialización la proximidad en el valor de inhibición de todas las regiones. Sin embargo, en situaciones reales, las regiones pueden variar ligeramente su posición y, además, aquellas que se encuentren cerca de los umbrales mínimos de selección podrían ser excluidas e incluidas de nuevo en determinados instantes del proceso, variando así el total de regiones que se deben tener en cuenta en la inicialización del mapa. En estas situaciones, una puesta a cero inmediata podría provocar la inicialización prematura del mapa, ocasionando que ciertas regiones no sean atendidas en ningún momento.

La figura 6.16 muestra el efecto de aplicar el ajuste de la expresión 6.43 sobre el mapa de inhibición del ejemplo. Dicho ajuste tiene lugar durante el incremento de inhibición de la región de mínima saliencia, una vez que las restantes regiones ya han sido atendidas. A medida que la inhibición sobre esta zona se hace mayor, las restantes pierden inhibición hasta anularse prácticamente. Una vez que cumple el tiempo de fijación sobre el foco de atención, todas las regiones, exceptuando la última, presentan una inhibición mínima, por lo que el ciclo de selección vuelve a empezar, comenzando de nuevo con la región de máxima saliencia.

#### 6.2.4.3. Proceso de selección global

El funcionamiento simultáneo de varios selectores de objetivo obliga a tener un selector global que decida qué foco de atención individual será finalmente atendido.

Cada selector de objetivo debe modularse con dos parámetros que denominaremos tiempo de concentración y nivel alerta. El tiempo de concentración es el tiempo durante el cual el selector mantiene el control de atención. Este valor es preservado por el selector global, otorgando el control de atención al mismo selector hasta que se supera el tiempo asociado con este parámetro. El nivel de alerta se refiere al grado en el que el selector debe mantener

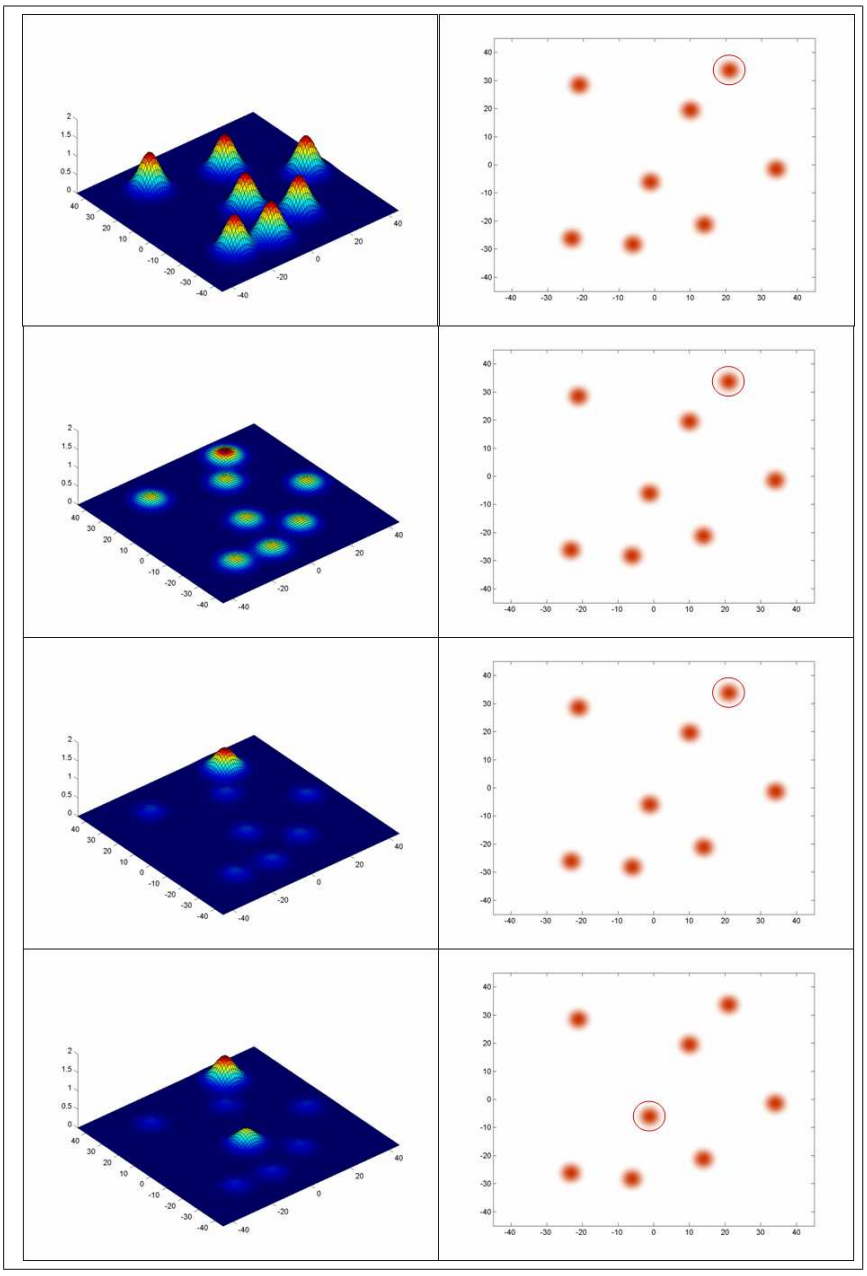


Figura 6.16: Reanudación del ciclo de selección

un estado de vigilancia sobre las regiones de la escena, imponiendo, en cierto modo, una prioridad en la toma de control global. Dicho parámetro es modelado como la relación entre el tiempo de concentración y el periodo del ciclo de activación.

La figura 6.17 muestra el funcionamiento ideal de 3 selectores de objetivo parametri-

zados con diferentes valores de alerta y concentración. El selector 1 se activa cada 3 u.t. (unidades de tiempo) manteniendo el control de atención durante 1 u.t., lo que le proporciona un nivel de alerta de 1/3 ; el selector 2 tiene asociado un mayor nivel de alerta (1/2), con un tiempo de concentración de 1 u.t.; por último, el selector 3 presenta un valor de alerta de 1/2 y un tiempo de concentración de 2 u.t. Los valores de tiempo asociados con estos tres selectores provocan que, en determinados instantes, se produzca una petición simultánea en la toma de control de atención, debiendo así realizar un reparto de tiempos que se ajuste lo máximo posible al funcionamiento ideal.

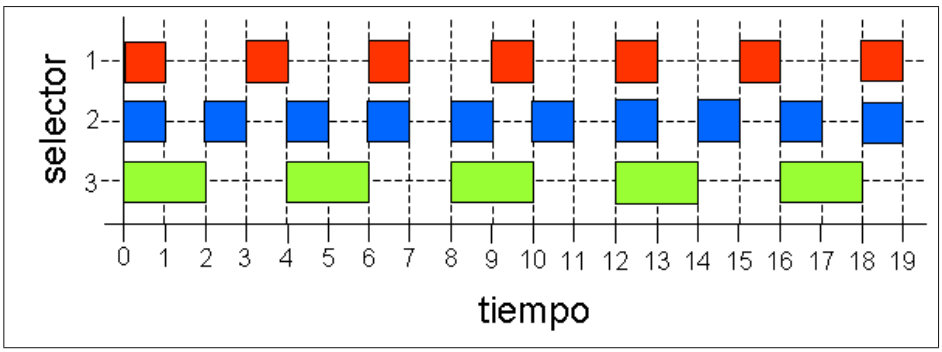



Figura 6.17: Tiempos ideales de activación de cada selector de objetivo

Una vez finalizado el proceso de selección individual, cada selector envía el foco de atención resultante al selector global. El selector global mantiene una marca de tiempo de cada selector individual que indica en qué momento debe cederse el control a dicho selector de acuerdo con su periodo de activación. Por ejemplo, si el periodo de activación de un selector es de 3, su marca se actualiza con el valor del instante en el que tomó el control más 3. La marca de tiempo sólo es actualizada tras la activación del selector, una vez que transcurre el tiempo asociado con el parámetro de concentración, lo que permite mantener marcas de tiempo bajas para aquellos selectores que llevan mucho tiempo sin tomar el control. En cada instante, el selector global analiza las marcas de tiempo de cada selector y escoge el foco de atención de aquel que tiene una marca de tiempo más antigua. Si existen varios selectores con la misma marca de tiempo, se cederá el control al que tenga un menor valor de concentración para que el perjuicio a los restantes selectores sea el mínimo posible. Si hay más de un selector con el mínimo valor de concentración, tendrá prioridad el que tenga un mayor nivel de alerta.

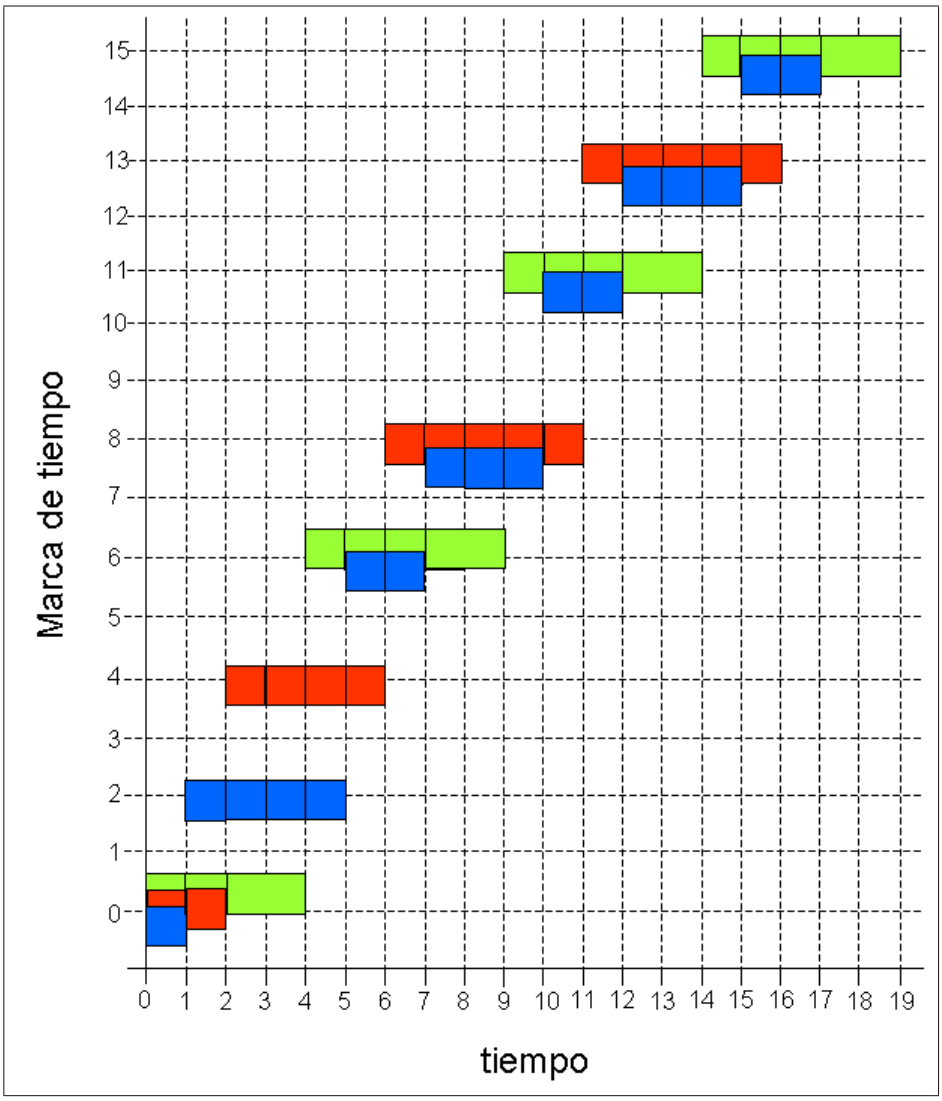



Figura 6.18: Marcas de tiempo de cada selector durante el proceso de selección

La figura 6.18 muestra las marcas de tiempo de cada selector del ejemplo durante varios instantes del proceso de selección global. Inicialmente, todos parten de una marca de 0, por lo que el control se cede al selector 2 que es el que presenta un menor tiempo de concentración y un mayor nivel de alerta. Este selector mantiene el control durante 1 u.t., momento en el cual actualiza su marca de tiempo a 2 u.t., esto es, el valor del instante de activación (0 u.t.) más el periodo de activación (2 u.t.). En el instante 1, toma el control el selector 1, tras lo cual coloca su marca de tiempo a 4 u.t. (instante de activación=1, periodo=3). En el instante 2, el selector 3 es el que mantiene la marca de tiempo más

antigua, por lo que toma el control durante su tiempo de activación (2 u.t.) y actualiza su marca a 6 u.t. (instante de activación=2, periodo=4). En el instante 4, la marca de tiempo más antigua es la del selector 2, por lo que éste es el que toma el control, actualizando dicha marca a 6 u.t. Lo mismo ocurre en el instante siguiente con el selector 1. En el instante 6, los selectores 2 y 3 presentan la misma marca de tiempo. Para que la penalización sea la menor posible, el control se cede al selector 2, que es el de menor tiempo de concentración, y, a continuación, al selector 3. Siguiendo este procedimiento de asignación del control, los 3 selectores ajustarían sus tiempos de activación consiguiendo un funcionamiento real lo más aproximado posible al funcionamiento ideal (figura 6.19).

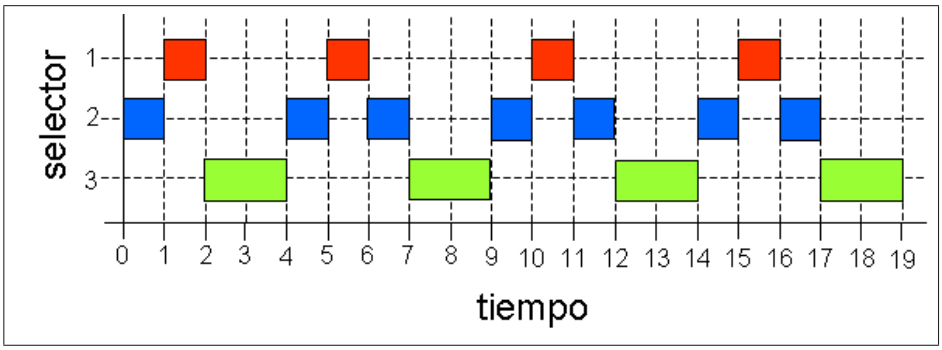



Figura 6.19: Tiempos reales de activación de cada selector de objetivo

### 6.2.5. Control de seguimiento del foco de atención

Una vez seleccionado el foco de atención, éste es fijado y mantenido hasta la selección del siguiente. Esta tarea es realizada por el controlador de seguimiento, encargado de localizar la región objetivo en la secuencia de imágenes capturadas a lo largo del tiempo por la cámara de control de atención.

El seguimiento de un objetivo visual requiere un proceso de búsqueda de correspondencias entre la región objetivo y la imagen obtenida de la captura actual. Puesto que las regiones son elementos distintivos de una imagen, su localización puede resolverse de manera adecuada mediante métodos basados en correlación. Como en un problema clásico de correspondencias entre características de imagen, el mantenimiento del foco de atención consistirá en la localización de la ventana de la nueva imagen que maximiza el coeficiente de correlación sobre la región objetivo. Ahora bien, si el único criterio es la similitud entre

ventanas de imagen, en muchos casos pueden aparecer falsos positivos que hay que descartar. Si existen varios máximos de valores próximos, es necesario utilizar algún criterio adicional que permita obtener una única región de máxima similitud con el objetivo (Guerra, 2002).

Cuando el contenido de una región no es lo suficientemente discriminatorio, la información sobre su entorno próximo puede ayudar a obtener una única zona de imagen coherente con la región objetivo. El análisis del entorno de una región puede llevarse a cabo aumentando el tamaño de la ventana de correlación y realizando así una comparación con zonas mayores del campo visual. Haciendo uso de la estructura multi-resolución obtenida por el procesador visual, este proceso es similar a una comparación en niveles de menor resolución manteniendo el tamaño de la ventana de correlación. Con esta idea, el seguimiento del foco de atención consiste en una localización ascendente de la región objetivo en el espacio de escala. El objetivo estará representado por un prisma multi-escala que contendrá información sobre la región de seguimiento y su entorno próximo. El proceso parte del nivel de mayor resolución y, a través de una búsqueda por correlación sobre la ventana asociada con la región objetivo, realiza una preselección de las zonas de imagen de mayor similitud. Cuando la ratio entre la correlación de una zona preseleccionada y el valor máximo de correlación sea cercano a 1, el grado de similitud obtenido no será suficiente para aceptar o rechazar dicha zona como nuevo punto de fijación. En este caso, es necesario realizar una nueva comparación en el siguiente nivel de resolución. Sin embargo, si la ratio no fuera lo suficientemente alta, es posible descartar la zona como región objetivo, puesto que, al menos, existe otra cuyo grado de similitud la convierte en mejor candidata. Este procedimiento se repite para cada nivel hasta que sólo exista una zona de máxima similitud o se alcance el último nivel de la estructura.

El algoritmo 1 muestra este procedimiento de seguimiento de la región objetivo. Los datos de entrada y salida del algoritmo son los indicados en la tabla 6.5. Inicialmente, se realiza el cálculo de correlación entre la región objetivo ($prismaR[0]$) y la imagen de mayor resolución ($piramide[0]$). Este cálculo se obtiene a partir de la función $cCorrel$, indicando, por cada imagen de entrada, la región de interés sobre la que debe realizarse la operación, mediante su posición central y sus dimensiones. La función devuelve una imagen ($iCorrel$) en la que cada píxel almacena el valor de correlación calculado entre la primera imagen y la ventana correspondiente de la segunda, centrada en dicha posición. A partir de esta

| **Algoritmo 1** Control de seguimiento del foco de atención | |
|---|---|
| **Entrada** | $prismaR$ = Prisma multi-resolución de la región objetivo |
| | $piramide$ = Pirámide multi-resolución obtenida de la captura actual |
| | $w_E$ = Ancho de cada imagen del prisma |
| | $h_E$ = Alto de cada imagen del prisma |
| | $W$ = Ancho de la imagen de mayor resolución |
| | $H$ = Alto de la imagen de mayor resolución |
| | $s$ = Factor de escala de la estructura multi-resolución |
| | $nLevels$ = Número de niveles de la estructura multi-resolución |
| | $\mu Correl$ = Umbral mínimo de correlación en la detección de máximos |
| | $\mu Ratio$ = Mínima ratio de correlación del proceso de selección |
| **Salida** | $(xTrack, yTrack)$ = Posición de la región objetivo en la captura actual |
| | $tracked$ = Indicador de éxito en el proceso de localización |

Tabla 6.5: Entrada y salida del algoritmo de seguimiento del foco de atención

imagen de correlación, se realiza una selección de aquellos puntos que mantienen un valor máximo en su entorno local y mayor que el umbral mínimo considerado para esta fase ($\mu Correl$). La función $LocalMaximaTh$ se encarga de esta tarea, devolviendo la lista de puntos que superan esta criba y el máximo global de correlación ($maxCorrel$). La lista obtenida contiene el conjunto de posiciones de mayor similitud con la región objetivo. Si hay más de una posición ($listMax.length() > 1$), comienza la segunda fase del proceso, en la que se realiza un recorrido ascendente en la estructura multi-resolución para intentar descartar posiciones que no concuerden, de acuerdo a su entorno, con la región objetivo. Este recorrido se mantendrá hasta que se encuentre la posición de máxima similitud, mientras existan niveles sobre los que realizar el análisis. El primer paso de esta fase es eliminar de la lista los puntos cuya ratio de correlación sea menor que el mínimo especificado ($\mu Ratio$). Para los restantes, debe obtenerse el valor de correlación entre su entorno y la imagen de la región objetivo en el siguiente nivel de la estructura ($sCorrel$). Dentro de este mismo paso, se obtiene además el máximo de correlación de todos los puntos para poder repetir el análisis de la ratio en la siguiente iteración del proceso. Una vez finalizada esta fase, se comprueba el número de posiciones contenidas en la lista de candidatos. Si sólo existe una, ésta será la nueva posición de la región objetivo ($xTracked, yTracked$). Si no, el proceso se repetirá hasta lograr la localización de la región o hasta que se reciba una nueva petición de seguimiento de otra región objetivo.

**Algoritmo 1** Control de seguimiento del foco de atención

$iCorrel \Leftarrow cCorrel(prismaR[0], w_E/2, h_E/2, w_E, h_E, piramide[0], W/2, H/2, W, H)$
$listMax \Leftarrow LocalMaximaTh(iCorrel, W, H, \mu Correl, maxCorrel)$
$l = 0$
**while** $listMax.length() > 1 \wedge l < nLevels$ **do**
  **for** $p = 0$ **to** $listMax.length() - 1$ **do**
    **if** $listMax[p].correl/maxCorrel < \mu Ratio$ **then**
      $listMax.delete(p)$
    **else**
      $p \Leftarrow p + 1$
    **end if**
  **end for**
  $l \Leftarrow l + 1$
  **if** $l < nLevels$ **then**
    $maxCorrel \Leftarrow 0$
    **for** $p = 0$ **to** $listMax.length() - 1$ **do**
      $x_C \Leftarrow (listMax[p].x) * s^l$
      $y_C \Leftarrow (listMax[p].y) * s^l$
      $correl \Leftarrow scCorrel(prismaR[l], w_E/2, h_E/2, w_E, h_E, piramide[l], x_C, y_C, w_E, h_E)$
      $listMax[p].correl \Leftarrow correl$
      **if** $correl > maxCorrel$ **then**
        $maxCorrel \Leftarrow correl$
      **end if**
    **end for**
  **end if**
**end while**
**if** $listMax.length() = 1$ **then**
  $tracked \Leftarrow TRUE$
  $xTracked \Leftarrow listMax[0].x$
  $yTracked \Leftarrow listMax[0].y$
**else**
  $tracked \Leftarrow FALSE$
**end if**

### 6.2.6. Control de vergencia

La función de los movimientos de vergencia en un sistema estéreo es la foveatización de un mismo objetivo visual en ambas cámaras. A partir de dichos movimientos, es posible fijar regiones de interés del espacio que varían en distancia con respecto al observador.

Un sistema de control de vergencia debe proporcionar una fijación binocular estable y

una respuesta suave y precisa a cambios del entorno. El diseño de una estrategia de control para un sistema de este tipo debe contemplar estos factores a la hora de definir las señales de entrada que guiarán su comportamiento. La fijación binocular puede cuantificarse utilizando una medida de similitud de las imágenes capturadas por ambas cámaras. Para obtener esta cuantificación del grado de fijación se ha optado por utilizar el coeficiente de correlación normalizada, al igual que en el proceso de seguimiento del foco de atención.

Partiendo de una cámara fija (en nuestro caso, la cámara de control de atención), el control de vergencia debe calcular el desplazamiento de la otra cámara que permita mantener centrado en la imagen el mismo objetivo visual. Con estas premisas, el control de vergencia puede tratarse como un problema de maximización del coeficiente de correlación entre la ventana central de imagen de la cámara fija y la franja epipolar homóloga de la otra imagen. Ahora bien, el tamaño adecuado de la ventana central es dependiente de las propiedades del mundo visual, lo que hace necesario controlar un nuevo parámetro: el tamaño de la ventana de correlación. Para solucionar este problema, la estrategia de control empleada hace uso de la estructura multi-escala obtenida por el procesador visual. Este esquema multi-escala permite extraer información acerca de la fijación binocular en diferentes extensiones del espacio visual. Admite la posibilidad de contrastar los valores de correlación de los distintos niveles de resolución, proporcionando un mecanismo para estimar de una forma precisa la posición correcta de vergencia (Bachiller et al., 2003).

En función de su posición en la estructura, cada nivel tiene asociado una función de correlación que se caracteriza por lo siguiente (figura 6.20):

- Los niveles de menor resolución son menos sensibles a pequeños cambios en la posición de vergencia, proporcionando valores de correlación más suaves que los niveles de mayor resolución. Esto permite evitar los máximos locales que pueden aparecer en la función de correlación de los niveles más bajos de la estructura.
- Los niveles de mayor resolución presentan valores más altos en los máximos globales de la función de correlación que los de menor escala. Por lo tanto, proporcionan información más precisa sobre la posición correcta de vergencia.

De acuerdo con esta observación, el control se lleva a cabo siguiendo una estrategia jerárquica en la que cada nivel actúa como selector de la ventana de imagen del siguiente con mayor probabilidad de contener la posición correcta de vergencia. Este proceso de

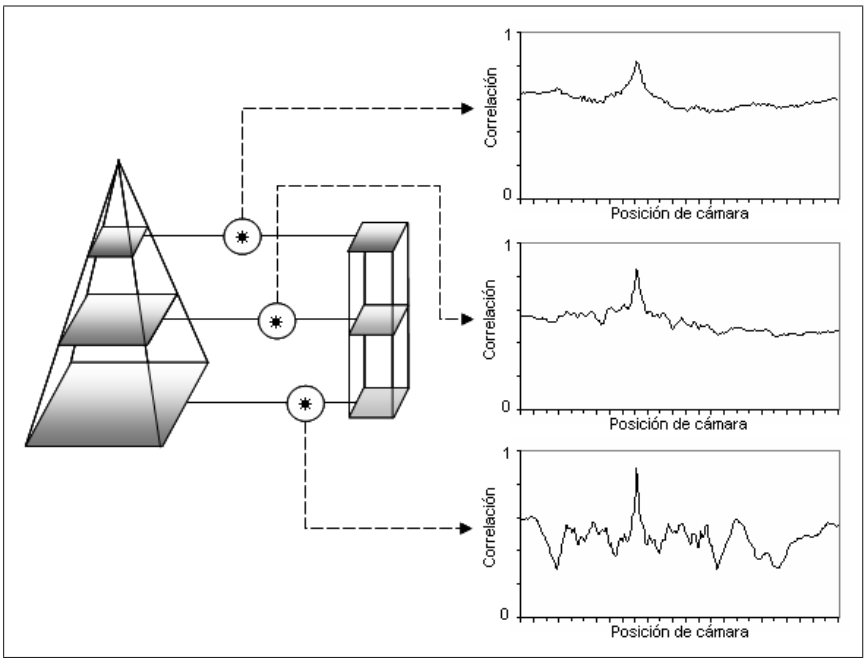



Figura 6.20: Función de correlación en distintos niveles de la estructura multi-escala

búsqueda es el inverso al utilizado en el seguimiento del foco de atención. El principal motivo para realizar una localización descendente en el espacio de escala, en lugar de una búsqueda ascendente como en el proceso anterior, es que, en el control de vergencia, no se supone conocimiento alguno sobre la extensión de la zona a la que ambas cámaras deben converger, por lo que no es posible realizar una búsqueda fiable partiendo del nivel de mayor resolución. Si se hiciera así, en los casos en los que la ventana de fóvea fuera demasiado reducida, la búsqueda ascendente podría no llegar a resolver el desempate entre varias posiciones candidatas, o bien, si la ventana fuera demasiado extensa, el nivel inferior no podría seleccionar ninguna posición inicialmente. Así, aunque una búsqueda descendente puede resultar más compleja computacionalmente, puesto que requiere un recorrido completo de toda la estructura, resulta más adecuada en este caso.

El método de localización jerárquica que se plantea en nuestro sistema de control de vergencia presenta ciertas similitudes con otros métodos de búsqueda de patrones basados en estructuras multi-resolución (Zhang y Wu, 2001). En las propuestas existentes, el objetivo viene dado por una representación piramidal que almacena la imagen del patrón de búsqueda en los distintos niveles de resolución. La localización piramidal, por lo tanto,

constituye un proceso de refinamiento en el que cada nivel permite ajustar de manera más precisa la posición del objetivo obtenida por el nivel anterior. En nuestro caso, el objetivo tiene dimensiones desconocidas y es representado por un prisma multiescala que permite considerar las posibles extensiones de la zona de vergencia. El recorrido jerárquico del método propuesto permite determinar la extensión de dicha zona y ajustar su posición en función de las posibilidades. Cada nivel debe mejorar la estimación del anterior, por lo que no todos los niveles de la estructura intervienen en el resultado final.

Partiendo del nivel de menor resolución, en el que la incertidumbre es máxima, la búsqueda se realiza en una ventana de imagen de máxima anchura. Esta búsqueda consiste en la localización de la posición de la ventana donde la función de correlación, entre un entorno de dicha posición y un entorno central de la imagen obtenida por la cámara fija, proporciona un valor máximo. Si se encuentra un máximo que supera un cierto umbral, se considera que la posición correcta de vergencia estará situada en un entorno cercano a la posición del máximo. Esto permite definir la ventana de imagen del siguiente nivel donde se debe realizar la búsqueda. Tras fijar la posición central y las dimensiones de la ventana, el proceso de búsqueda se repite sobre cada nivel, hasta concretar la posición de vergencia en la imagen de mayor resolución. Si, en algún nivel, el máximo obtenido no es lo suficientemente alto, el resultado de dicho nivel no se tendrá en cuenta en el siguiente; es decir, el cálculo de la ventana de búsqueda se realizará a partir de las especificaciones del último nivel válido. Tampoco se utilizará el resultado de aquellos niveles que no aumenten el valor máximo de correlación obtenido hasta el momento. Esto favorece el control global en situaciones en las que la falta de textura de las zonas de mayor resolución no permite mejorar la estimación realizada por los niveles de menor resolución.

El algoritmo 2 muestra los pasos de este método de control de vergencia. En la tabla 6.6 se describen los datos de entrada y salida utilizados en él. La ventana de imagen central de la cámara fija está representada en los distintos niveles de resolución a través de un prisma multi-escala ($prismaF$). De manera similar, la imagen completa de la cámara de control de vergencia viene dada por la correspondiente pirámide multi-resolución ($piramideV$). El control se lleva a cabo a través de un proceso iterativo que recorre estas dos estructuras multi-escala de manera descendente. Cada iteración constituye un paso de búsqueda por nivel de la posición de la cámara móvil que presenta la mayor correspondencia con la fóvea de la cámara fija. Esta búsqueda consiste en la localización de la posición que

| **Algoritmo 2** Control de vergencia | |
|---|---|
| **Entrada** | $prismaF$ = Prisma central de la cámara fija de tamaño $rxr$ |
| | $piramideV$ = Pirámide multi-resolución de la cámara móvil |
| | $wEnv$ = Ancho de la ventana de búsqueda en caso de éxito |
| | $hWin$ = Alto de la ventana de búsqueda |
| | $w$ = Ancho de la imagen de menor resolución |
| | $h$ = Alto de la imagen de menor resolución |
| | $s$ = Factor de escala de la estructura multi-resolución |
| | $nLevels$ = Número de niveles de la estructura multi-resolución |
| | $\mu Correl$ = Umbral mínimo de correlación en la detección de máximos |
| **Salida** | $(xVerg, yVerg)$ = Posición de vergencia de la cámara móvil |
| | $maxCorrel$ = Máximo valor de correlación en toda la estructura |

Tabla 6.6: Entrada y salida del algoritmo de control de vergencia

**Algoritmo 2** Control de vergencia

$hS \Leftarrow hWin$
$wS \Leftarrow w$
$xS \Leftarrow w/2$
$yS \Leftarrow h/2$
$maxCorrel \Leftarrow 0$
**for** $l = nLevels - 1$ **downto** $0$ **do**
  $iCorrel \Leftarrow cCorrel(prismaF[l], r/2, r/2, r, r, piramideV[l], xS, yS, wS, hS)$
  $correlLevel \Leftarrow maximumXY(iCorrel, xS, yS, wS, hS, xMax, yMax)$
  **if** $correlLevel > \mu Correl \wedge correlLevel > maxCorrel$ **then**
    $maxCorrel \Leftarrow correlLevel$
    $xS \Leftarrow xMax/s$
    $yS \Leftarrow yMax/s$
    $wS \Leftarrow wEnv$
  **else**
    $xS \Leftarrow xS/s$
    $yS \Leftarrow yS/s$
    $wS \Leftarrow wS/s$
  **end if**
**end for**
$xVerg \Leftarrow xS * s$
$yVerg \Leftarrow yS * s$

maximiza el coeficiente de correlación entre las imágenes del prisma y la pirámide del nivel que corresponda (función $maximumXY$). En cada nivel, la búsqueda se realiza sobre una ventana de la imagen de la cámara móvil. Dicha ventana vendrá determinada por

su posición central $(xS, yS)$, el alto $(hS)$ y el ancho$(wS)$. El alto se mantiene constante $(hWin)$ para conservar la proporción entre las dimensiones de la fóvea y el alto de la zona de búsqueda, con independencia del nivel. La posición central de la ventana será aquella para la que se haya obtenido el valor máximo de correlación y superior al umbral $\mu Correl$ en algún nivel anterior. Esta posición es transformada a la escala actual a través del factor $s$. El ancho de la ventana representa un posible error de desplazamiento de la posición de vergencia determinada por el nivel anterior. Es actualizado, en caso de éxito en la localización de cada nivel, a un ancho mínimo de búsqueda $(wEnv)$, o bien, si en la iteración actual no se encontró una posición válida, al ancho de búsqueda actual en el siguiente nivel de resolución $(wS/s)$. Siguiendo este procedimiento, en cada iteración del proceso, se determina cuál es la zona de imagen donde es más probable localizar la posición correcta de vergencia. Una vez recorridos todos los niveles, la posición válida $(xVerg, yVerg)$ será la que haya determinado el nivel de mayor resolución. Puesto que es posible que dicha localización no tenga éxito, el algoritmo retorna también el máximo valor de correlación obtenido al alcanzar el último nivel, para poder así aceptar o desechar la posición calculada.

Cuando existe un cambio brusco de atención, la imagen asociada con la cámara de control de vergencia puede no ser coherente con la situación actual. En estas circunstancias, el procedimiento anterior no permitirá localizar una posición correcta de vergencia y será necesario poner en marcha un proceso de exploración dinámica a través de movimientos de giro sobre el eje vertical de la cámara que permitan ampliar la zona de búsqueda del objetivo. Esta acción constituye una forma de localización dinámica que elimina las limitaciones asociadas con la extensión del campo visual.

## 6.3. Control basado en la atención

El control inteligente del robot se lleva a cabo a través de las relaciones establecidas entre el conjunto de posibles comportamientos de alto nivel y el sistema visual atencional. Dentro de esta arquitectura, un comportamiento se define como una unidad ejecutora de una función, cuyo resultado puede ser una acción motora o interna, y que se encuentra enlazado con el sistema atencional, modulándolo de manera específica para cumplir sus objetivos. Los comportamientos pueden mantener además conexiones con otros sistemas sensoriales que complementen la información procedente del sistema visual (información exteroceptiva) o simplemente informen sobre el estado del robot (información propiocep-

tiva e interoceptiva).

Cada comportamiento de alto nivel pone en marcha un mecanismo de control de atención coherente con sus objetivos. Esto es, activa y modula un selector de objetivo del sistema de atención que le proporcione la secuencia apropiada de información visual para cumplir y mantener sus objetivos. Cuando un selector de objetivo toma el control motor, el foco de atención seleccionado es enviado a las distintos comportamientos, permitiendo que aquellos compatibles con la región visual foveatizada realicen el procesamiento que corresponda y generen la acción adecuada. Las posibles acciones pueden ser directas o indirectas. Las acciones directas son comandos de movimiento enviados al sistema de control motor. Las indirectas se traducen en la activación de otros comportamientos que permitan resolver la situación. La activación de comportamientos puede entenderse como un proceso jerárquico, aunque, durante su funcionamiento, cada comportamiento responde de manera independiente a la información recibida.

Desde el punto de vista de la actuación, las relaciones entre los componentes de control de atención y los comportamientos en ejecución proporcionan un medio de serialización de las acciones que resuelven una determinada situación. El sistema de atención asegura en todo momento la selección de un objetivo que actúa como única entrada visual en todos los comportamientos de alto nivel. Esta selección sensorial se traduce en el disparo de un subconjunto de acciones posibles de entre el total de acciones que pueden llevarse a cabo por el grupo de comportamientos en ejecución. Partiendo de que el conjunto de comportamientos mantienen un control cooperativo en la consecución de un objetivo global, la secuencia de acciones producida por la serialización de información visual proporciona una posible vía para alcanzar el resultado deseado. Ahora bien, la solución obtenida no tiene porqué ser la más eficiente, tomando como medida de eficiencia el tiempo requerido para alcanzar el objetivo o el número de acciones necesarias. Esto dependerá de que la modulación de tiempos de los selectores de objetivo, realizada por los comportamientos correspondientes, sea la más adecuada. No obstante, no parece posible realizar a priori una asignación de tiempos que resuelva cualquier situación de la manera más eficiente. La opción más plausible sería dotar al sistema de una cierta capacidad de aprendizaje que le permitiera ajustar los tiempos de activación de cada selector de acuerdo con la experiencia. Esta opción no ha sido probada y se plantea como una ampliación del sistema, a desarrollar como posible evolución de esta tesis.

El sistema visual propuesto modela distintos tipos de atención, proporcionando cierta flexibilidad a la hora de definir las posibles formas de control de alto nivel:

- Atención *bottom-up*: la selección de un foco de atención se lleva a cabo a partir de la observación de las propiedades de las regiones presentes en una escena. Si alguna región mantiene características coherentes con los criterios de control de los selectores activos, la región es focalizada. Desde este punto de vista, puede hablarse de una atención *bottom-up* en el sentido de que el sistema proporciona una respuesta atencional ante ciertos estímulos que, aunque no se esperen, pueden presentarse en cualquier momento.

- Atención *top-down*: el sistema de atención es modulado desde los procesos de alto nivel mediante la activación de determinados selectores de objetivo acordes con la situación actual. Desde esta perspectiva, la selección de un foco de atención está influida por el contexto de actuación a través de un esquema atencional guiado por objetivo.

- Atención abierta: el sistema visual asegura en todo momento la selección de un foco de atención que actúa como única fuente de información visual. Este proceso selectivo proporciona una respuesta inmediata al estímulo atendido por parte de los procesos de alto nivel, obviando la información irrelevante para la tarea en curso.

- Atención encubierta: la coexistencia de múltiples selectores de objetivo permite mantener simultáneamente una focalización "mental" sobre varios estímulos. Esto implica que, aunque sólo se pueda fijar atención abierta sobre un único objetivo visual en un instante dado, es posible mantener un estado de alerta en el sistema que permite, en cualquier momento, desplazar la atención, no sólo hacia otra zona del espacio visual, sino hacia objetivos con diferentes propiedades y para diferentes propósitos conductuales.

Para tratar de esclarecer los distintos aspectos expuestos anteriormente, se presenta a continuación un ejemplo de un sistema de control compuesto por varios comportamientos y selectores de objetivo que permiten resolver un problema de navegación siguiendo un grupo de balizas. El sistema global viene descrito por los componentes y relaciones que se muestran en la figura 6.21. Desde los objetivos a la dinámica de atención y acción, cada comportamiento se define de la siguiente forma:


Figura 6.21: Componentes de control en una tarea de navegación con balizas

- **NAVEGACIÓN CON BALIZAS**:
  - Objetivos: navegar con seguridad siguiendo una secuencia de balizas.
  - Dinámica de atención: atender a la baliza más cercana que aún no haya sido alcanzada.
  - Dinámica de acción: avanzar hacia la baliza atendida, evitando los obstáculos situados en la dirección de avance. En ausencia de entrada visual, explorar en busca de la baliza.
- **IR A PUNTO**:
  - Objetivos: alcanzar la zona de referencia evitando los obstáculos del camino.
  - Dinámica de atención: seleccionar la región más cercana en la dirección de avance.
  - Dinámica de acción: si la región atendida es la región de referencia, avanzar hacia ella. Si no, evitar la región favoreciendo el avance hacia la referencia.

- **EXPLORAR**:
  - Objetivos: explorar el espacio visual manteniendo la alerta ante situaciones de peligro.
  - Dinámica de atención: seleccionar zonas del campo visual que no hayan sido atendidas recientemente.
  - Dinámica de acción: estado de parada mientras no ocurra nada que pueda afectar a la seguridad.
- **EVITAR EN PARADA**:
  - Objetivos: responder de forma segura a situaciones de peligro.
  - Dinámica de atención: seleccionar regiones en movimiento hacia la posición actual.
  - Dinámica de acción: evitar objetos en aproximación.

Cada comportamiento del ejemplo define una dinámica del sistema de atención a través de la modulación de un selector de objetivo. No obstante, su ejecución efectiva dependerá de la activación del comportamiento correspondiente, por lo que no todos actuarán al mismo tiempo sobre el control de atención global. A través de un esquema de control cooperativo, en un instante dado, sólo un subconjunto de comportamientos que resuelven la situación actual se encuentra en ejecución.

En el esquema de control del ejemplo se consideran dos posibles situaciones. La primera se da cuando existe una baliza localizada a la que el robot debe aproximarse. En este caso, los comportamientos en ejecución son *NAVEGACIÓN CON BALIZAS* e *IR A PUNTO*. Ambos están relacionados con el sistema de atención a través de un selector de baliza y un selector de obstáculo, respectivamente. El selector de baliza recibe información de aspecto de la baliza a localizar a través de descriptores conocidos. La frecuencia de atención sobre la baliza es especificada por el comportamiento de *NAVEGACIÓN* de manera inversamente proporcional a la distancia de ésta. Así, cuanto menor sea la distancia a la posición objetivo, mayor será la dedicación del sistema de atención a la focalización de dicha posición. Cuando el comportamiento de *NAVEGACIÓN* recibe información visual acorde con la baliza esperada, reprograma el selector de baliza para que funcione a la frecuencia adecuada y envía la posición objetivo al comportamiento *IR A PUNTO* para conseguir que el robot

se desplace hacia dicha posición. Para cumplir con sus objetivos, este comportamiento se encuentra conectado a un selector de obstáculos, que permite mantener la atención sobre aquellas regiones del entorno que se encuentran en la trayectoria del robot hacia la posición destino del desplazamiento (ver ejemplo de selección individual de la sección 6.2.4.1). El comportamiento *IR A PUNTO* interpreta la información visual recibida como la región más cercana que puede interferir en su camino hacia la posición objetivo. Así, lleva a cabo diferentes acciones en función de la situación de la región atendida con respecto a la de la posición de referencia, de manera que, si ambas coinciden, se considera que la zona atendida es la zona objetivo, por lo que responde con un avance hacia dicha posición. Sin embargo, si existe cierta distancia entre ambas zonas, supone que la región atendida representa un obstáculo hacia la posición objetivo, por lo que reacciona variando su trayectoria de avance de forma que el obstáculo sea evitado. Para que el tiempo de reacción ante situaciones de peligro sea el menor posible, el selector de obstáculos es modulado con una frecuencia de activación proporcional a la velocidad del robot.

La segunda situación contemplada por el sistema del ejemplo se da cuando, en su estado inicial, el comportamiento de *NAVEGACIÓN* no obtiene información visual de la baliza objetivo o cuando, tras haberla localizado, transcurre demasiado tiempo desde la última vez que se focalizó. En ambos casos, es necesario poner en marcha una actividad centrada en la búsqueda de la baliza para lo cual el comportamiento de *NAVEGACIÓN* desactiva el comportamiento *IR A PUNTO* y activa el de *EXPLORACIÓN*. La función de este último es acceder visualmente a zonas del entorno que no hayan sido vistas recientemente, función que realiza a través de su conexión con un selector de *zona no atendida* del sistema visual. Este selector mantiene una dinámica puramente *bottom-up* sobre las regiones detectadas en el entorno, que se hace efectiva por el mecanismo de inhibición de retorno presente en todos los selectores de objetivo. A fin de mantener la alerta ante situaciones de peligro, como puede ser el acercamiento excesivo de algún objeto en movimiento, el comportamiento de *EXPLORACIÓN* activa a su vez al comportamiento *EVITAR EN PARADA* que permite responder adecuadamente ante dichas situaciones. Para ello, mantiene una conexión con el sistema visual a través de un selector de *zona en avance*, que permite centrar la atención sobre regiones en movimiento que avanzan hacia la posición del robot. Este selector es modulado con una frecuencia alta para obtener así un tiempo de respuesta reducido por parte del comportamiento.

Las dos situaciones descritas anteriormente son resueltas a través de un control de atención multi-objetivo en el que se observan distintos tipos de atención que dan lugar a diferentes formas de control de alto nivel. Por un lado, existe una atención *top-down* sobre la baliza que permite que el sistema decida entre avanzar a una cierta posición del entorno o explorar en su búsqueda. En la situación de avance, la atención se reparte entre la baliza y los posibles obstáculos, manteniendo atención abierta sobre uno de los dos objetivos y atención encubierta sobre el otro. Esto permite que el robot reaccione a tiempo a los diferentes obstáculos a la vez que su posición con respecto a la baliza es recuperada con rapidez. En la situación de exploración, la búsqueda se lleva a cabo mediante un control de atención *bottom-up* en el que no existen especificaciones de objetivo. Para lograr la localización, el selector de baliza se mantiene activo funcionando a una frecuencia alta para responder de forma inmediata a la visualización del objetivo. Mientras este estado exploratorio se mantiene, el robot tiene la capacidad de reaccionar ante situaciones de peligro a través de un tercer objetivo visual en el que se mantiene una atención *top-down* por sus especificaciones y *bottom-up* por la situación real del entorno. De nuevo, la existencia de atención encubierta sobre el objetivo no focalizado permite mantener cierta continuidad en las acciones del robot a la vez que proporciona flexibilidad en sus respuestas de acuerdo con las condiciones del entorno.

Este esquema constituye un comportamiento básico de navegación que puede ser ampliado añadiendo nuevas dinámicas de atención-acción. Como ejemplo de estas ampliaciones, podríamos considerar las siguientes formas de actuación:

- **Recarga de baterías**: se trataría de incluir una dinámica de atención que permitiera al robot detectar zonas del entorno donde sea posible recargar sus baterías. Esta dinámica estaría activa en todo momento con una frecuencia de activación determinada por sus niveles de carga.

- **Recepción de órdenes**: este comportamiento actuaría como un módulo de comunicación que supondría una forma de interacción entre el robot y un humano. La dinámica de atención asociada se encargaría de dirigir la atención con prioridad alta hacia señales visuales asociadas con instrucciones conocidas por el robot. Dichas instrucciones estarían asociadas con la activación o desactivación de otros comportamientos incluidos en el sistema.

Los comportamientos descritos están limitados a las capacidades de actuación impuestas

por el cuerpo del robot en el que se encuentra implantado el sistema. La ampliación de estas capacidades, añadiendo nuevos elementos como, por ejemplo, un manipulador, dotarían al robot de nuevas formas de actuación, que permitirían incluir nuevas dinámicas de control basadas en la atención. En el caso concreto de un manipulador, la atención permitiría localizar y mantener las zonas de agarre adecuadas del objeto para la correcta ejecución de acciones posteriores sobre él. Estas cuestiones serán tratadas en posteriores trabajos incluidos dentro de las ampliaciones de esta tesis.

# Capítulo 7

# Experimentos

En este capítulo se muestran los resultados experimentales obtenidos a partir de una serie de pruebas reales destinadas a validar el sistema propuesto. En primer lugar, se presenta un conjunto de experimentos de evaluación independiente de ciertos módulos del sistema. Se trata de los componentes dedicados a la extracción de propiedades y al control de movimientos de cámara. Estos experimentos permiten mostrar el funcionamiento de determinados mecanismos cuyo rendimiento no puede ser apreciado en su totalidad en pruebas globales del sistema. Asimismo, se mostrarán partes de estos procesos dirigidas a aclarar los diferentes aspectos de los métodos empleados. Tras este primer grupo de pruebas, se describen varios experimentos de navegación que muestran diferentes dinámicas de atención-acción proporcionadas por el sistema al completo.

## 7.1. Extracción de propiedades

### 7.1.1. Propiedades de aspecto

Este primer grupo de experimentos está dirigido a probar las cualidades expuestas del conjunto de descriptores que constituyen las propiedades de aspecto de las regiones de una imagen. Se presentan los resultados obtenidos para regiones de dos imágenes en las que se comprobará, por un lado, la invariabilidad a escala y a rotaciones de los descriptores RIFT y Spin y, por otro lado, su capacidad discriminatoria. Este conjunto de pruebas constituyen una pequeña muestra cuyo objetivo no es en ningún caso realizar un análisis profundo de este componente de procesamiento, sino mostrar sus posibilidades como parte integrante del sistema atencional.

La figura 7.1 muestra la imagen utilizada para el primer experimento. En ella se han etiquetado las 5 regiones que constituyen este primer caso de estudio. Se trata de una estrella de 5 picos similares a través de los cuales comprobaremos la invariabilidad a rotaciones de los descriptores utilizados en el sistema.

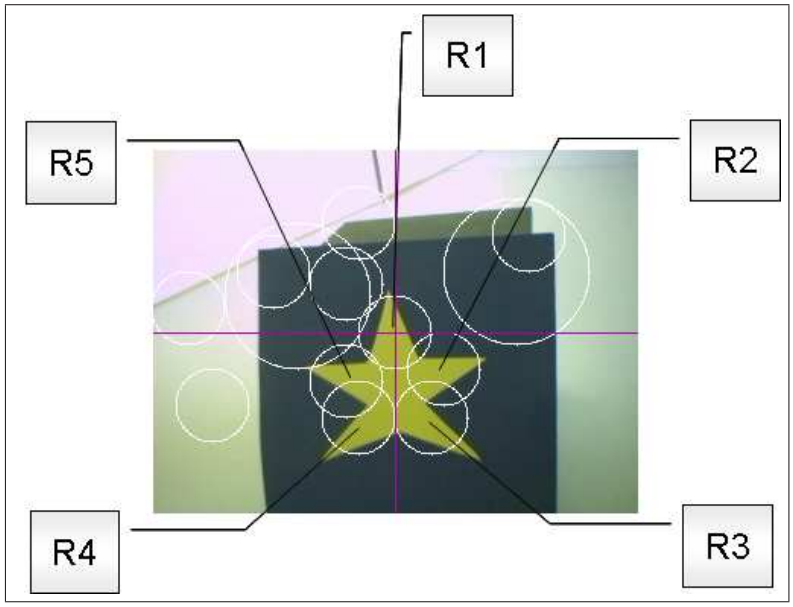
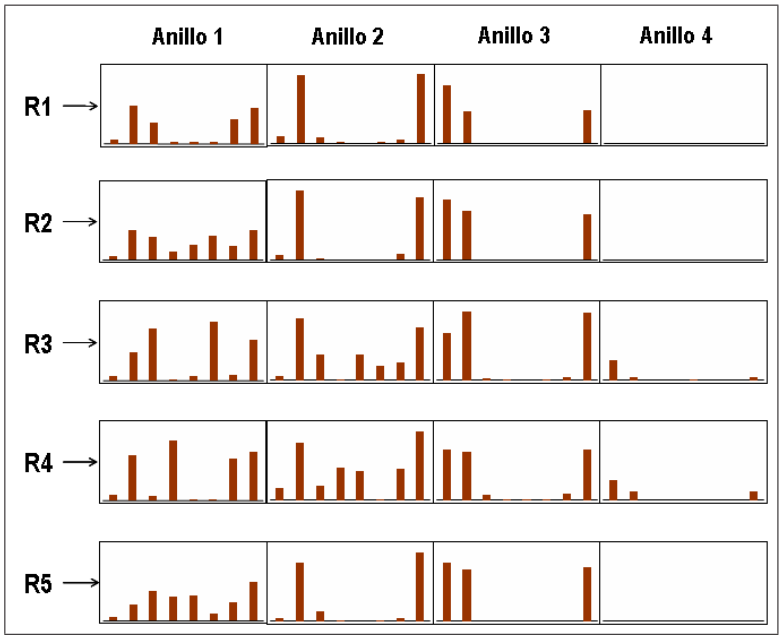


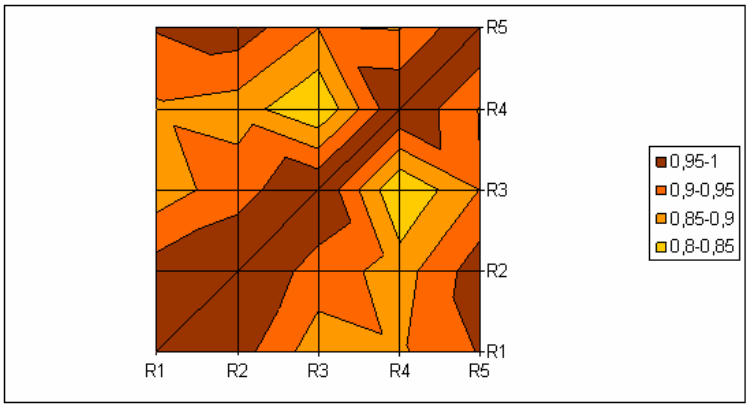

Figura 7.1: Regiones del primer experimento de extracción de descriptores

Cada descriptor ha sido extraído a partir de una división del parche de imagen correspondiente en 4 anillos, considerando el anillo 1 como el más exterior y el 4, el más cercano al centro de la región. Para el descriptor RIFT se han tomado 8 intervalos de orientación, lo que proporciona un total de 32 entradas. Los descriptores Spin han sido calculados para los tres planos RGB, utilizando 10 intervalos de intensidad en cada uno de ellos. Se han aplicado, además, los parámetros de suavizado $\alpha$ y $\beta$ en el cálculo de descriptores Spin para obtener resultados menos sensibles a pequeñas variaciones de iluminación o deformaciones producidas por cambios del punto de vista. En concreto, el parámetro de suavizado de distancia $\alpha$ ha sido fijado al 20 % del radio de la región y el de suavizado de intensidad $\beta$ a un valor de 10.

La figura 7.2 muestra los resultados para las regiones del primer experimento en relación a sus descriptores RIFT. En la figura (a) se presentan los histogramas obtenidos para cada




(a) Histogramas resultantes del proceso




(b) Correlación entre descriptores

Figura 7.2: Resultados de la extracción de descriptores RIFT de las regiones de la imagen 7.1

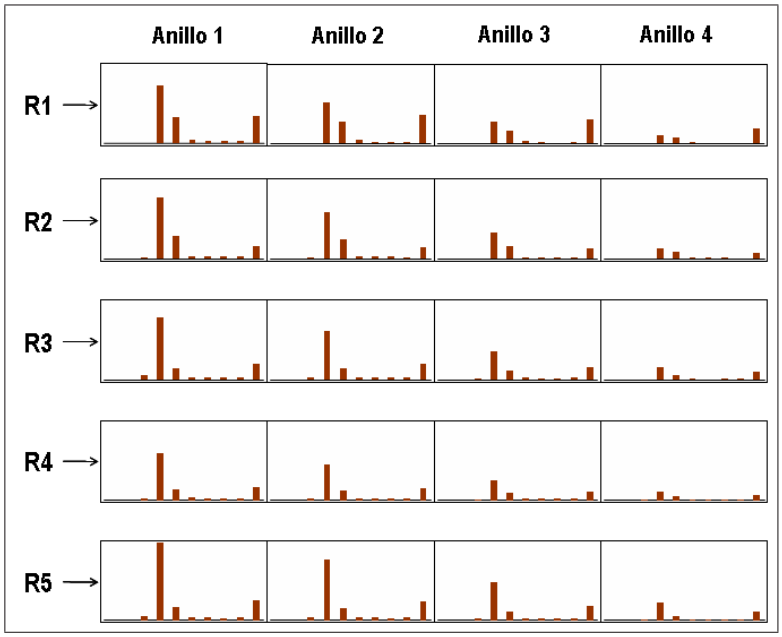
región. La gráfica (b) representa la relación entre los descriptores extraídos, medida a partir del coeficiente de correlación de Pearson entre cada par de descriptores. Dicho coeficiente ($cP$) se calcula dividiendo la covarianza ($\sigma_{didj}$) entre dos descriptores $di$ y $dj$ por el producto de sus desviaciones típicas ($\sigma_{di}$ y $\sigma_{dj}$):

$$cP = \frac{\sigma_{didj}}{\sigma_{di}\sigma_{dj}} \tag{7.1}$$

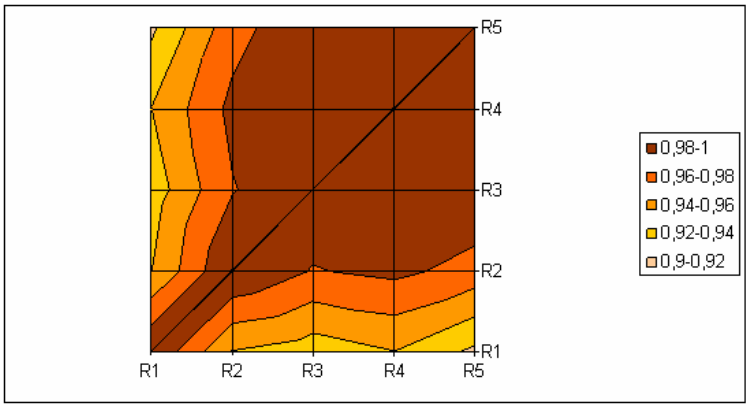
Este índice proporciona una medida de la similitud entre dos descriptores. Además, es independiente de la escala, lo que hace posible comparar descriptores de regiones de tamaños diferentes. El valor de este coeficiente varía en el intervalo $[-1, 1]$. Un valor cercano a 1 indica una semejanza alta entre dos descriptores, mientras que un valor próximo a -1 puede interpretarse como una relación prácticamente opuesta en todas las entradas que componen dichos descriptores.

Los resultados obtenidos en esta primera prueba (figura 7.2) indican un alto grado de similitud en todos los descriptores del experimento que puede apreciarse por la distribución de medidas de los histogramas resultantes, pero que, además, se confirma por los valores de correlación representados en la gráfica correspondiente. Este resultado viene dado por la propiedad de invariabilidad a rotaciones de los descriptores RIFT. No obstante, como se observa en la figura, las regiones $R3$ y $R4$ muestran una menor semejanza entre ellas, así como con las restantes regiones. Esto puede explicarse por la existencia de un cierto solapamiento en la imagen con sus regiones vecinas, lo que implica que los parches de imagen asociados no se corresponden exactamente con esos dos picos de la figura de estrella.

El efecto del solapamiento de regiones de $R3$ y $R4$ se ve reducido en los descriptores Spin RGB, tal y como puede apreciarse en las figuras 7.3, 7.4 y 7.5. Esto es debido al suavizado de histogramas proporcionado por los parámetros $\alpha$ y $\beta$ utilizados en el cálculo de descriptores. Como se observa en estas figuras, los resultados obtenidos muestran una mayor semejanza entre todas las regiones del experimento.

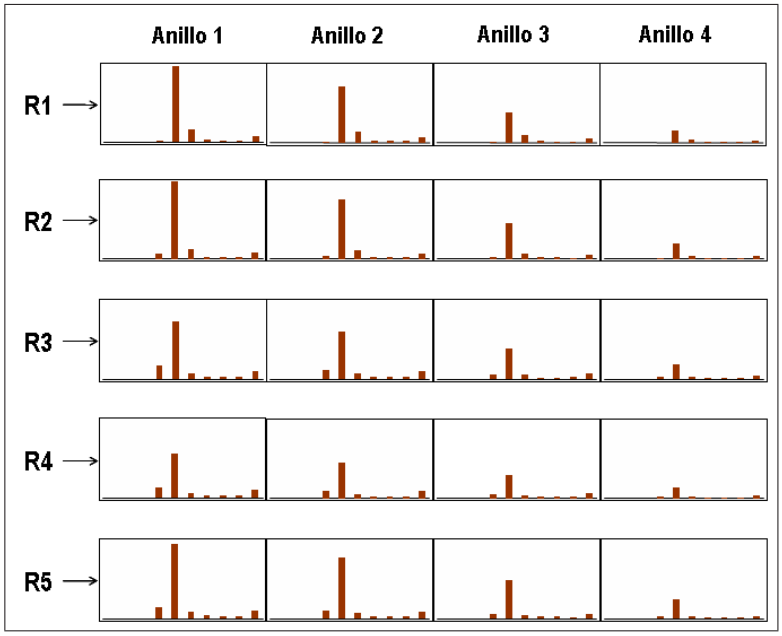



(a) Histogramas resultantes del proceso

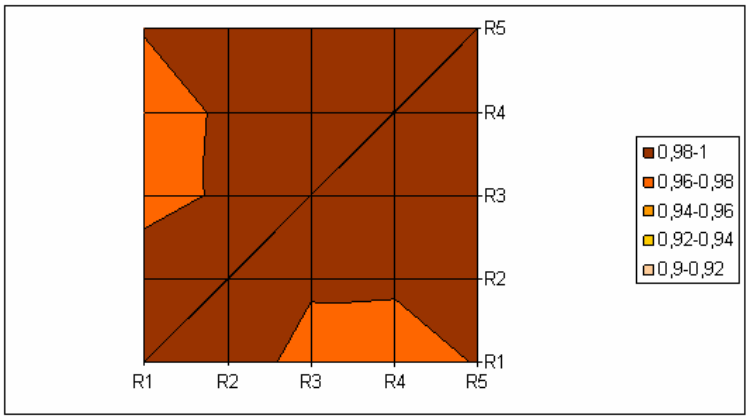



(b) Correlación entre descriptores

Figura 7.3: Resultados de la extracción de descriptores Spin (plano rojo) de las regiones de la imagen 7.1

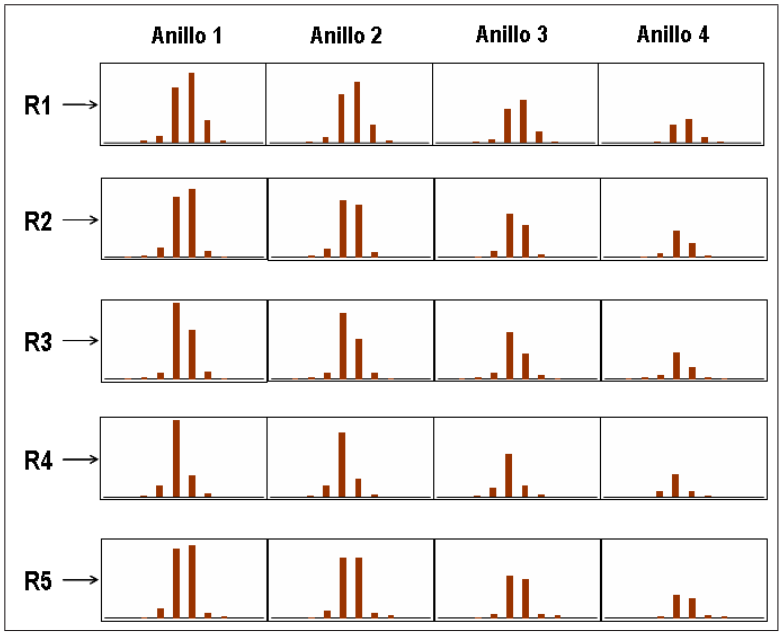



(a) Histogramas resultantes del proceso

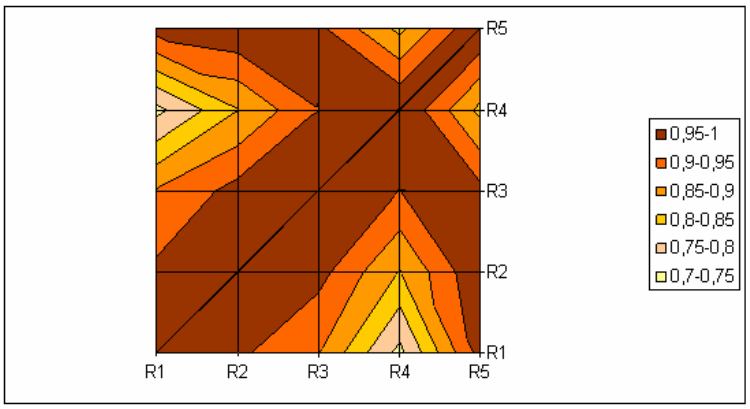



(b) Correlación entre descriptores

Figura 7.4: Resultados de la extracción de descriptores Spin (plano verde) de las regiones de la imagen 7.1




(a) Histogramas resultantes del proceso




(b) Correlación entre descriptores

Figura 7.5: Resultados de la extracción de descriptores Spin (plano azul) de las regiones de la imagen 7.1

El segundo experimento está destinado a comprobar la invariabilidad a la escala de los descriptores utilizados en el sistema, así como su potencial de generalización como elementos representativos de la apariencia de regiones. Con este fin, se ha aplicado el proceso de extracción de descriptores al grupo de regiones de la figura 7.6. Se trata de 4 regiones de apariencia similar en cuanto a la forma y/o la distribución de colores de su interior. Las regiones $R1$ y $R2$ contienen la misma imagen en diferentes escalas. La imagen de $R3$ mantiene una forma similar a la de $R1$ y $R2$, aunque sus atributos de color son completamente diferentes. Por último, la distribución de colores de la imagen de $R4$ presenta cierta semejanza con $R1$ y $R2$, aunque no puede afirmarse lo mismo de su estructura.

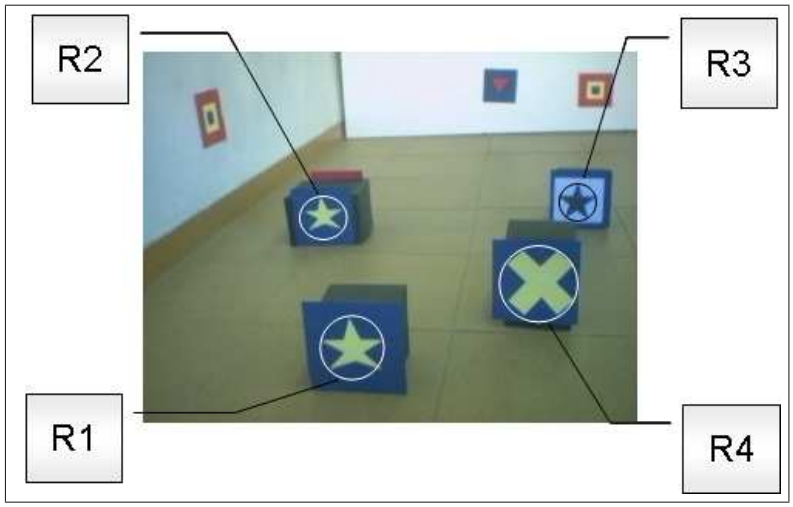
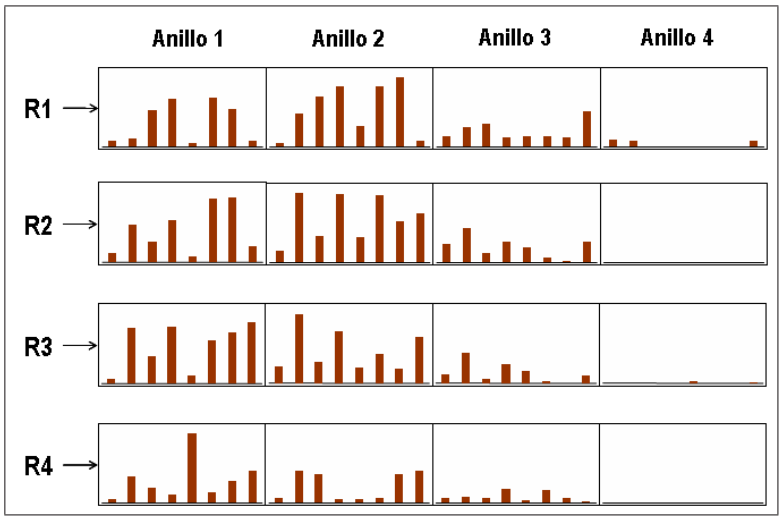


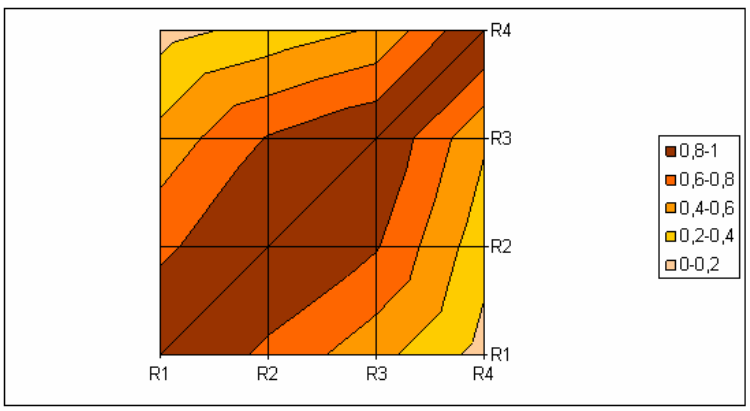

Figura 7.6: Regiones del segundo experimento de extracción de descriptores

Como primer resultado de este experimento, en la figura 7.7 se muestran los datos correspondientes a los descriptores RIFT de estas 4 regiones. Recordemos que el descriptor RIFT se basa en la distribución de orientaciones de gradiente en el interior de la región, por lo que una imagen y su imagen negativa producen histogramas opuestos. Para evitar esta circunstancia, cuando la orientación del gradiente en un punto sea negativa, se considerará la dirección inversa, sumándole a dicha orientación un ángulo de $180^o$. Esto proporciona descriptores independientes de la relación de intensidades, positiva o negativa, entre los píxels situados a cada lado de un borde. De esta forma, se obtiene una descripción de la región relacionada únicamente con la estructura de su interior y separada completamente de la distribución de colores, aspecto que ya es considerado en la representación proporcionada por los descriptores Spin.

La consideración anterior sobre el descriptor RIFT proporciona los resultados esperados en las regiones del experimento. Como se observa en la figura 7.7, las 2 primeras regiones presentan un alto grado de similitud, lo que es coherente con la invariabilidad a la escala propia de estos descriptores. Además, $R3$ mantiene una cierta relación de semejanza con $R1$ y con $R2$, de mayor grado con esta última. Por último, la representación de $R4$ muestra diferencias notables con las restantes regiones, proporcionando un resultado congruente con las imágenes del experimento.




(a) Histogramas resultantes del proceso

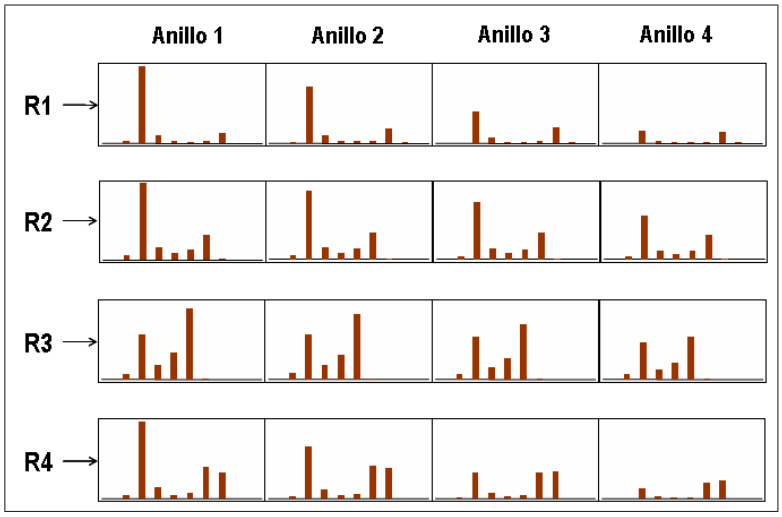



(b) Correlación entre descriptores

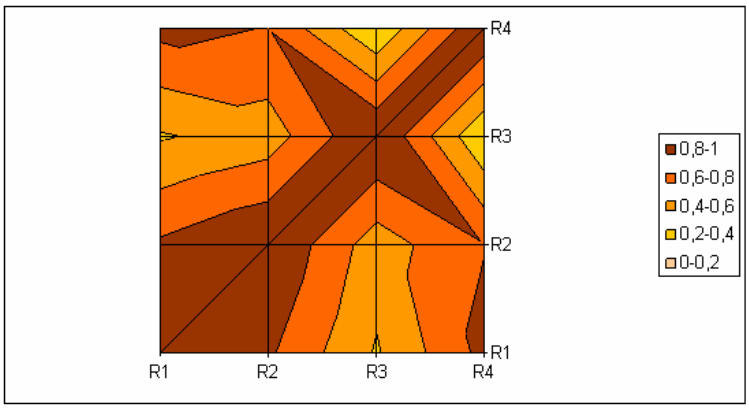
Figura 7.7: Resultados de la extracción de descriptores RIFT de las regiones de la imagen 7.6

Con respecto a los descriptores Spin (figuras 7.8, 7.9 y 7.10), los resultados muestran la semejanza, en cuanto a la distribución de colores, entre $R1$, $R2$ y $R4$. No ocurre así con $R3$,

como cabía esperar, que presenta correlaciones prácticamente nulas con los descriptores de las otras 3 regiones. La similitud entre $R4$ y las regiones $R1$ y $R2$ muestra la dificultad de los descriptores Spin para discriminar entre imágenes que presenten los mismos colores de fondo e interior, por lo que su uso aislado no es conveniente. La combinación de los dos tipos de descriptores, RIFT y Spin, permitirá resolver casos en los que cada descriptor individual no proporciona resultados completamente fiables.

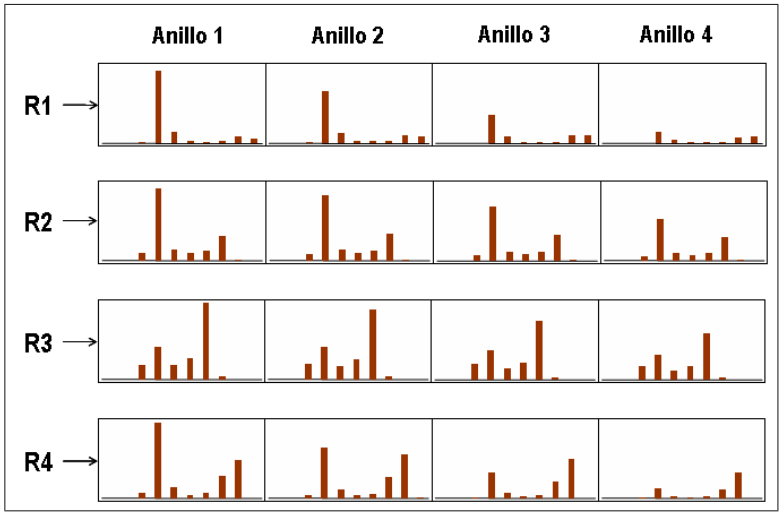



(a) Histogramas resultantes del proceso

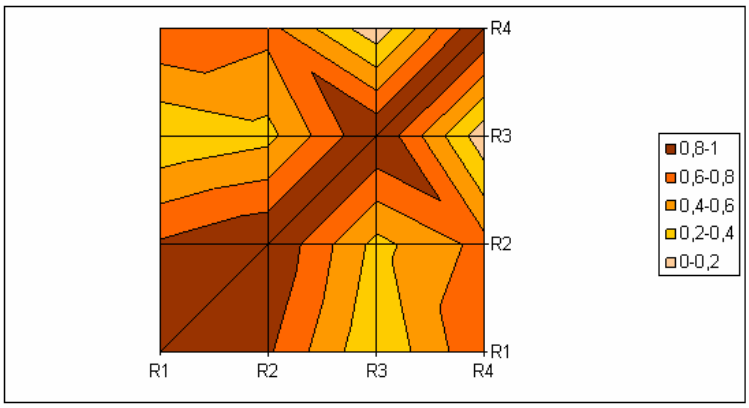



(b) Correlación entre descriptores

Figura 7.8: Resultados de la extracción de descriptores Spin (plano rojo) de las regiones de la imagen 7.6

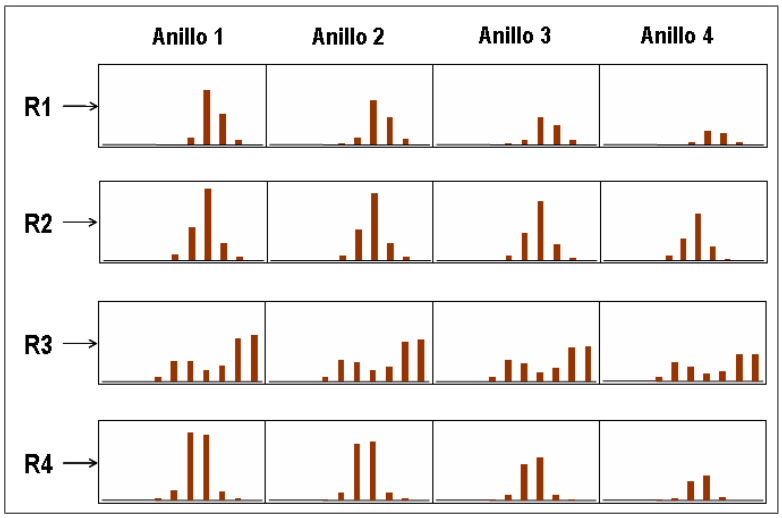



(a) Histogramas resultantes del proceso

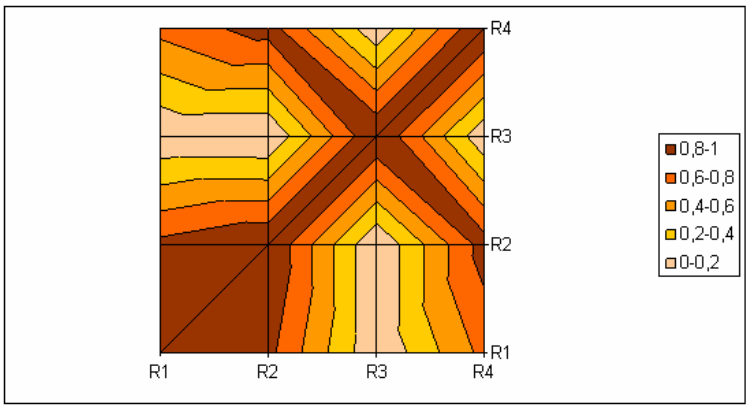



(b) Correlación entre descriptores

Figura 7.9: Resultados de la extracción de descriptores Spin (plano verde) de las regiones de la imagen 7.6




(a) Histogramas resultantes del proceso




(b) Correlación entre descriptores

Figura 7.10: Resultados de la extracción de descriptores Spin (plano azul) de las regiones de la imagen 7.6

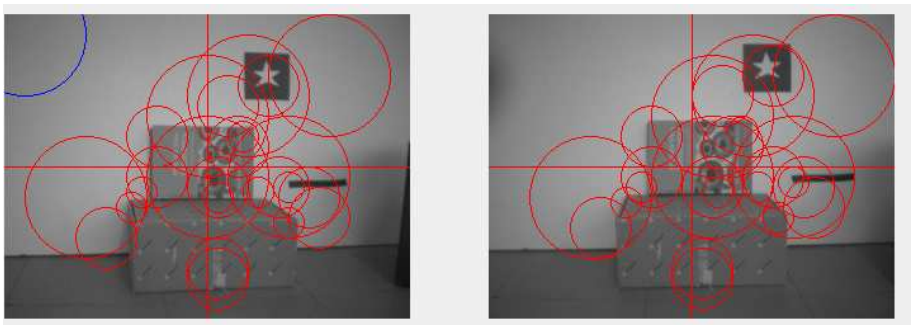
### 7.1.2. Propiedades espaciales

Se muestra en esta sección un grupo de resultados proporcionados por el subsistema de extracción de propiedades espaciales. Nos centraremos en el cálculo de posición y orientación espacial de regiones a través de dos experimentos.

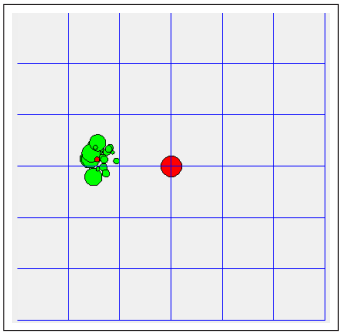
En primer lugar, se presentan los resultados de la localización espacial de regiones del entorno desde 4 situaciones distintas del robot (figuras 7.11, 7.12, 7.13 y 7.14). Se muestra, por cada vista, la correspondencia de regiones entre las imágenes de las dos cámaras y la proyección en planta de las posiciones obtenidas. Para que esta segunda representación sea más informativa, cada región localizada ha sido representada como un objeto esférico de dimensiones proporcionales a su extensión en la imagen. Se muestra también una imagen real de la escena, capturada por una cámara situada en el techo. Se ha marcado en rojo la zona del entorno visible por las cámaras en cada instante y en verde la posición real del robot.

La imagen (a) de cada figura muestra el resultado del emparejamiento entre regiones. La imagen de la izquierda contiene las regiones resultantes de la detección por Harris-Laplace. En la de la derecha, se indican las regiones homólogas de las de la izquierda. Se han marcado en rojo las regiones emparejadas y en azul las que no han proporcionado resultado en el proceso de búsqueda de correspondencias. Como se observa en la distintas situaciones, el porcentaje de éxito de este proceso de búsqueda es elevado.

La gráfica (b) representa la proyección en planta de las regiones homólogas de la imagen (a). La posición del robot (círculo rojo) en cada situación ha sido calculada a partir de la estimación de odometría proporcionada por los componentes de control de la base. La posición de cada región dentro de esta representación se obtiene de la transformación de las coordenadas 3D estimadas, relativas a la posición del robot, a un sistema de referencia global del mundo. Para distinguirla del resto, se ha representado en rojo la región que constituye el foco de atención actual. Los resultados obtenidos en este experimento muestran una estimación del 3D coherente con cada situación. Las variaciones producidas por los cambios de posición del robot en el cálculo de posición espacial de cada región son poco significativas, proporcionando representaciones estables de los elementos que componen la escena.



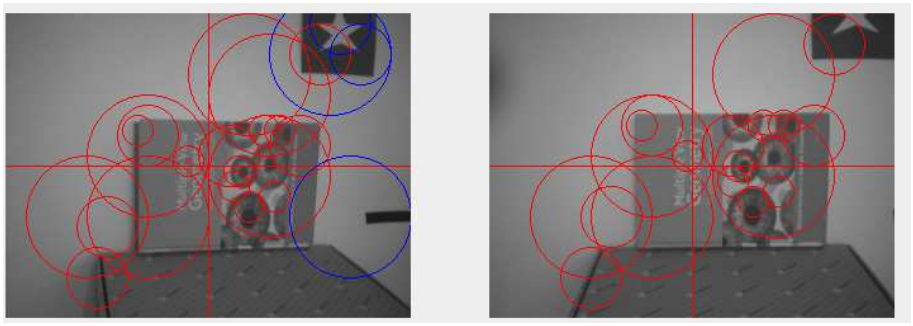
(a)



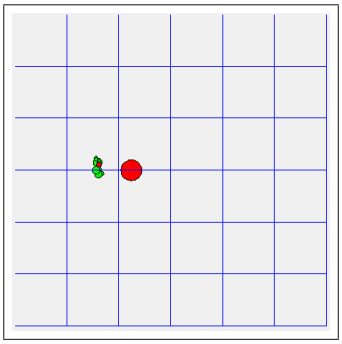
(b)

(c)

Figura 7.11: Correspondencia y 3D de regiones (situación 1)



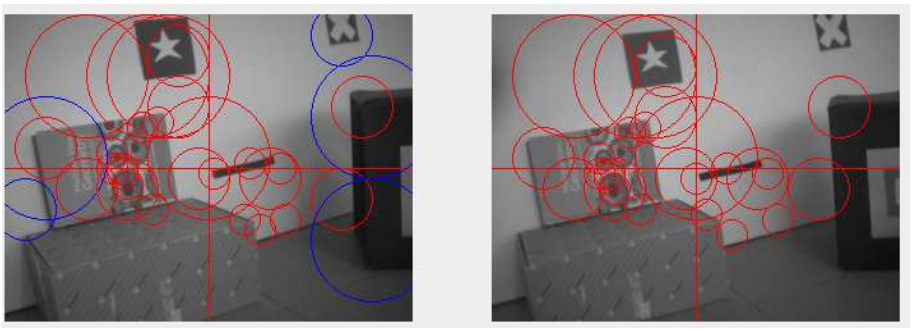
(a)



(b)

(c)

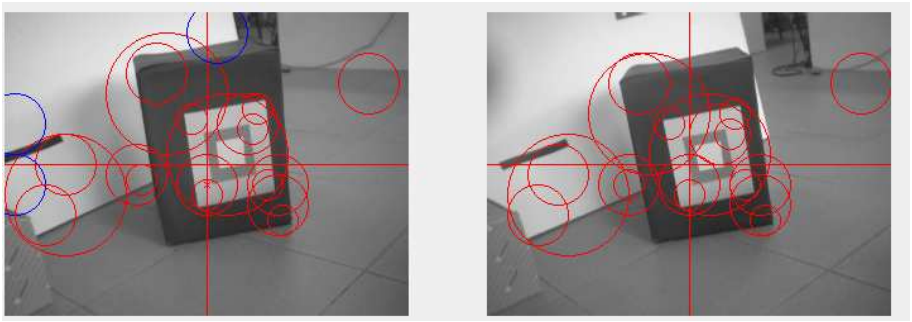
Figura 7.12: Correspondencia y 3D de regiones (situación 2)



(a)

(b) (c)

Figura 7.13: Correspondencia y 3D de regiones (situación 3)

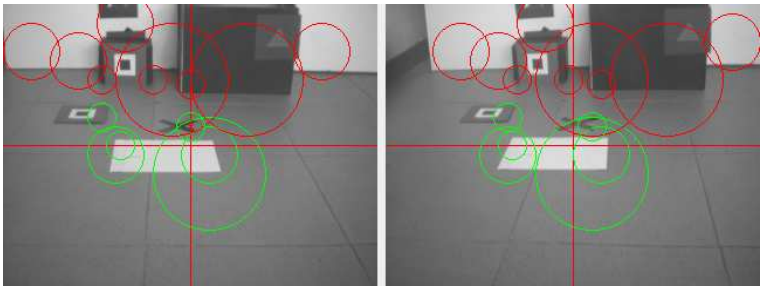


(a)

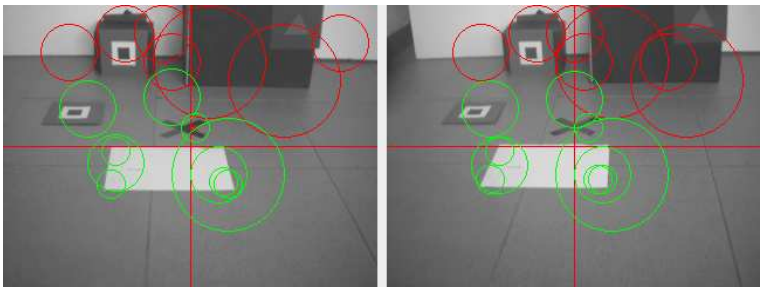
(b) (c)

Figura 7.14: Correspondencia y 3D de regiones (situación 4)

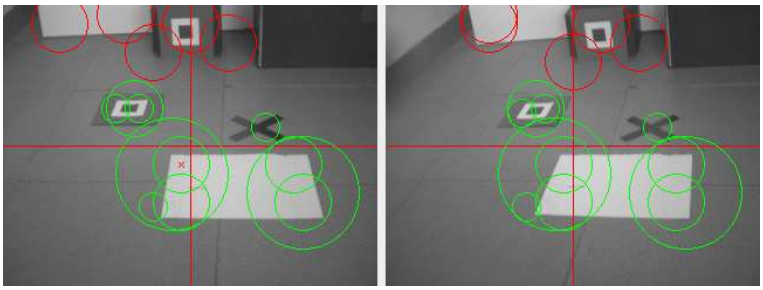
El segundo experimento de extracción de propiedades espaciales está dirigido a analizar el método propuesto sobre estimación de orientaciones de superficies. Para ello se ha diseñado una prueba en la que el robot debe distinguir entre regiones planas con orientaciones paralelas y perpendiculares al suelo.



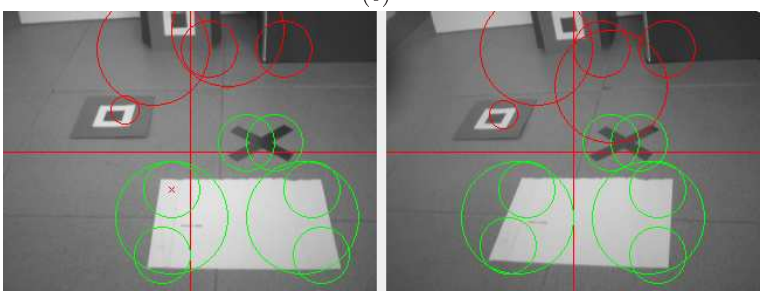
(a)



(b)



(c)



(d)

Figura 7.15: Detección de superficies planas sobre el suelo

Los resultados de esta prueba se presentan en la figura 7.15. Se trata de un grupo de imágenes obtenidas en varios instantes de una secuencia de avance del robot. Cada imagen representa las dos vistas del par estéreo. En función de la orientación estimada, cada región es considerada una superficie paralela (marcadas en verde) o perpendicular al suelo (representadas en rojo). Los resultados muestran la bondad del método propuesto. Como puede observarse, aparecen varios casos en los que la falta de textura no permitiría un cálculo directo de la homografía. La estimación inversa del proceso empleado permite resolver dichos casos adecuadamente.

## 7.2. Control de movimientos de cámara

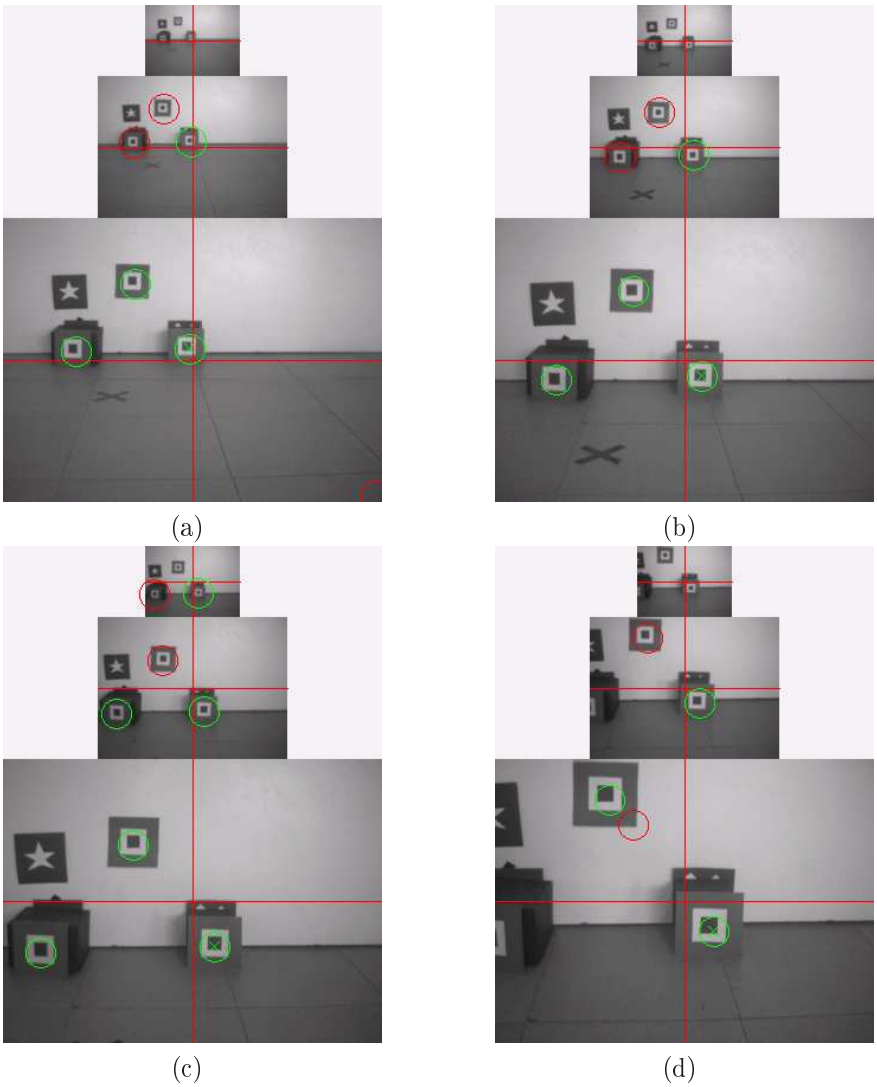
Se presentan en esta sección varios experimentos de evaluación de los métodos propuestos para el control de movimientos de cámara. En primer lugar, se presentan pruebas separadas de control de seguimiento y vergencia. A continuación, se muestran los resultados de un experimento de control conjunto de ambos movimientos.

### 7.2.1. Seguimiento de un objetivo

Para tratar de esclarecer los distintos aspectos del proceso encargado del control de seguimiento visual, se presenta a continuación el resultado de la localización de un objetivo en varios instantes de una secuencia de avance del robot. La figura 7.16 muestra este resultado. Cada subfigura incluye la representación piramidal de la escena en un instante determinado. Esta representación permite analizar el proceso de localización ascendente propuesto.

En la escena del experimento aparecen 3 regiones de aspecto similar, de las cuales, la situada más a la derecha constituye el objetivo que debe ser foveatizado a lo largo de la secuencia. Por cada nivel de la pirámide, se han marcado las regiones que mantienen una correlación alta con la región objetivo. De entre ese grupo de regiones, las marcadas en verde presentan además un valor de correlación muy próximo al máximo obtenido, mientras que las señaladas en rojo, a pesar de presentar una alta similitud con la región objetivo, muestran una correlación inferior al máximo en un porcentaje significativo. Por cada nivel, sólo aquellas regiones de correlación alta y cercana al valor máximo son consideradas zonas candidatas a la fijación visual. Si sólo existe una región, el proceso finaliza y, como resultado, se produce un cambio de posición en la cámara que permita fijar la región en el centro de la imagen. Si hay más de una región candidata, el proceso de búsqueda se repite en el

siguiente nivel de la pirámide, tal y como puede observarse en la secuencia de imágenes de la figura 7.16. El ascenso de nivel supone un aumento de tamaño de la región objetivo por lo que la búsqueda desde un nivel superior permite descartar regiones seleccionadas desde niveles inferiores. Este hecho aparece reflejado en las imágenes de la figura 7.16. En los instantes de la secuencia representados en las imágenes (a), (b) y (d) el proceso de localización requiere el acceso a los dos primeros niveles de la pirámide, tras lo cual se obtiene la nueva posición de la región objetivo. En la imagen (c), la búsqueda en los dos niveles inferiores no es suficiente para encontrar un único candidato, por lo que el proceso se repite en el nivel superior que proporciona finalmente el resultado deseado.

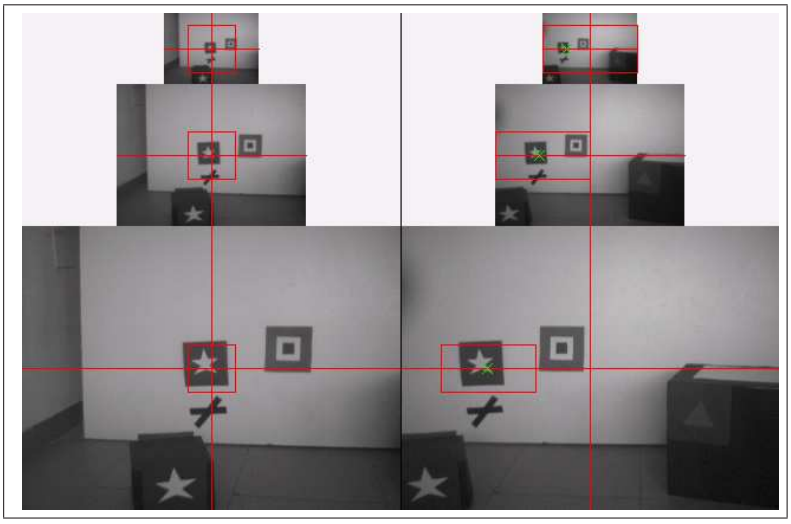


(a) (b) (c) (d)

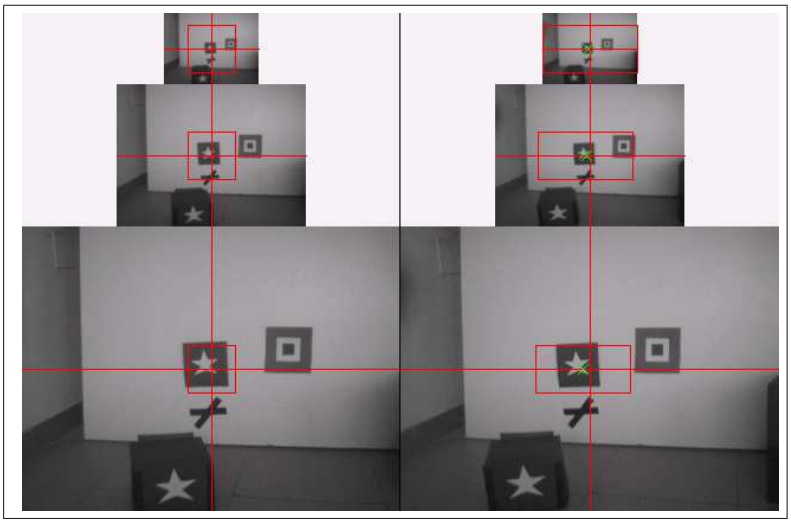
Figura 7.16: Seguimiento de una región en una secuencia de avance

### 7.2.2. Control de vergencia

El segundo bloque de experimentos de esta sección está dedicado al análisis del controlador de vergencia incluido en el sistema. Se muestran los resultados obtenidos a partir de 3 situaciones diferentes: control de vergencia en una situación favorable, vergencia hacia una zona sin textura y vergencia hacia una zona situada fuera del campo de visión.



(a) Cálculo de posición

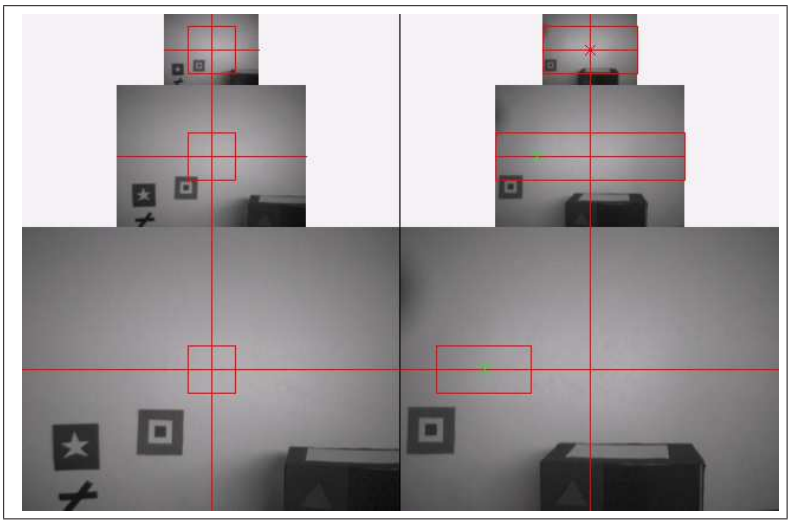


(b) Posición alcanzada

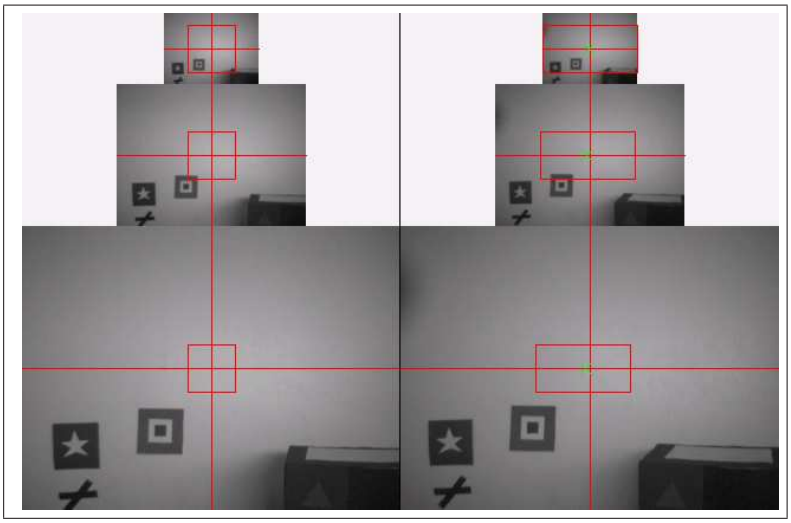
Figura 7.17: Vergencia en una situación favorable

En la primera situación (figura 7.17), la cámara fija (imagen de la izquierda) se encuentra centrada sobre una zona texturada y visible por la cámara de control de vergencia

(imagen de la derecha). La localización de la posición de vergencia arranca desde el nivel superior de la pirámide y es ajustada siguiendo un proceso de búsqueda descendente por esta estructura multi-resolución. La imagen (a) de la figura 7.17 muestra este proceso. La ventana de búsqueda en cada nivel ha sido recuadrada en rojo. Dentro de dicha ventana se ha marcado con una "X" la posición de vergencia obtenida en ese nivel. Como se observa en las imágenes de la figura, la posición localizada en cada nivel selecciona la ventana de búsqueda del siguiente hasta completar el recorrido por la pirámide y obtener la posición final en el nivel inferior. Tras finalizar el proceso, la posición obtenida es centrada en la imagen a través del giro correspondiente de la cámara (imagen (b) de la figura 7.17).



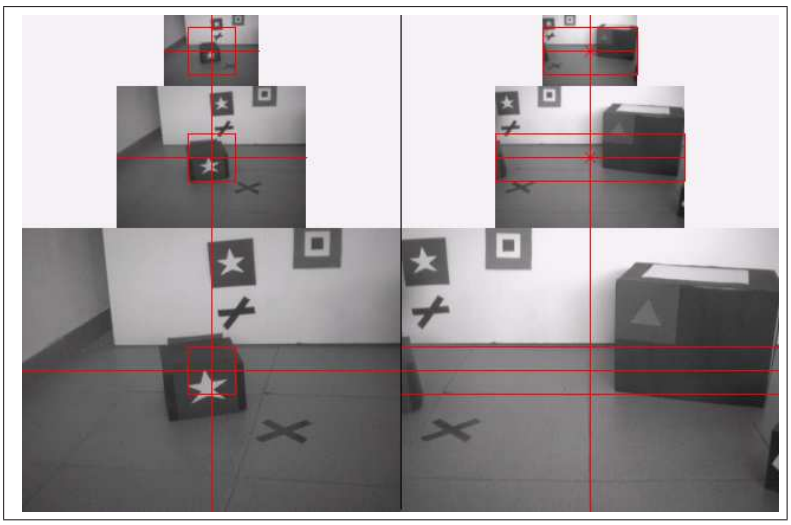
(a) Cálculo de posición



(b) Posición alcanzada

Figura 7.18: Zona de vergencia sin textura

El segundo experimento está dirigido a probar el comportamiento del subsistema de control de vergencia en una situación en la que la zona a la que la cámara debe converger carece de textura. La figura 7.18 muestra el resultado de esta prueba. Inicialmente, el nivel superior no consigue localizar ninguna correspondencia debido a que la zona de vergencia de la cámara fija no está completamente incluida en la imagen de la cámara de vergencia (imagen (a)). En el siguiente nivel, la posición de vergencia es localizada con éxito, lo que permite ajustar la ventana de búsqueda del nivel inferior. En este último nivel, la máxima correlación obtenida no mejora el valor de correlación obtenido previamente, por lo que la posición final de vergencia es fijada a la calculada por el nivel anterior. La imagen (b) muestra el resultado de esta localización tras el cambio de posición de la cámara.

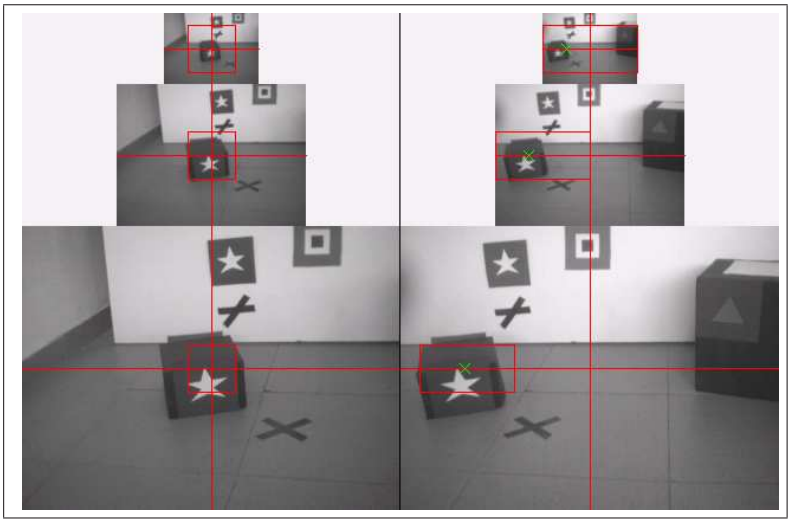


(a) Sin posición de vergencia. Se inicia el proceso de búsqueda dinámica

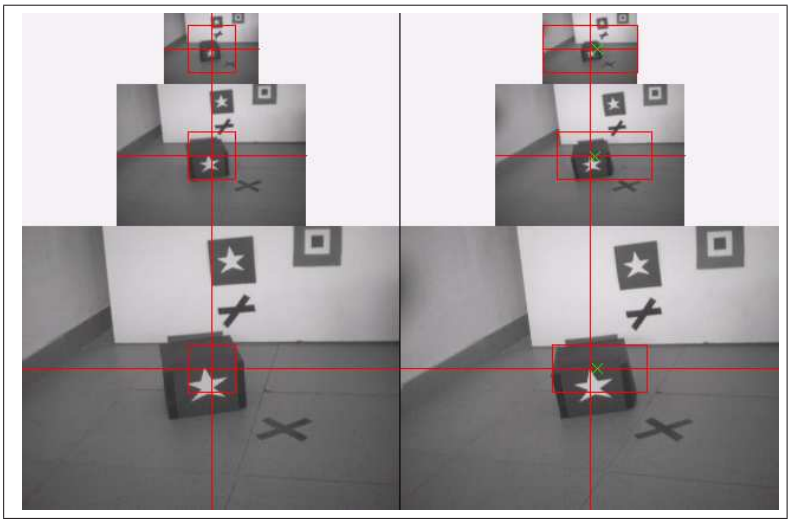
Figura 7.19: Zona de vergencia fuera del campo de visión

Para finalizar, se muestra una última situación en la que la zona de vergencia especificada por la cámara fija se encuentra fuera del campo visual de la cámara sobre la que se realiza el control. Esta situación es la indicada en la figura 7.19. En la imagen (a) de esta figura se muestra la aplicación del procedimiento de localización descendente. La falta de resultado provoca la puesta en marcha de un proceso de búsqueda dinámica consistente en modificar la posición de la cámara de manera que sea posible cubrir el rango completo de vergencia. Por cada nueva posición, la búsqueda piramidal se inicia de nuevo hasta lograr la localización de la posición correcta. El resultado de esta dinámica puede observarse en la imagen (b) de la figura 7.19. Tras modificar la posición de la cámara por la búsqueda

dinámica, el procedimiento jerárquico de localización consigue obtener la posición de vergencia adecuada, que es finalmente alcanzada en la imagen (c).



(b) Posición de vergencia encontrada



(c) Posición alcanzada

Figura 7.19: Zona de vergencia fuera del campo de visión (continuación)

### 7.2.3. Control conjunto de seguimiento y vergencia

Para concluir la serie de experimentos dedicados al control de movimientos de cámara, se presentan los resultados obtenidos en una prueba de funcionamiento simultáneo de los subsistemas de control de seguimiento y de vergencia. Se trata de un experimento de fijación binocular de un objetivo en movimiento. La figura 7.20 muestra varios instantes de la secuencia de control obtenida en esta prueba. Las imágenes de la izquierda se corresponden con las dos vistas de la escena obtenidas por las cámaras del par estéreo. Las de la derecha muestran una vista general de la escena capturada en el mismo instante que las imágenes situadas a su izquierda. En este último grupo de imágenes, se ha marcado en rojo la posición del robot que realiza el control y en verde la posición de otro robot que constituye el objetivo en movimiento.

Como se observa en las imágenes de la figura, los cambios de posición del objetivo producen cambios inmediatos de posición en las cámaras del robot que realiza el control. Estos cambios se realizan de manera independiente en cada cámara, guiados por su propio sistema de control. En la cámara de la izquierda, el control de seguimiento visual provoca giros sobre los dos ejes de rotación de la cámara que permiten mantener centrado el objetivo. En la cámara de la derecha, las disparidades entre las imágenes del par estéreo se traducen en giros sobre su eje vertical guiados por el sistema de control de vergencia. La necesidad de disparidad implica que el control de vergencia se inicia en un tiempo posterior al de seguimiento.

La velocidad del control permite capturar de forma continua los cambios visuales producidos en la región objetivo. Esto proporciona respuestas de seguimiento efectivas ante variaciones sustanciales del objetivo, tal y como puede apreciarse en las últimas imágenes de la secuencia.

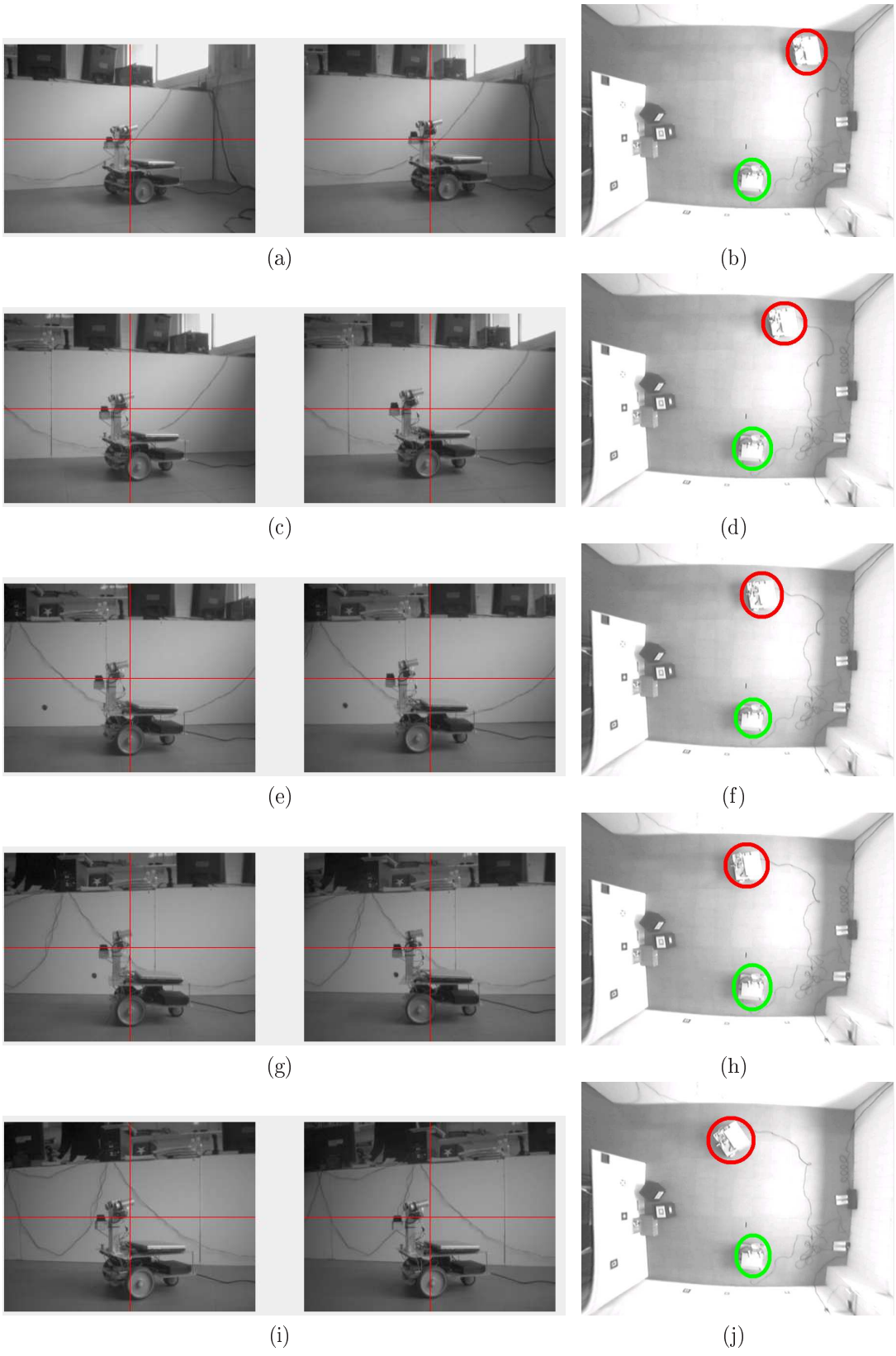
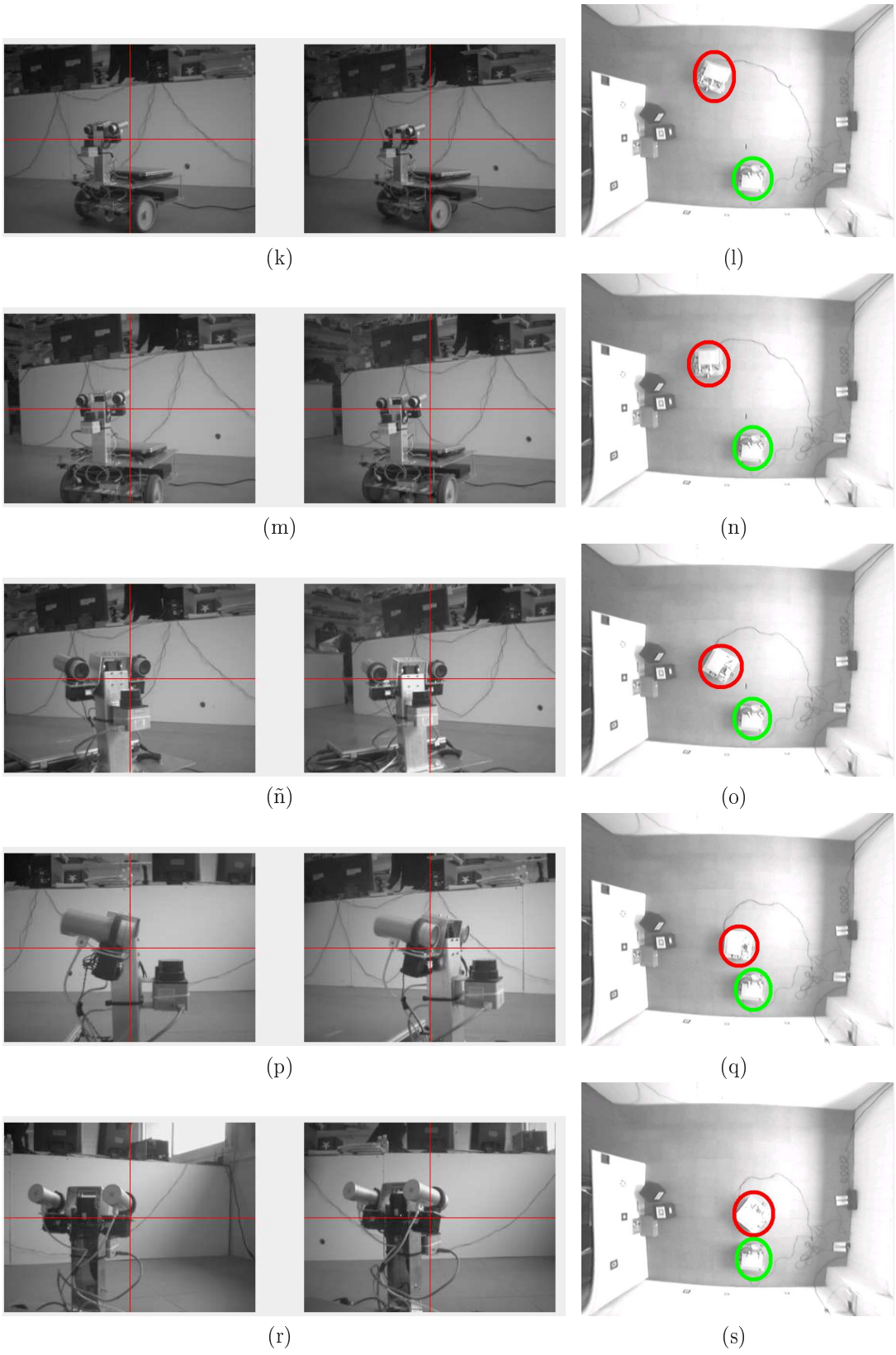



Figura 7.20: Seguimiento y vergencia de un objetivo en movimiento (primera parte)



(k) (l)

(m) (n)

(ñ) (o)

(p) (q)

(r) (s)

Figura 7.20: Seguimiento y vergencia de un objetivo en movimiento (segunda parte)

## 7.3. Dinámicas de atención-acción

El sistema global ha sido probado a través de varios experimentos de navegación autónoma. Se presentan los resultados obtenidos para 4 pruebas, mostrando capturas de varios instantes de las secuencias de navegación. En una primera figura se muestra, para cada prueba, la secuencia de imágenes capturadas desde el robot por las dos cámaras del par estéreo. En ellas puede observarse la zona de atención seleccionada en cada instante. En segundo lugar, para cada instante reflejado en las imágenes de la primera figura, se muestra, en una segunda figura, una vista general captada por una cámara situada en el techo que permite apreciar la situación real de la escena en ese momento. En las imágenes de esta segunda figura, se ha marcado en rojo la zona atendida en cada instante para facilitar la asociación entre los dos grupos de imágenes.

Para poner en marcha esta batería de pruebas se ha empleado un sistema de control compuesto por 3 comportamientos de alto nivel enlazados a 3 selectores de objetivo del sistema atencional. Los 3 comportamientos de alto nivel son *NAVEGACIÓN CON BALIZAS*, *IR A PUNTO* y *EXPLORAR*, que modulan, respectivamente, a los selectores atencionales de *baliza*, *obstáculo* y *zona no atendida*. Estos componentes son los mismos que fueron expuestos en el ejemplo de control basado en la atención de la sección 6.3. El comportamiento de *NAVEGACIÓN CON BALIZAS* envía al selector asociado un grupo de descriptores que permiten identificar el aspecto de la baliza que debe ser seleccionada. El selector de baliza selecciona la zona del campo visual que mantiene una máxima relación con los descriptores especificados. Si no es posible realizar esta selección, el comportamiento de *NAVEGACIÓN* asume que la baliza no es visible y pone en marcha un comportamiento de *EXPLORACIÓN*. Este último activa al selector de *zona no atendida* que, a través del mecanismo de inhibición de retorno, permite explorar visualmente la escena. El funcionamiento simultáneo de los selectores de *baliza* y de *zona no atendida* da lugar a que, en el momento en que la baliza aparezca en el campo visual, ésta sea seleccionada como foco de atención convirtiéndose en entrada visual única de los comportamientos de alto nivel. Como consecuencia, el comportamiento de *NAVEGACIÓN CON BALIZAS* desactiva la exploración y activa el comportamiento *IR A PUNTO* indicando como posición objetivo la posición de la baliza. Además, modifica el funcionamiento del selector de *baliza* para que éste se active con una frecuencia inversamente proporcional a la distancia de la posición objetivo. El comportamiento *IR A PUNTO*, por su parte, activa el selector atencional de obstáculo y le envía, como posición de referencia, la posición de la baliza. Este selector se

encarga de seleccionar las zonas del entorno próximas a la trayectoria hacia el objetivo. En función de los datos recibidos a través del flujo de información visual, el comportamiento *IR A PUNTO* lleva a cabo diferentes acciones que permiten el acercamiento hacia la baliza sin colisionar. Si, en un momento determinado, el selector de obstáculos no consigue seleccionar ninguna región, el flujo de información se interrumpe y el comportamiento asociado modifica la posición de los sensores produciendo un barrido vertical que permite que el sistema atencional vuelva a recuperar el control. Esta situación puede darse inicialmente por una posición elevada de la baliza con respecto a la altura del robot que provoca la ausencia de objetivos coherentes con las especificaciones del selector de obstáculos. Transcurrido un cierto tiempo, el selector de baliza vuelve a recuperar el control provocando que todo el ciclo descrito comience de nuevo. Una vez alcanzada la baliza, el comportamiento de *NAVEGACIÓN* reinicia el proceso enviando a su selector la descripción de una nueva baliza.

### 7.3.1. Experimento 1: aproximación a una baliza con sorteo de obstáculos

En el primer experimento de prueba de este sistema de control, el robot debe alcanzar una baliza que contiene la imagen de una estrella, situada en una pared frontal a su orientación inicial. Existe un obstáculo situado frente a él que debe evitar en su trayectoria de aproximación a la baliza. La secuencia de imágenes de las figuras 7.21 y 7.22 muestran los resultados de esta prueba. Inicialmente (imagen (a)), la atención es fijada sobre la baliza, provocando la puesta en marcha del comportamiento *IR A PUNTO*. En ese momento, el control atencional es dirigido por el selector del obstáculos, que actúa fijando la atención sobre la zona de máximo peligro en el camino hacia la baliza (imagen (b)). A partir de ese instante, la atención se centra en las regiones situadas en posiciones cercanas a la zona de paso que, junto con la información odométrica, provocan acciones de giro y avance, permitiendo al robot sortear los obstáculos sin desviarse en exceso de la zona objetivo (imágenes (c)-(k)). Transcurrido un cierto tiempo, el selector de baliza toma de nuevo el control atencional, permitiendo actualizar la posición objetivo (imagen (l)). De acuerdo a esta nueva posición, el selector de obstáculos actualiza sus especificaciones de control, permitiendo un acercamiento más preciso a la baliza (imágenes (m)-(ñ)) hasta alcanzar la posición deseada (imagen (o)).

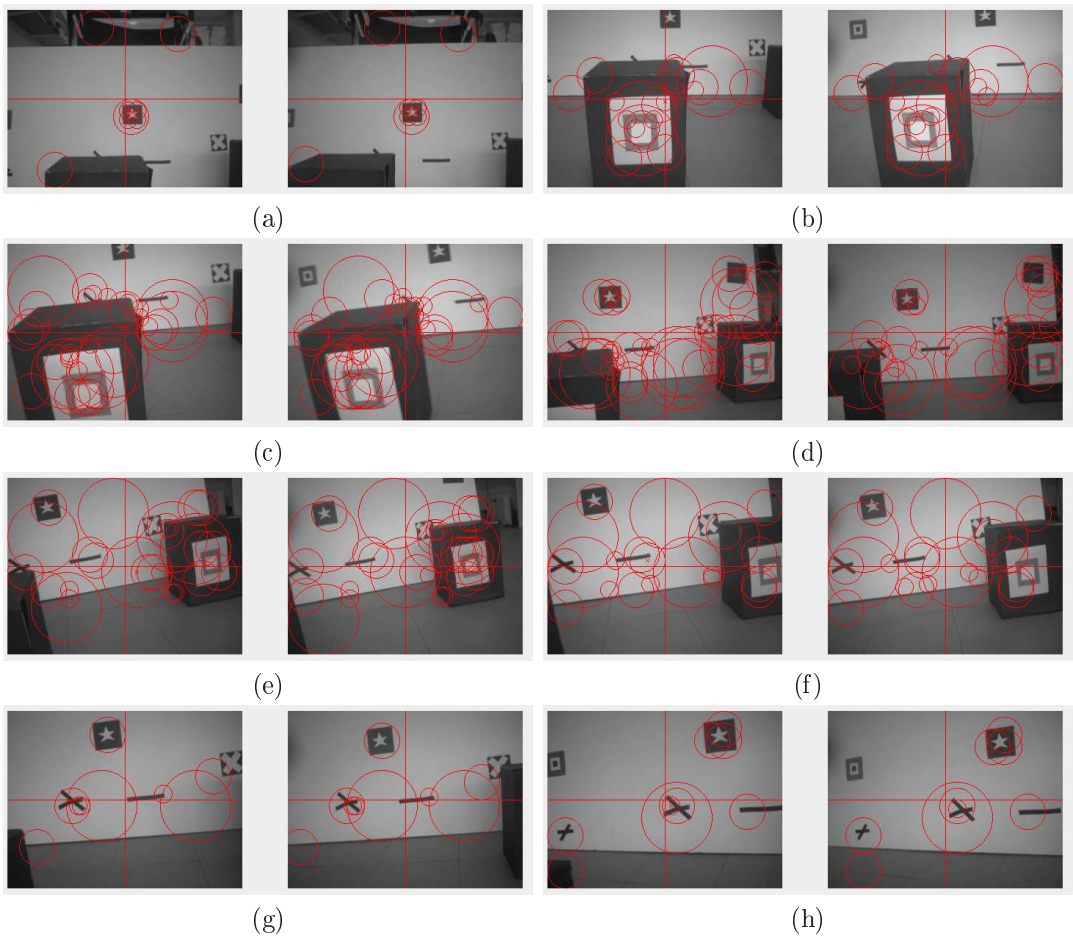



Figura 7.21: Vista de la escena desde el robot en el primer experimento de navegación

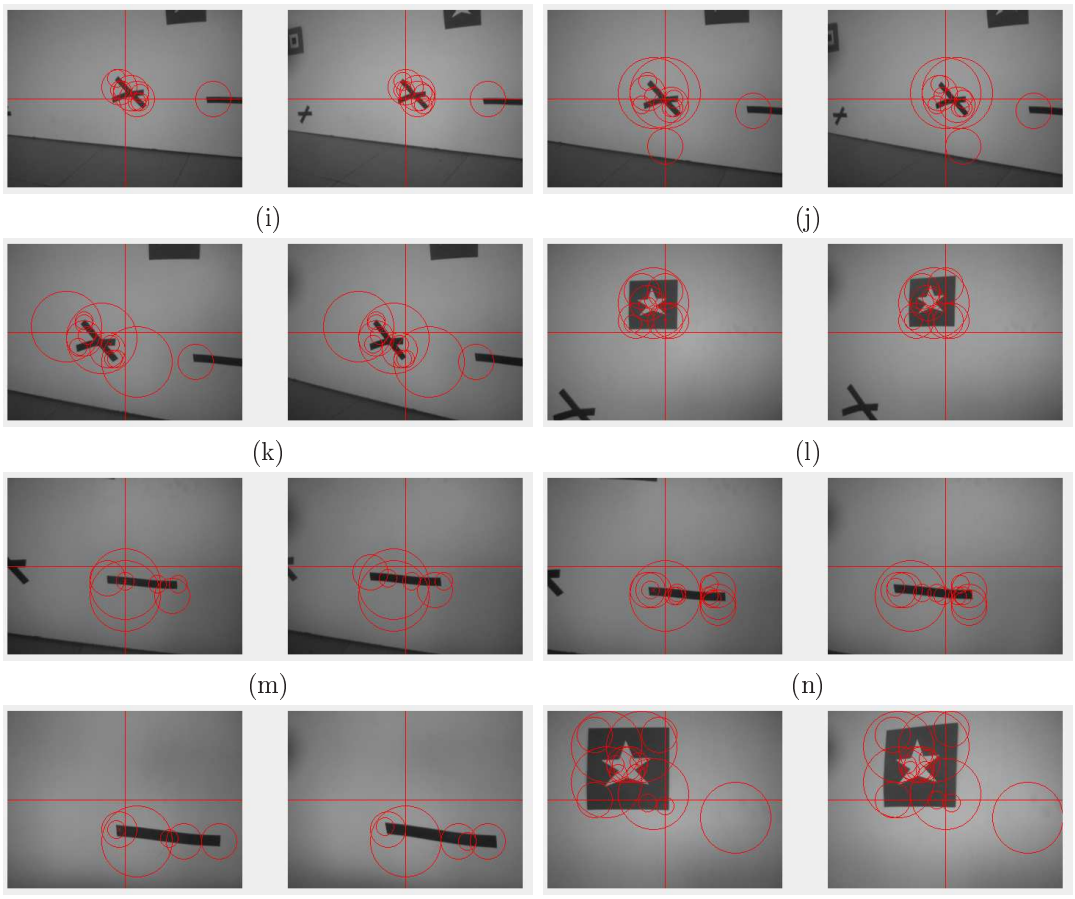
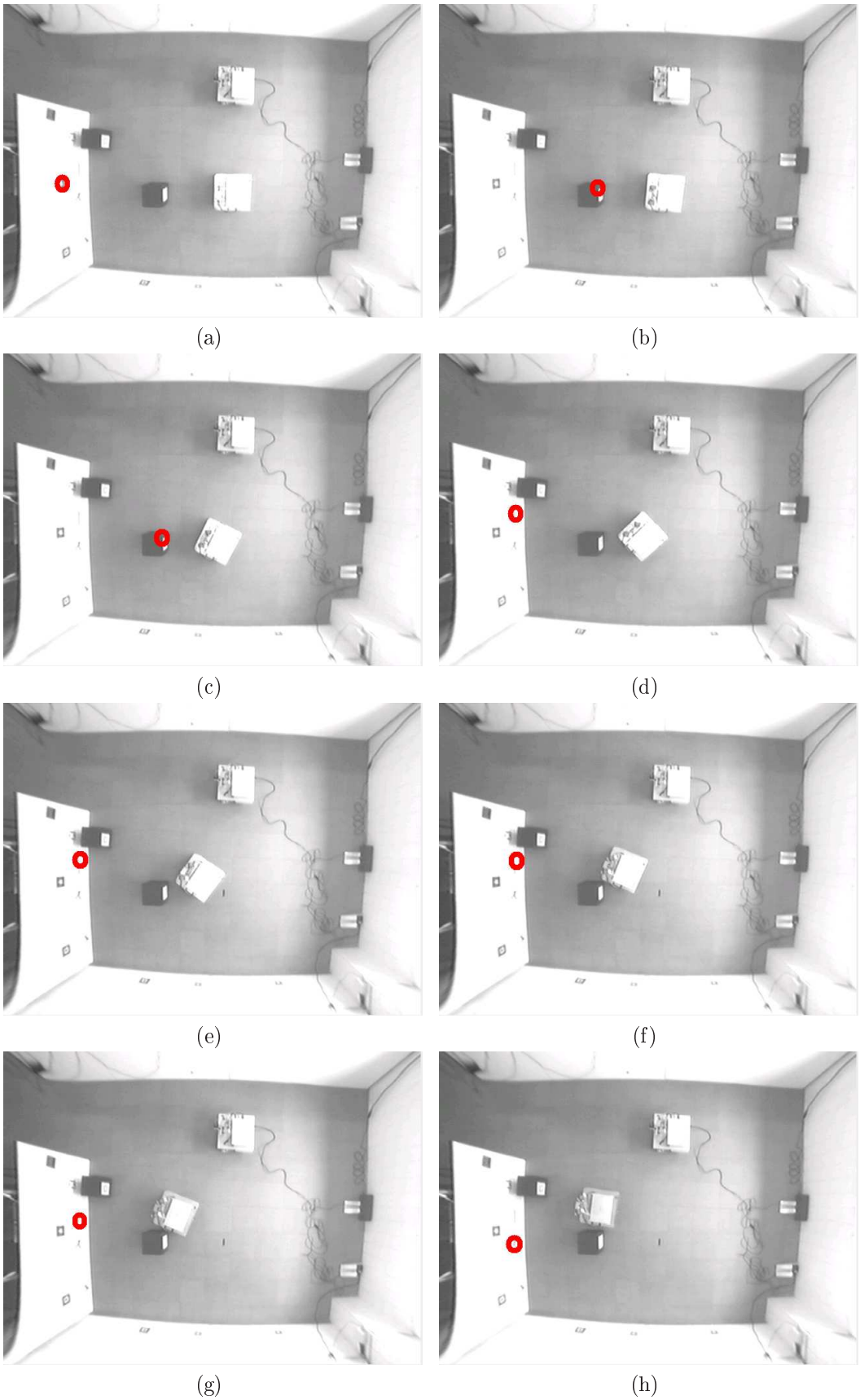
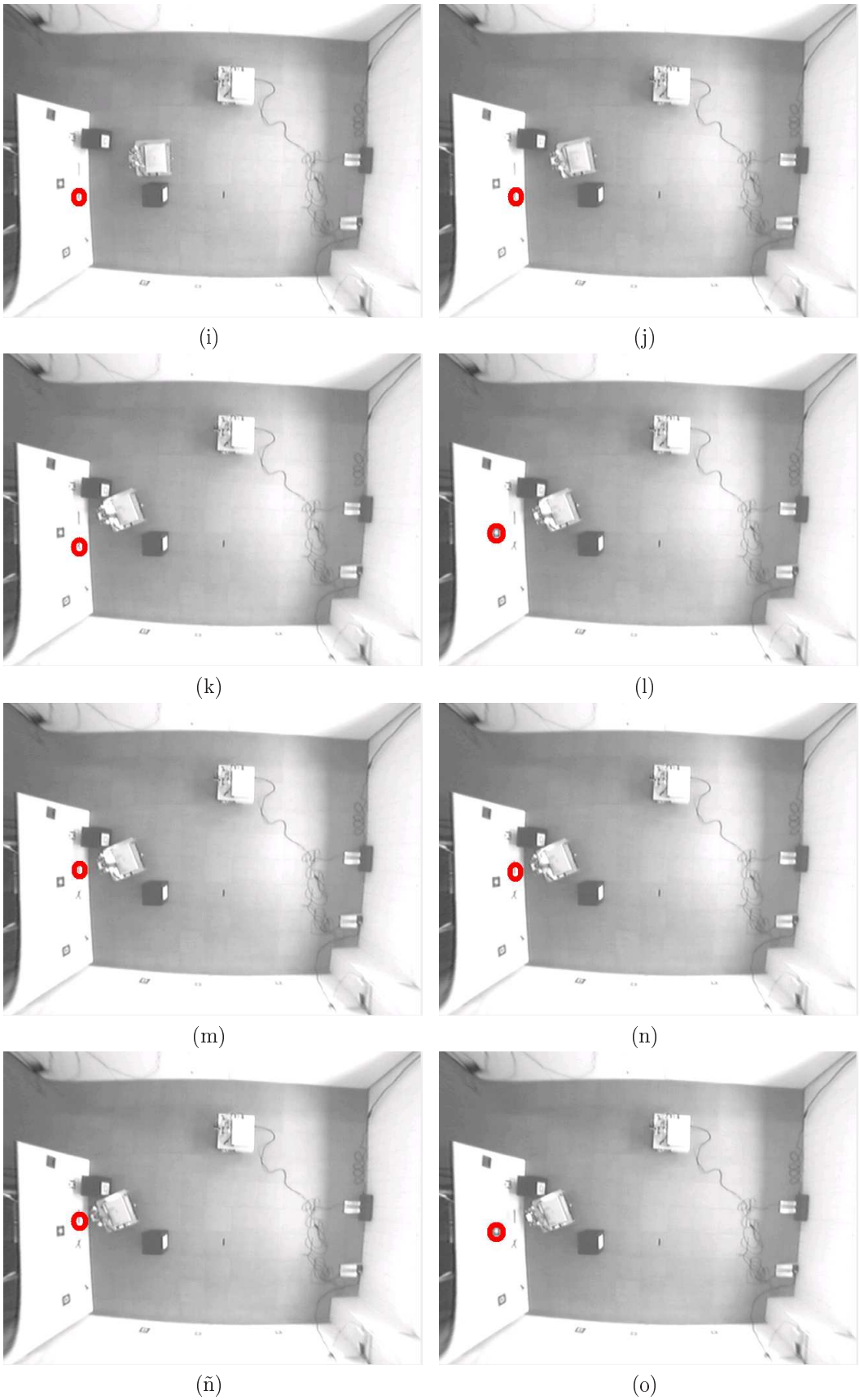



Figura 7.21: Vista de la escena desde el robot en el primer experimento de navegación



(a) (b) (c) (d) (e) (f) (g) (h)

Figura 7.22: Vista general de la escena en el primer experimento de navegación



(i) (j)

(k) (l)

(m) (n)

(ñ) (o)

Figura 7.22: Vista general de la escena en el primer experimento de navegación

### 7.3.2. Experimento 2: detección de suelo durante la tarea de navegación

En el segundo experimento, el robot debe pasar entre dos obstáculos para alcanzar la posición objetivo marcada por una baliza con forma de estrella. Tras los dos obstáculos se han situado una serie de objetos planos sobre el suelo que el robot debe detectar y rebasar sin evitarlos. Las figuras 7.23 y 7.24 muestran el resultado de esta prueba. En la secuencia de imágenes capturadas por el robot (figura 7.23) se han marcado en verde las zonas detectadas como objetos planos sobre el suelo. Recordemos que el cálculo de orientación sólo se realiza para un conjunto de regiones próximas al foco de atención, por lo que no todas las regiones que cumplen las condiciones de suelo son detectadas.

Como en el experimento anterior, en la situación inicial el robot se encuentra atendiendo a la baliza (imagen (a)) lo que permite poner en marcha una actividad de aproximación hacia su posición. La fijación sobre el primer obstáculo provoca acciones que permiten evitarlo produciendo pequeñas desviaciones en la trayectoria hacia el objetivo (imágenes (b)-(f)). La proximidad del segundo obstáculo da lugar a un cambio de atención que permite la fijación sobre la zona correspondiente del entorno. Esto se traduce, en el nivel de las acciones, en variaciones de la dirección de avance del robot que conllevan el correcto sorteo del obstáculo (imágenes (g)-(j)). Tras orientarse adecuadamente, la atención cambia de nuevo hacia zonas cercanas al objetivo (imágenes (k) y (l)), produciendo nuevas acciones de aproximación. El reinicio del ciclo de activación del selector de baliza permite que éste recupere el control atencional (imagen (m)) y que se actualice así la posición objetivo. Esto provoca nuevos cambios de atención, dirigidos por el selector de obstáculos, que dan lugar a la fijación de regiones próximas a su posición actual (imágenes (n) y (ñ)). En concreto, la atención se centra en las regiones asociadas con uno de los dos objetos planos situados sobre el suelo, el más posterior. El anterior no es seleccionado porque el alcance de los sensores no lo permite. No obstante, esta circunstancia no tiene porqué suponer un problema, dado que la imposibilidad para acceder a ese grupo de regiones puede interpretarse como ausencia de peligro en esa zona del entorno. Si se tratara de una zona que obstaculizara realmente, las regiones asociadas habrían sido seleccionadas en instantes previos proporcionando el comportamiento adecuado. La fijación atencional sobre la nueva zona (imágenes (ñ)-(s)) da lugar a la correcta detección de las regiones seleccionadas como partes del propio suelo, produciendo un tratamiento adecuado de la situación a través de avances hacia la posición objetivo. Tras un nuevo cambio atencional hacia la baliza (imagen (t)) se producen nuevas

aproximaciones (imágenes (u) y (v)) que permiten alcanzar la posición final (imagen (w)).

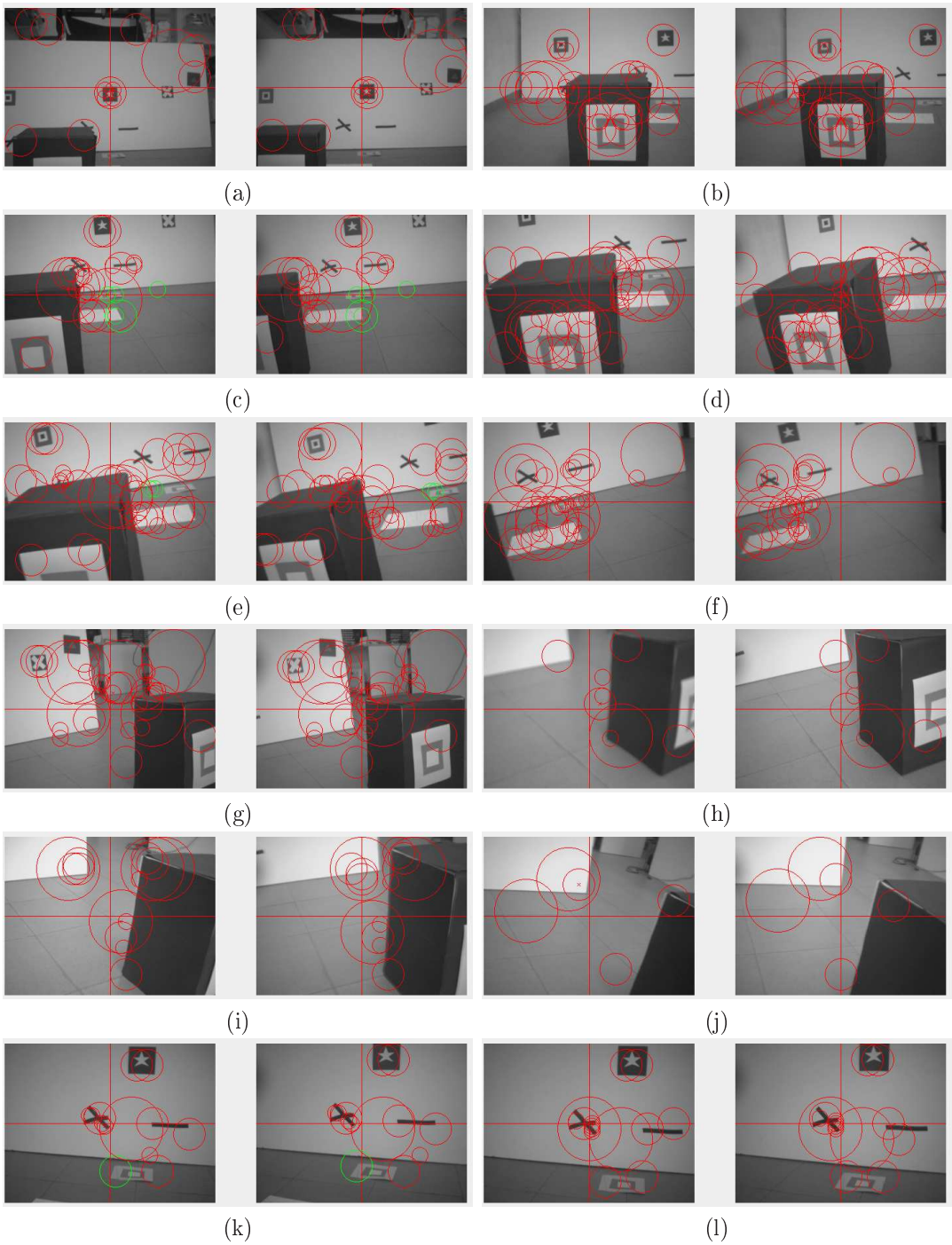



Figura 7.23: Vista de la escena desde el robot en el segundo experimento de navegación

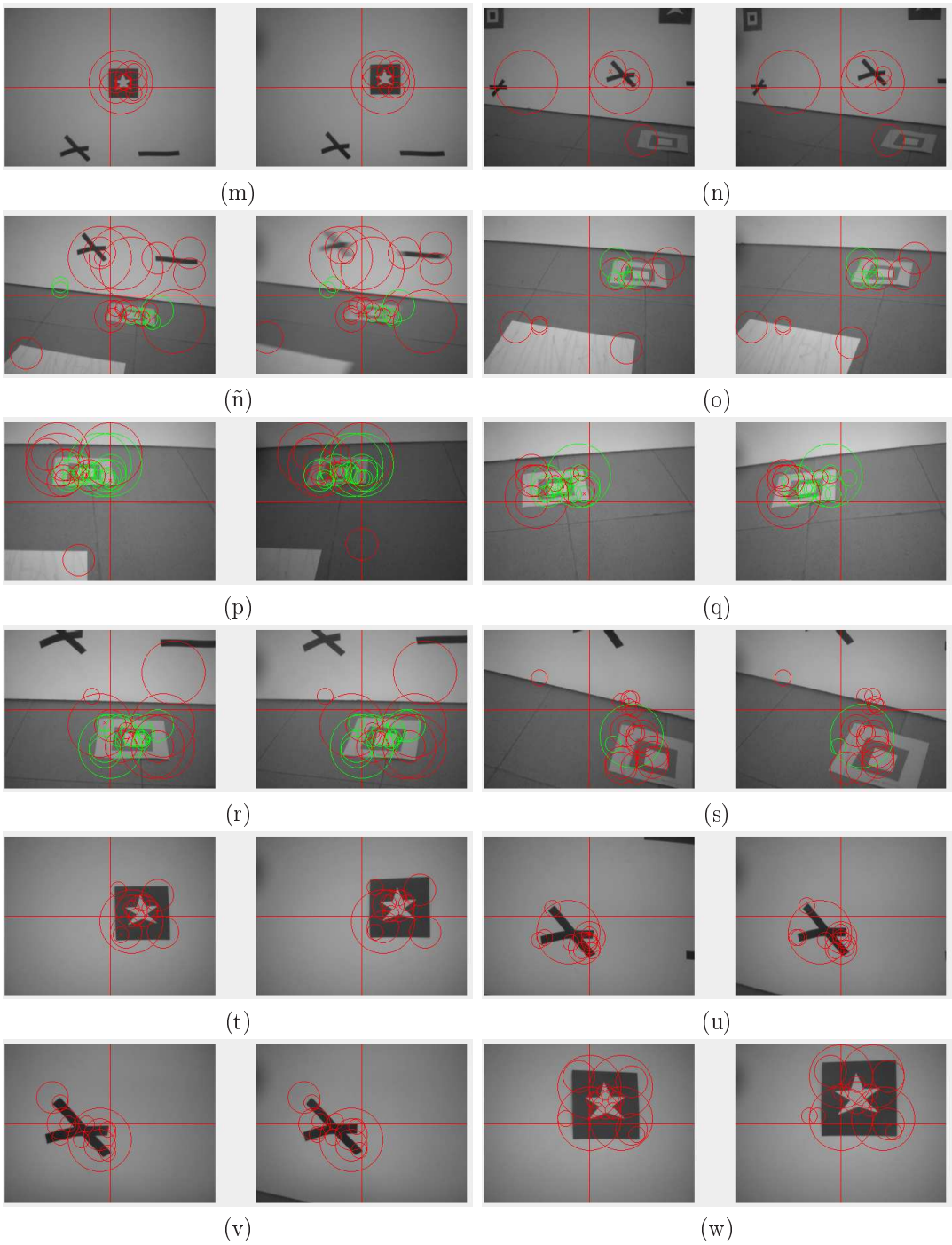
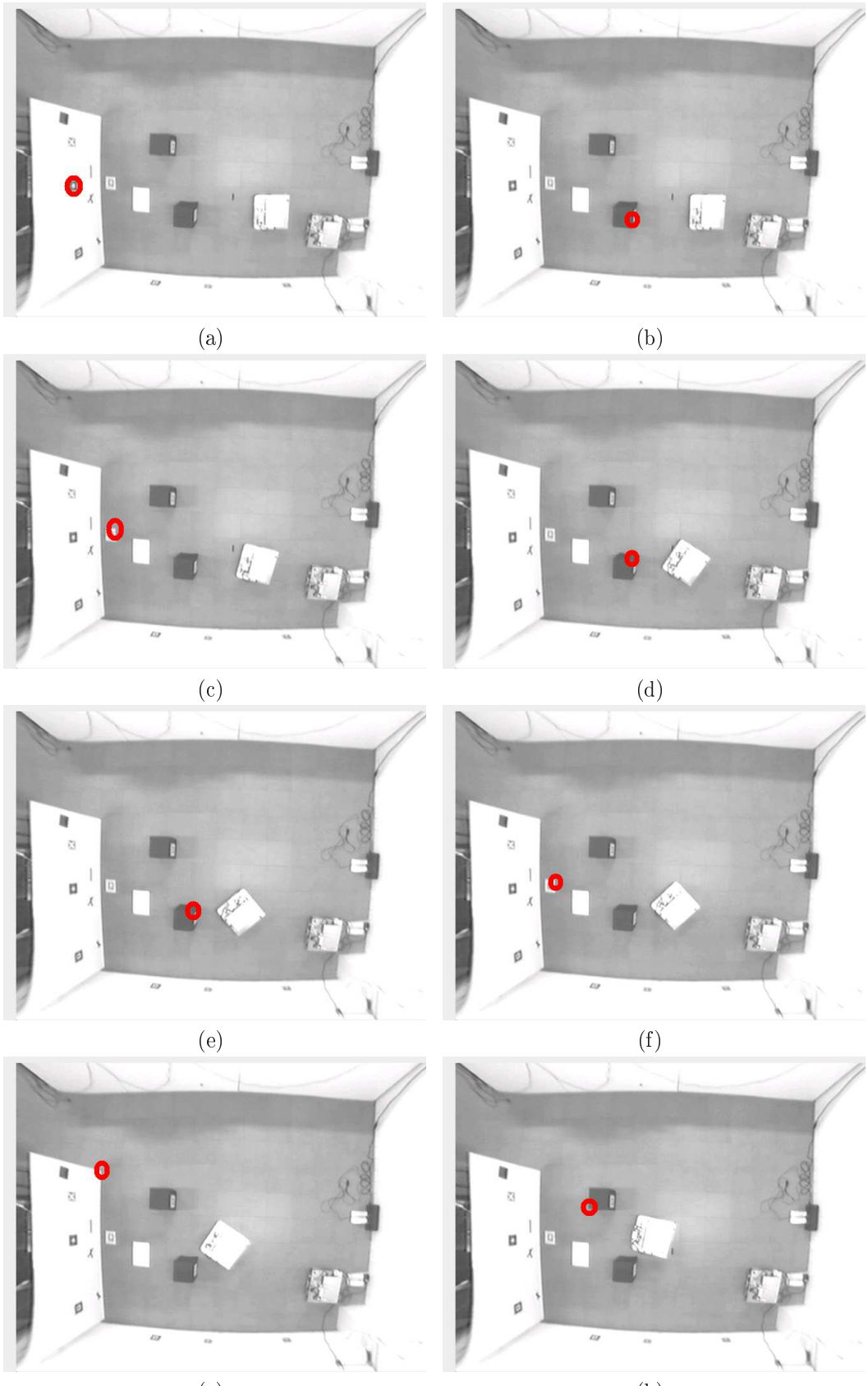
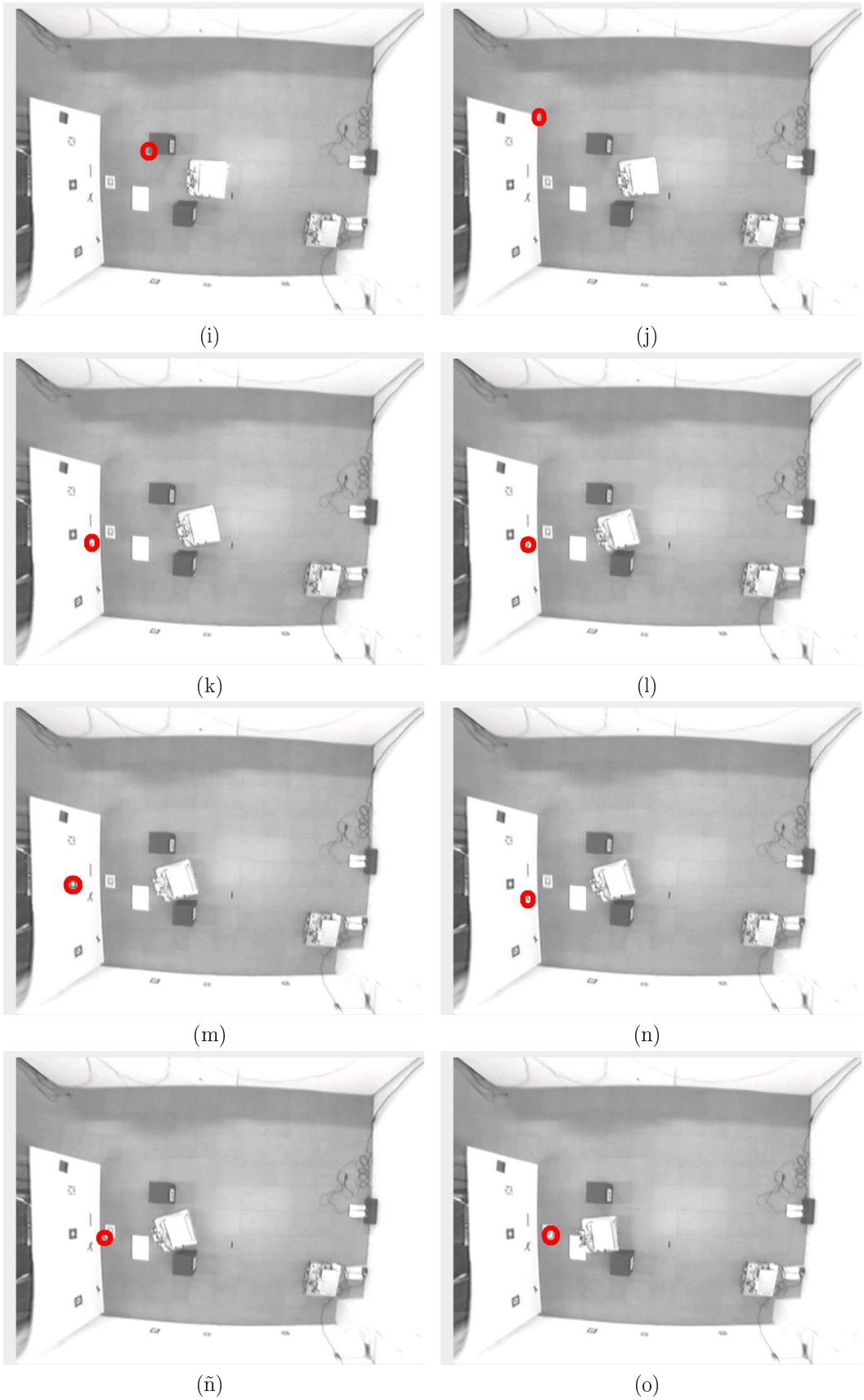



Figura 7.23: Vista de la escena desde el robot en el segundo experimento de navegación



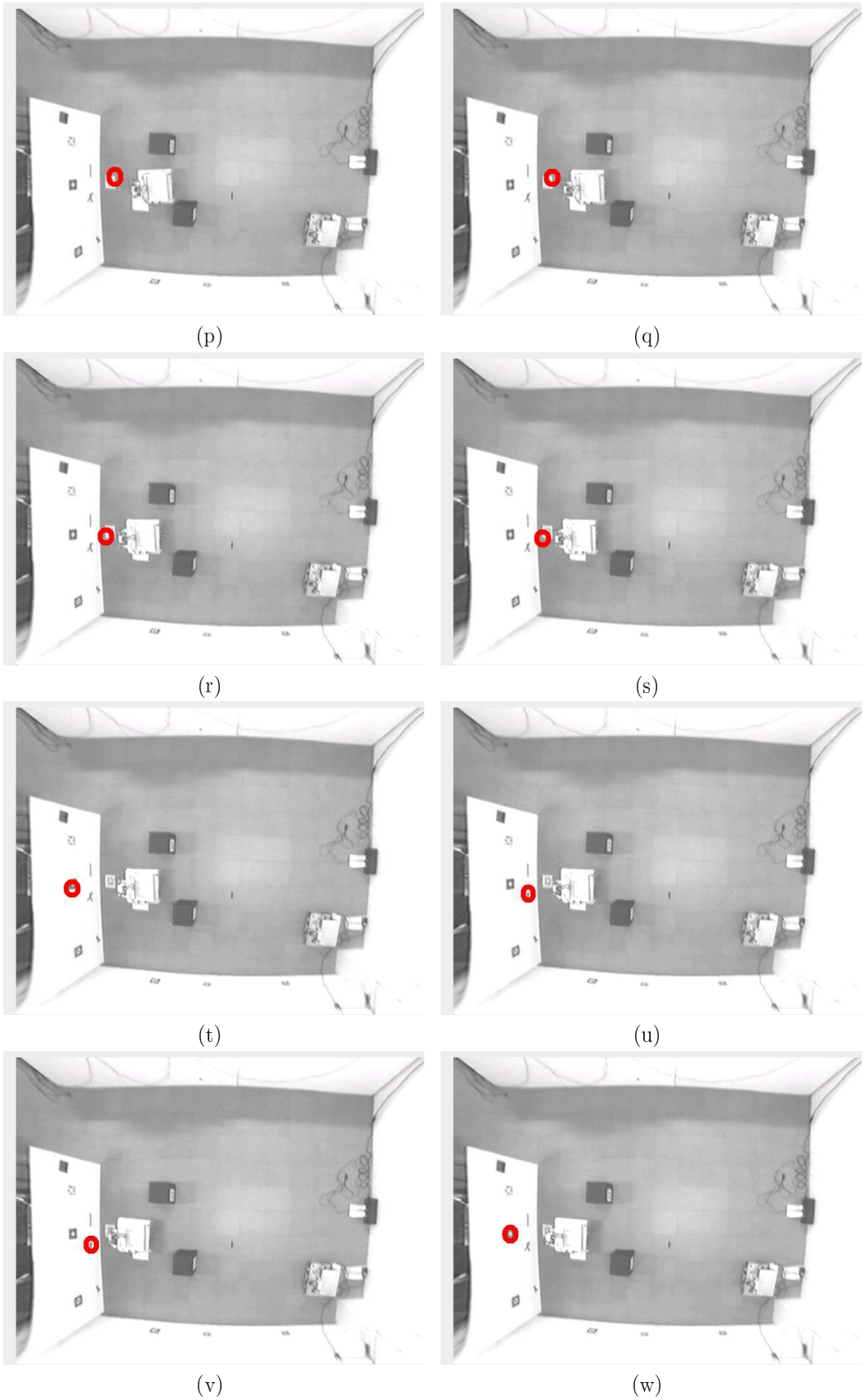
(a) (b) (c) (d) (e) (f) (g) (h)

Figura 7.24: Vista general de la escena en el segundo experimento de navegación



(i) (j) (k) (l) (m) (n) (ñ) (o)

Figura 7.24: Vista general de la escena en el segundo experimento de navegación



(p) (q)

(r) (s)

(t) (u)

(v) (w)

Figura 7.24: Vista general de la escena en el segundo experimento de navegación

### 7.3.3. Experimento 3: navegación con varias balizas de orientaciones similares

Los dos últimos experimentos se centran en el análisis del comportamiento del robot en tareas de navegación con varias balizas.

En la primera situación, existen dos balizas que el robot debe alcanzar secuencialmente. Ambas balizas se encuentran en direcciones visuales próximas, lo que permite la rápida localización de la segunda una vez alcanzada la primera. Las imágenes 7.25 y 7.26 muestran los resultados obtenidos. La primera baliza está formada por 3 cuadrados concéntricos de diferentes colores. La segunda es la baliza de estrella utilizada en los dos experimentos anteriores. Al inicio del experimento, el robot mantiene la fijación atencional sobre la primera baliza (imagen (a)). La ausencia de obstáculos permite un acercamiento continuo hacia la posición objetivo (imágenes (b)-(g)). Tras alcanzar dicha posición (imagen (g)), el selector de baliza es reprogramado para conseguir la localización de la segunda baliza. Al no existir ninguna región visible compatible con las propiedades de esta segunda baliza, se pone en marcha un comportamiento exploratorio guiado por el selector atencional de *zona no atendida* (imagen (h)). El comportamiento atencional implicado permite el acceso visual a la segunda baliza (imagen (k)) una vez recorridas otras zonas del entorno (imágenes (i) y (j)). A partir de ese instante, la primera baliza se concibe como un obstáculo que el robot debe evitar para alcanzar la nueva posición objetivo. El selector de obstáculos proporciona esta interpretación de la situación fijando de nuevo la atención sobre las regiones asociadas con la primera baliza, aunque ahora para propósitos conductuales diferentes (imágenes (l)-(n)). Tras orientarse adecuadamente para evitar la colisión con el obstáculo en su camino a la nueva baliza (imagen (ñ)), el control atencional se centra en regiones próximas al objetivo. Esto da lugar a una serie de acciones de acercamiento (imágenes (o)-(r)), que se interrumpen en un instante dado por la adquisición del control del selector de balizas (imagen (s)). Tras la renovación de la posición objetivo y ante la falta de zonas de obstaculización, se producen nuevas aproximaciones (imágenes (t)-(v)) que permiten finalmente alcanzar la posición deseada (imagen (w)).

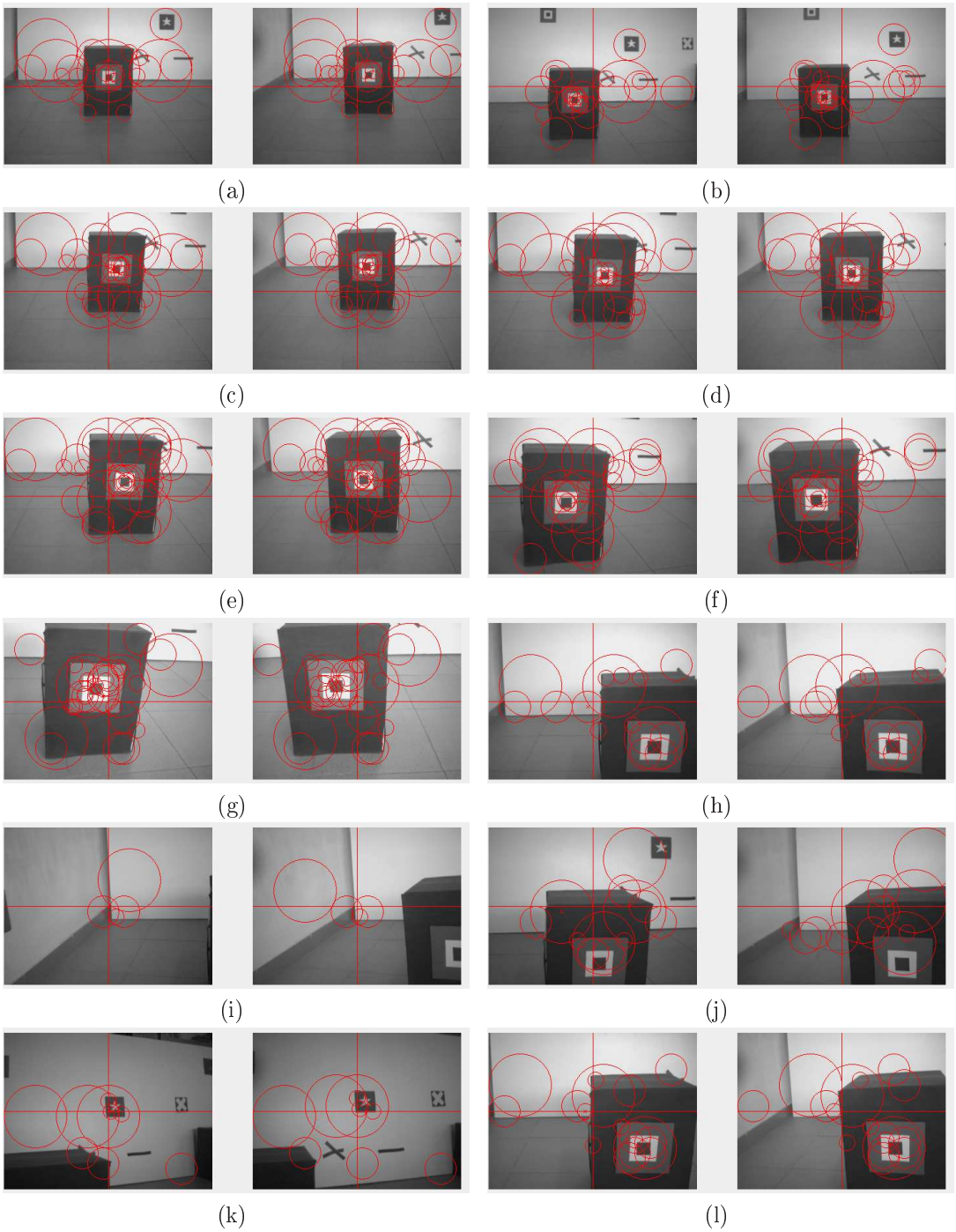



Figura 7.25: Vista de la escena desde el robot en el tercer experimento de navegación

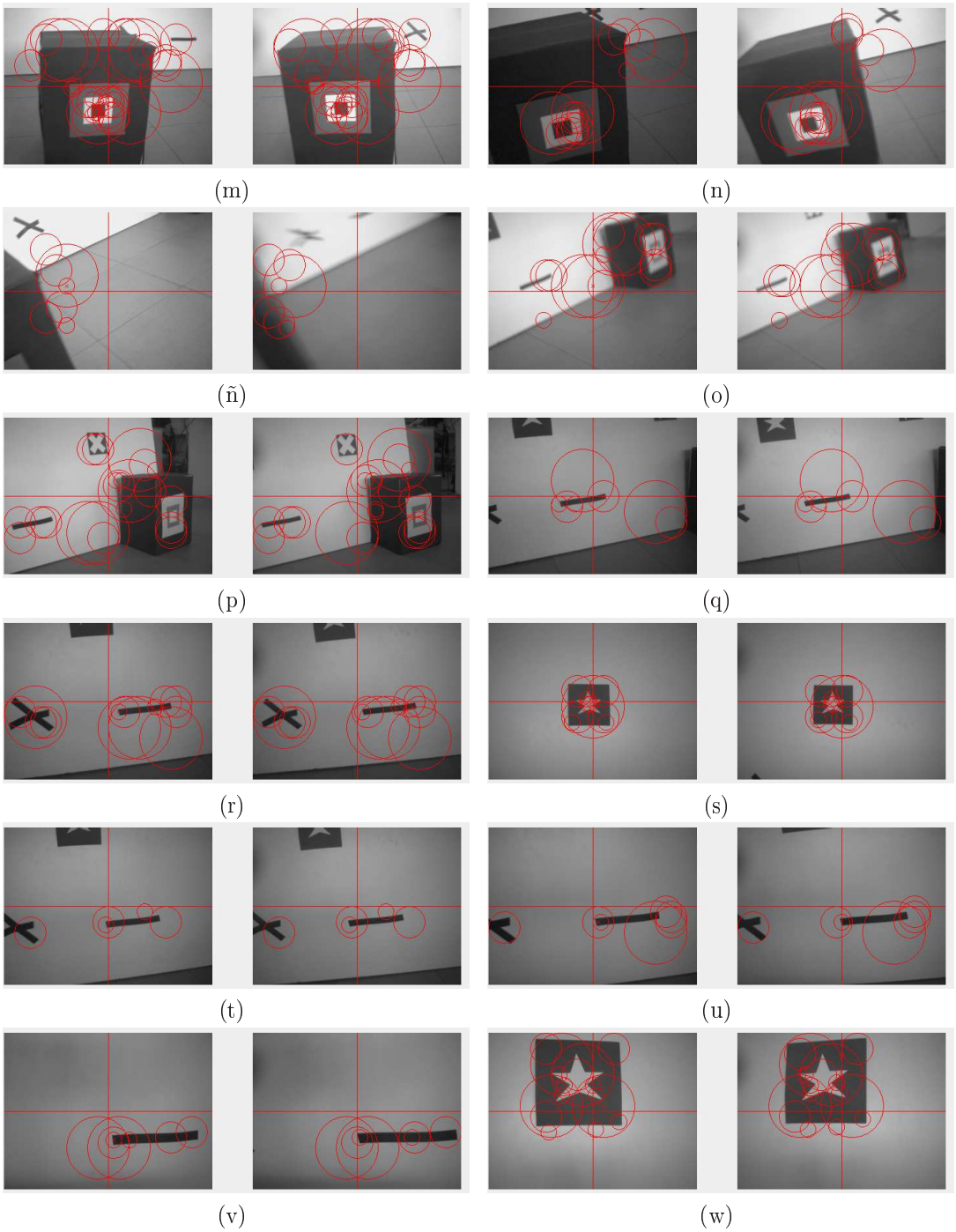
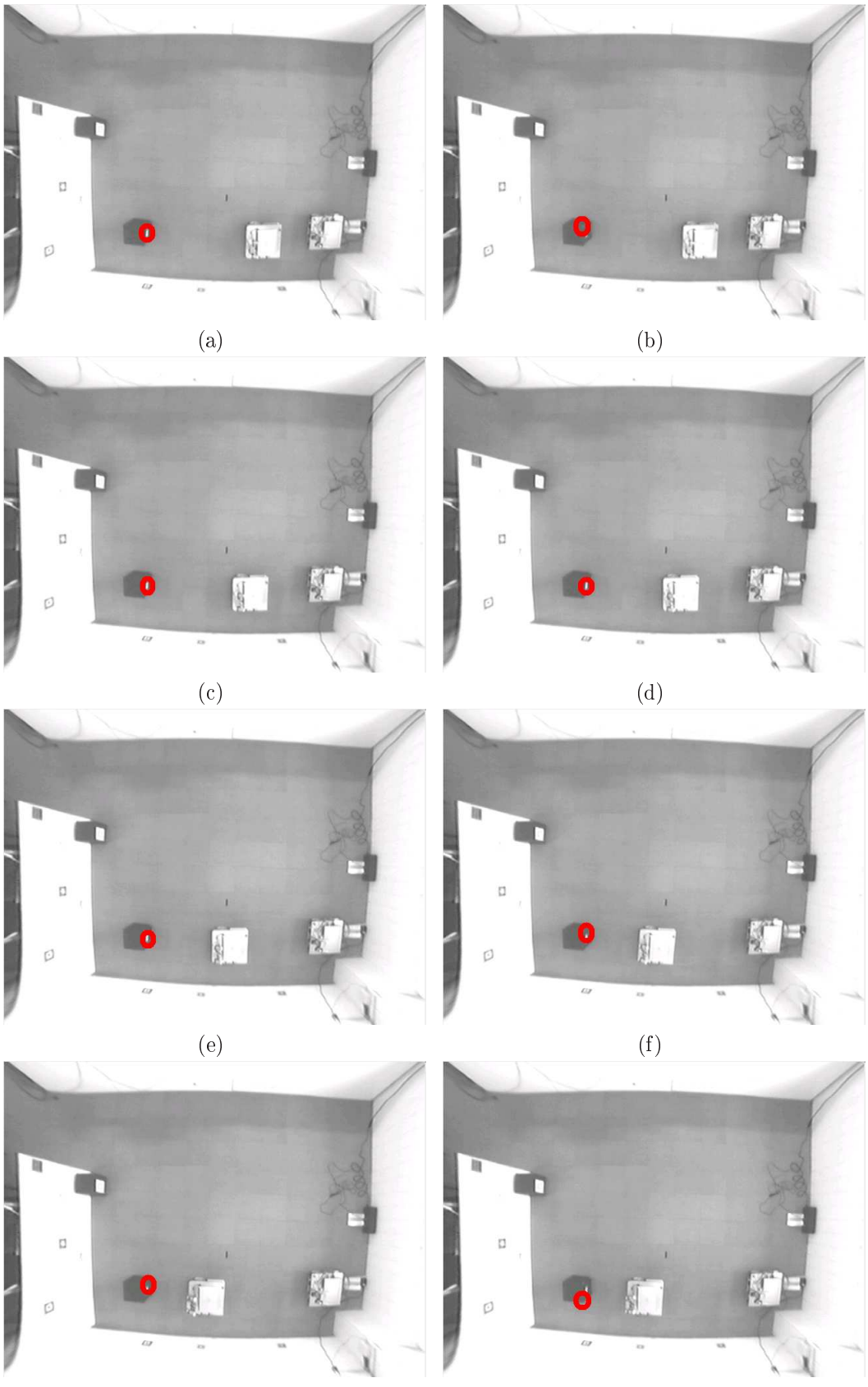
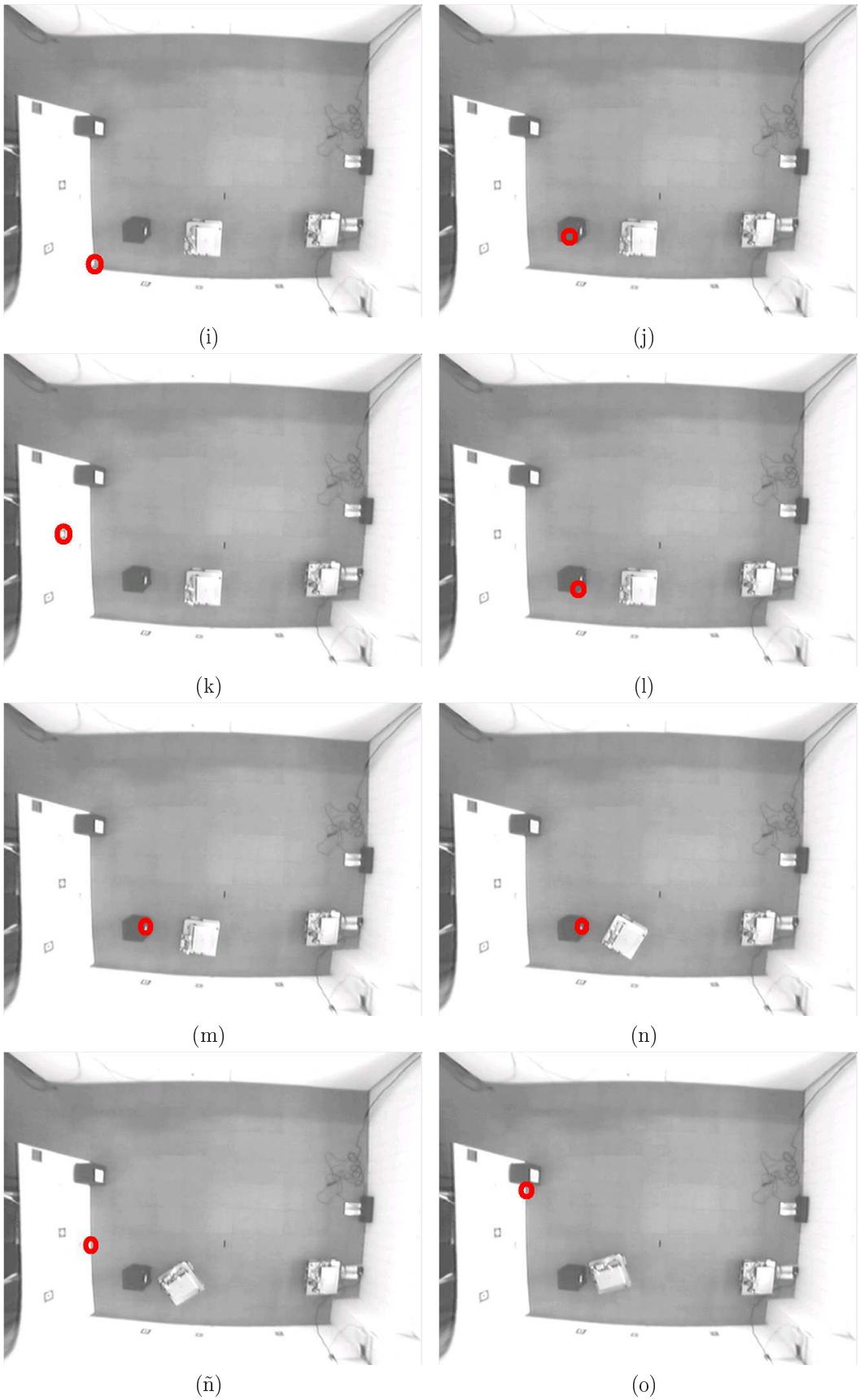



Figura 7.25: Vista de la escena desde el robot en el tercer experimento de navegación



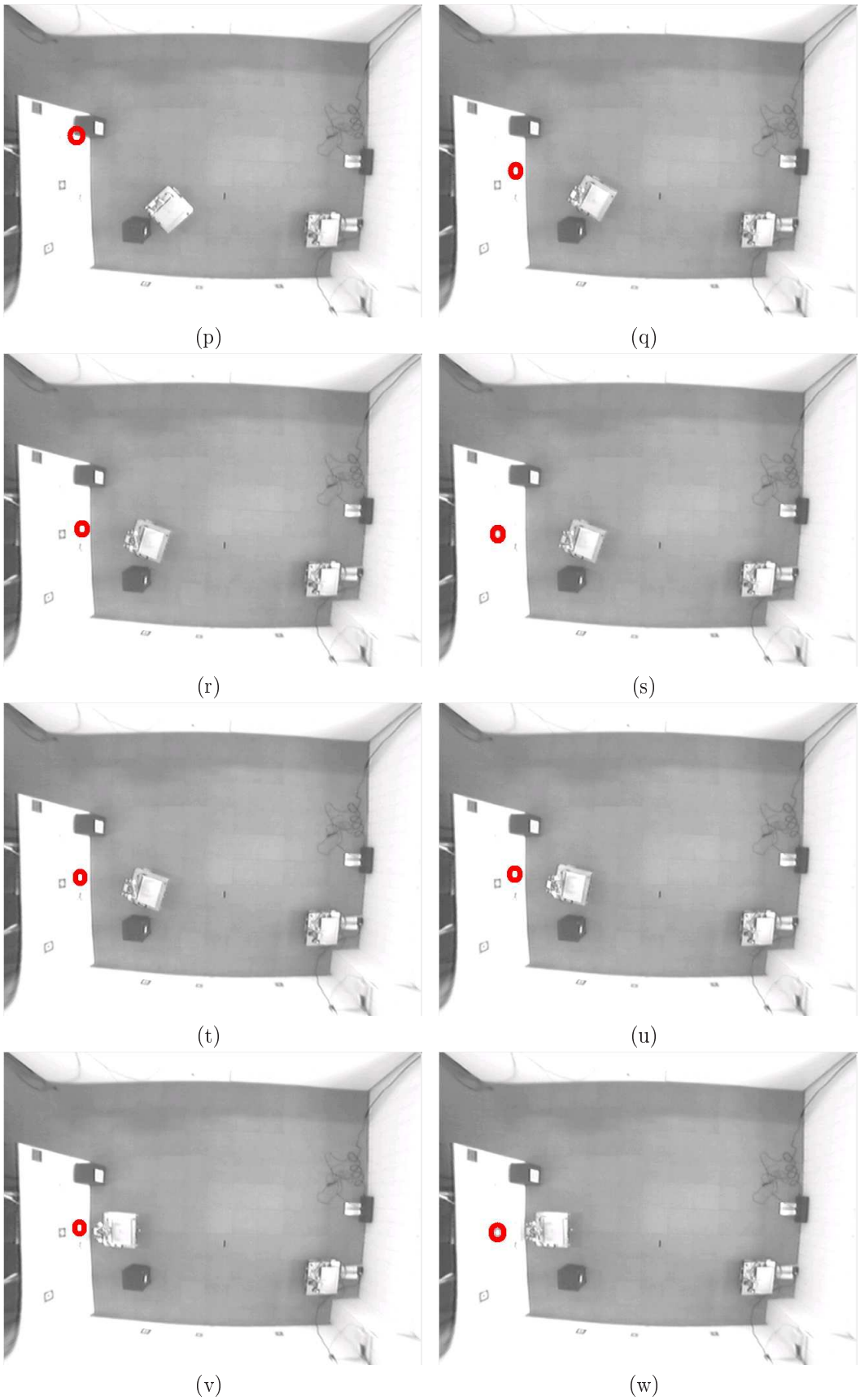
(a) (b) (c) (d) (e) (f) (g) (h)

Figura 7.26: Vista general de la escena en el tercer experimento de navegación



(i) (j) (k) (l) (m) (n) (ñ) (o)

Figura 7.26: Vista general de la escena en el tercer experimento de navegación



(p) (q)

(r) (s)

(t) (u)

(v) (w)

Figura 7.26: Vista general de la escena en el tercer experimento de navegación

### 7.3.4. Experimento 4: navegación con varias balizas de orientaciones dispares

Este último experimento describe el comportamiento del sistema en una situación en la que, como en el experimento anterior, el robot debe alcanzar dos balizas en secuencia. En esta ocasión, las balizas se han situado en posiciones que requieren una exploración más extensa del entorno para lograr su localización. Los resultados de esta prueba pueden observarse en las figuras 7.27 y 7.28.

Partiendo de una situación en la que la fijación atencional se centra en la primera baliza (imagen (a)), comienza la secuencia de acciones que permiten la aproximación a este primer objetivo. La orientación inicial del robot provoca la selección atencional de zonas que, si bien no suponen un obstáculo en la trayectoria hacia el objetivo, se encuentran en posiciones cercanas a su zona de paso (imágenes (b)-(f)). Una vez que el robot se ha situado en la dirección del objetivo (imagen (g)), éste es seleccionado produciendo acciones de acercamiento más precisas (imágenes (g)-(i)). Tras alcanzar una posición próxima al primer objetivo (imagen (i)), se produce el cambio de baliza provocando la exploración visual del entorno. La situación de la segunda baliza con respecto a la orientación actual del robot da lugar a una búsqueda de mayor duración que la descrita en el experimento anterior. El control atencional asociado con el comportamiento exploratorio amplía paulatinamente la zona de búsqueda (imágenes (j)-(o)) hasta que finalmente permite la localización de la segunda baliza (imagen (p)). La proximidad del primer objetivo provoca su selección (imagen (q)) y a partir de ese instante se suceden las acciones que proporcionan el acercamiento al segundo objetivo (imágenes (q)-(t)). La reducida distancia entre los dos objetivos produce que, tras breves instantes desde el comienzo de la acción de aproximación hacia la segunda baliza, el objetivo sea seleccionado de nuevo (imagen (u)). Esta nueva selección permite que el robot corrija su orientación (imagen (v)), logrando así el alcance de la posición final (imagen (w)).

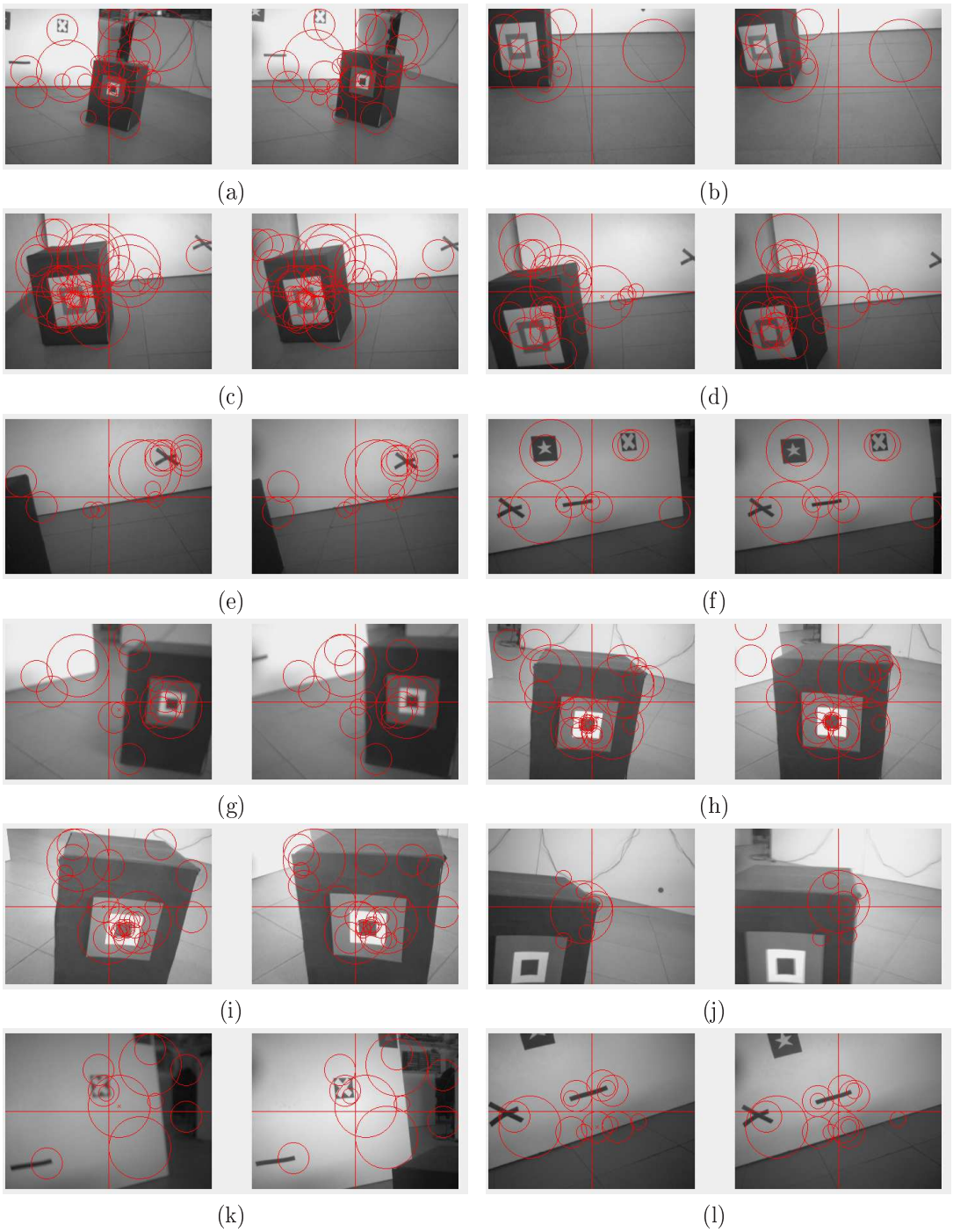



Figura 7.27: Vista de la escena desde el robot en el cuarto experimento de navegación

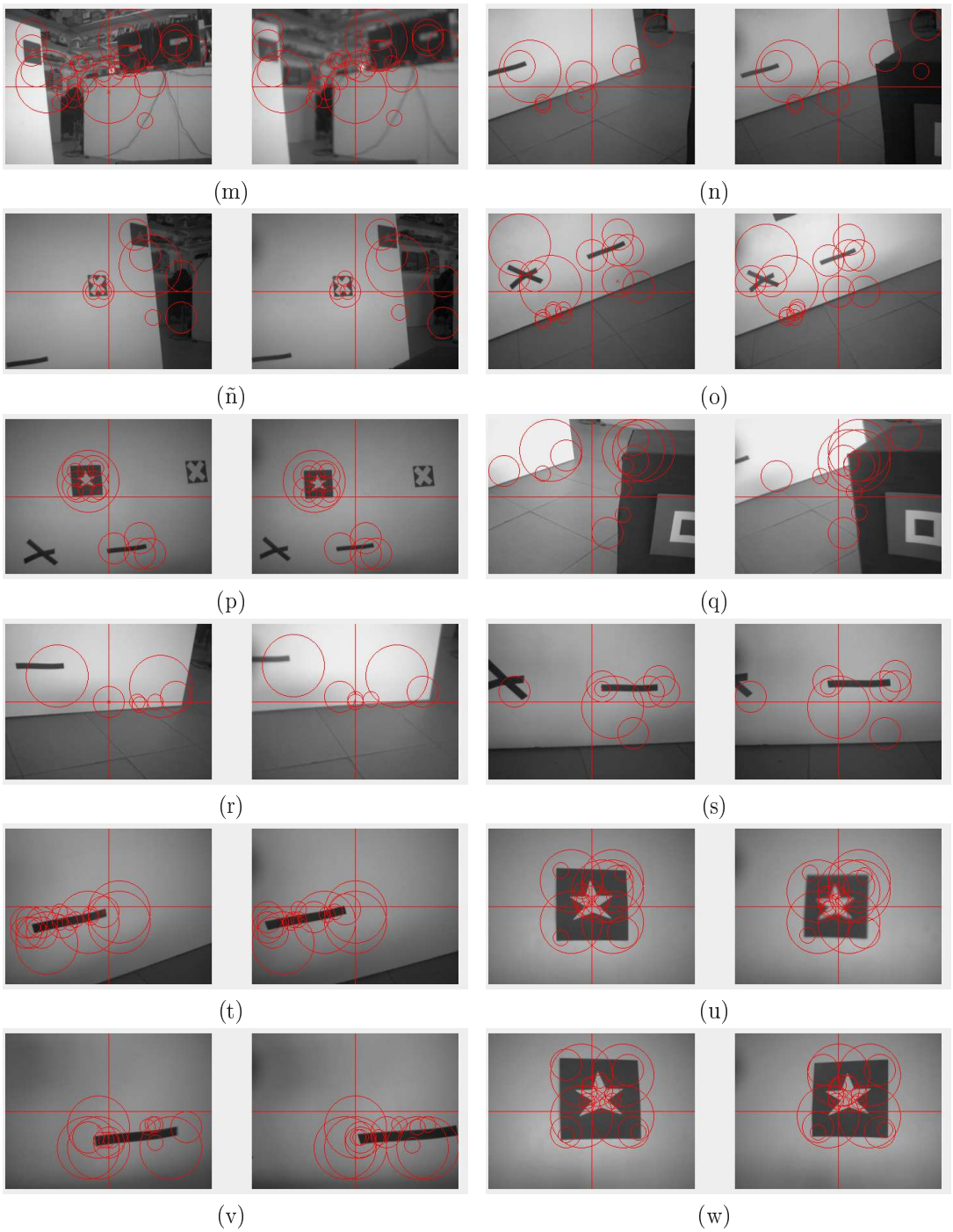



Figura 7.27: Vista de la escena desde el robot en el cuarto experimento de navegación

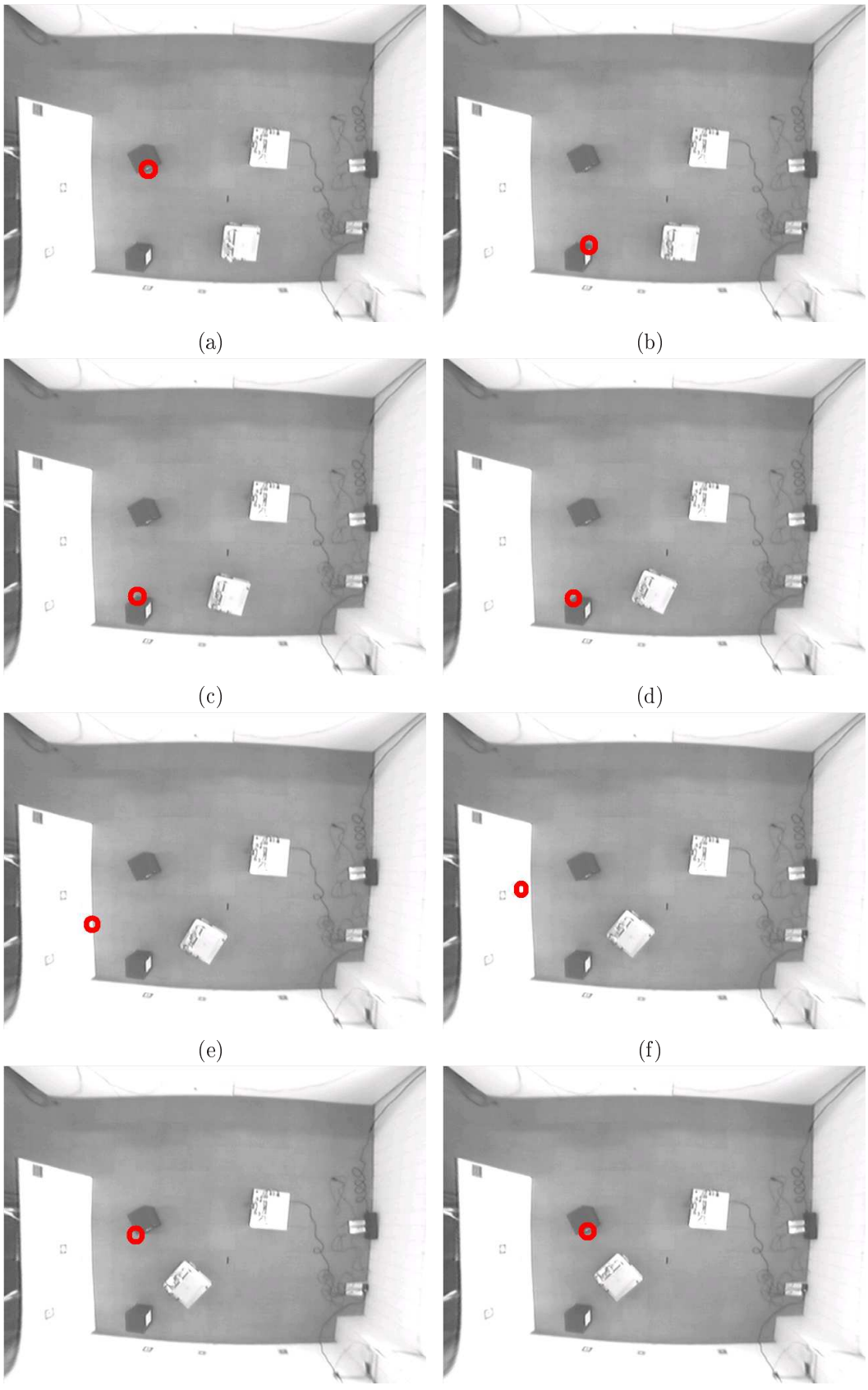
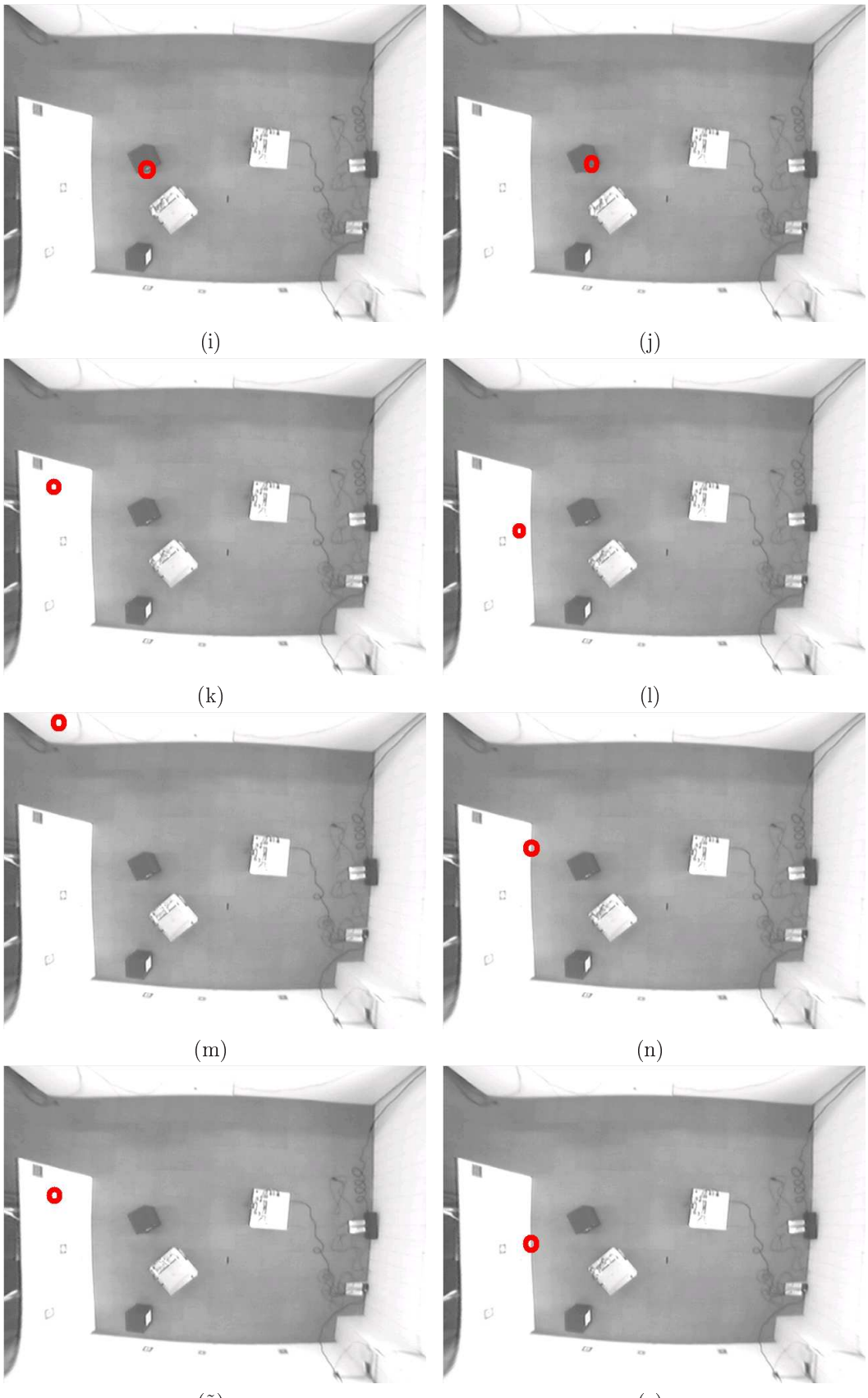


Figura 7.28: Vista general de la escena en el cuarto experimento de navegación



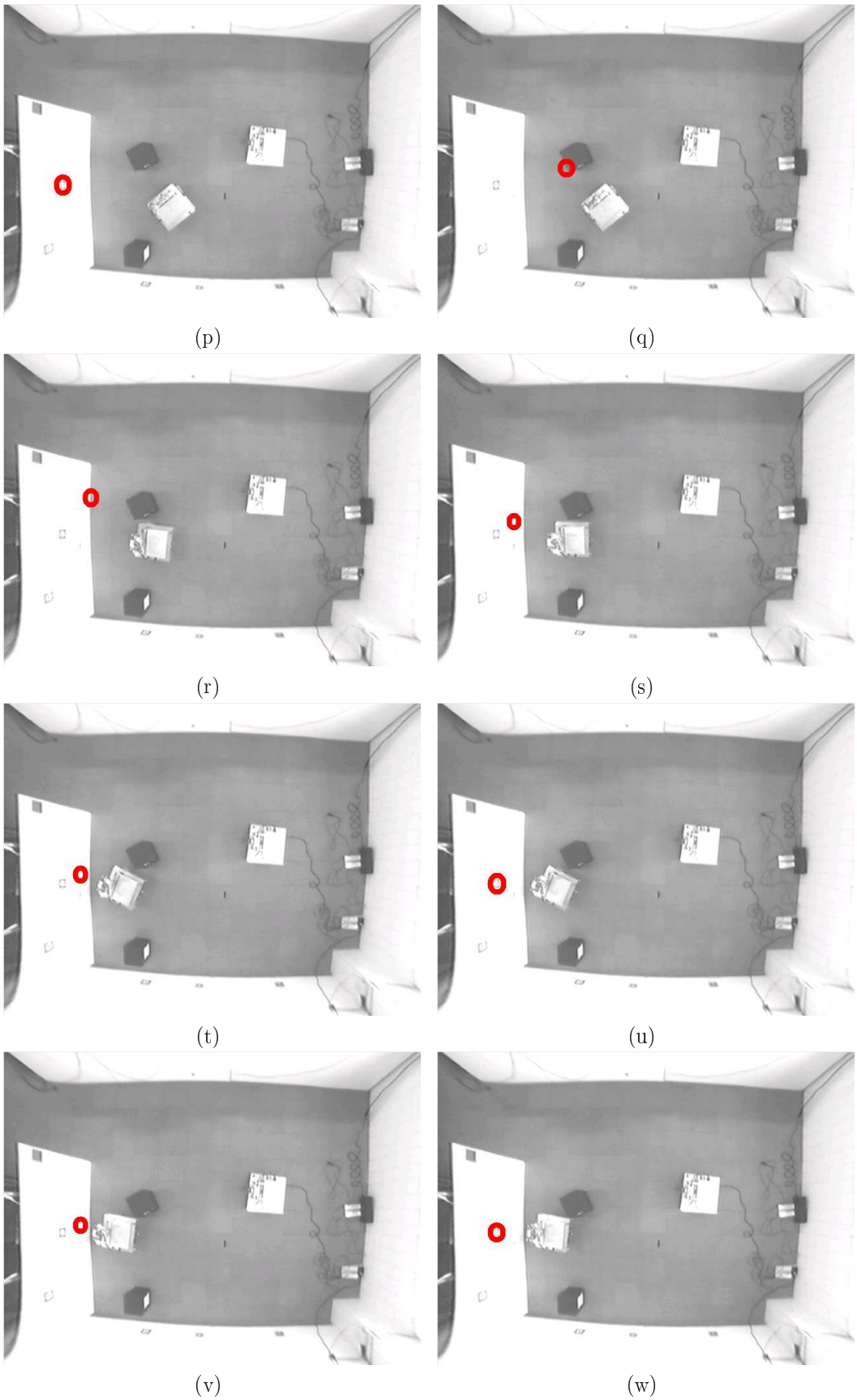
(i) (j)

(k) (l)

(m) (n)

(ñ) (o)

Figura 7.28: Vista general de la escena en el cuarto experimento de navegación



(p) (q)

(r) (s)

(t) (u)

(v) (w)

Figura 7.28: Vista general de la escena en el cuarto experimento de navegación

# Capítulo 8

# Conclusiones

Para cerrar esta memoria, se presentan en este capítulo las conclusiones principales de esta tesis y se proponen diferentes líneas de continuación y mejora del trabajo desarrollado.

## 8.1. Conclusiones y aportaciones principales

Esta tesis aborda el estudio de la atención como mecanismo de conexión entre la percepción visual y el control autónomo de robots. Para centrar las cuestiones fundamentales de este estudio, se ha propuesto un sistema de control basado en la atención que permite generar comportamiento autónomo en un robot móvil a partir de la definición de relaciones entre atención y acción. Las bases principales de nuestro enfoque son:

- *La atención actúa como mecanismo de selección para la acción*: la atención dirige el proceso perceptivo en función de las necesidades de actuación, seleccionando la información sensorial relevante para ejecutar la acción.
- *La atención proporciona un medio de selección de la acción*: la selección atencional restringe las posibles acciones que pueden llevarse a cabo en cada situación. La secuencia de estímulos proporcionada por el mecanismo atencional provoca la ejecución secuencial de las acciones, actuando así como medio de selección de la acción.

Para explotar estas ideas se ha desarrollado un sistema de atención visual que rompe con la línea clásica de la hipótesis de capacidad limitada adoptada por la mayoría de los modelos computacionales surgidos en los últimos años. A diferencia de las propuestas existentes, en nuestro sistema la atención y la acción se encuentran estrechamente relacionadas siguiendo

las dos líneas expuestas anteriormente. El enlace entre atención y acción se hace efectivo a partir de una serie de consideraciones sobre el sistema atencional:

- *Especificaciones descendentes*: la atención debe ser modulada desde los comportamientos en función de las necesidades de actuación. No basta, por lo tanto, con considerar las propiedades del mundo visual. Es necesario incluir la influencia de la acción en la selección atencional.
- *Control distribuido*: el control atencional no debe estar centralizado, sino que debe encontrarse distribuido en múltiples unidades individuales de control que permitan mantener varios objetivos visuales simultáneamente. Esto da lugar a que la atención pueda ser modulada desde múltiples comportamientos con diferentes necesidades de información.
- *Atención abierta y encubierta*: el sistema debe proporcionar control abierto y encubierto de la atención, permitiendo así que la selección atencional alterne entre múltiples objetivos en función de las propiedades externas del mundo y de las necesidades internas del robot.

De todas estas consideraciones nace el sistema de atención visual propuesto que se plantea como una colección de componentes funcionales que colaboran para llevar a cabo la selección atencional. Cada componente permite resolver una parte del proceso atencional que comprende la siguientes fases: detección y mantenimiento de regiones visuales; extracción de propiedades de las regiones detectadas; selección de múltiples objetivos visuales mediante varios componentes de control modulados desde distintos comportamientos; selección atencional abierta sobre un único objetivo en función de las necesidades conductuales; fijación binocular del objetivo. Para cada una de estas fases se han propuesto métodos específicos que constituyen una parte importante de las aportaciones de esta tesis:

1. Se ha propuesto un método de detección de regiones en el que se plantea un tratamiento desigual de las diferentes zonas de la imagen en función de su excentricidad. La propuesta consiste en la aplicación del método Harris-Laplace sobre un prisma multi-escala, que permite limitar el tamaño mínimo de las regiones que son detectadas en cada parte de la imagen. Como resultado, el proceso proporciona un efecto de detalle en la parte central de la imagen y de información más general en la periferia, simulando la estructura de una superficie retiniana y proporcionando una reducción significativa de los tiempos de procesamiento en relación al método original.

2. La fase dedicada a la extracción de propiedades se ha separado en dos bloques que proporcionan dos flujos de procesamiento paralelos e independientes. Estos dos flujos de procesamiento se corresponden con la separación entre el "qué" y el "cómo" propuesta en la neurociencia. El subsistema dedicado al "qué" extrae propiedades de aspecto de las regiones a partir de varios descriptores basados en histogramas que muestran propiedades de invariabilidad ante diferentes condiciones visuales. El subsistema relacionado con el "cómo" se encarga de calcular diversas propiedades espaciales de las regiones visuales, tales como posición 3D u orientación. Dentro de este segundo grupo de propiedades, cabe destacar el método propuesto de estimación de orientaciones en superficies planas. El método consiste en suponer diferentes orientaciones y confirmar o desechar cada hipótesis a partir de la proyección proporcionada por la transformación homográfica correspondiente. La aplicación recursiva de este proceso permite reducir, en cada iteración, el rango de posibles orientaciones proporcionando finalmente una aproximación de la orientación real. La mayor ventaja del método propuesto es su independencia del grado de textura de las regiones sobre las que se aplica, proporcionando resultados coherentes en casos donde no es posible estimar directamente la transformación homográfica.

3. La propuesta distribuida del control atencional se ha materializado a través de la definición de un grupo de componentes del sistema, a los que hemos denominado selectores de objetivo. Cada selector de objetivo es modulado por un comportamiento de alto nivel de manera que éste debe responder centrando la atención, a una frecuencia determinada, sobre zonas del entorno que presenten propiedades coherentes con los objetivos conductuales. La integración de las especificaciones descendentes con las propiedades de las regiones visuales del entorno da lugar a un mapa de saliencia que permite determinar la zona de mayor interés en la selección atencional. Para llevar a cabo esta fase de integración, se ha propuesto la utilización de técnicas de lógica borrosa. Esta metodología proporciona una definición sencilla de los criterios de selección que caracterizan a un selector de objetivo y un manejo adecuado de las propiedades de regiones utilizadas en el proceso de selección. Así, cada selector de objetivo se construye a partir de un sistema de reglas borrosas que permiten determinar la saliencia de las regiones del entorno. A partir de la salida final del sistema de reglas, el selector determina cuáles son las regiones del entorno de mayor interés y cuáles deben ser completamente descartadas en el proceso de selección. Se ha propuesto, además, un método de asignación de consecuentes a las reglas del sistema

que garantiza resultados coherentes con todas las especificaciones de funcionamiento del selector, a la vez que proporciona una vía de reducción del número de reglas del sistema.

4. Cada selector de objetivo tiene la capacidad de alternar entre varias regiones candidatas a la fijación atencional. Para ello, se ha presentado un mecanismo de inhibición de retorno que garantiza la selección en orden decreciente de saliencia de las diferentes zonas candidatas y la reanudación del ciclo de selección tras fijar la atención sobre el total de regiones.

5. Los selectores de objetivo mantienen simultáneamente atención encubierta sobre sus correspondientes objetivos visuales. La atención abierta, en un instante dado, es dirigida por un único selector. La frecuencia con la que cada selector adquiere el control de la atención abierta es modulada por los comportamientos de alto nivel en función de sus necesidades. No obstante, pueden darse peticiones simultáneas de control abierto por parte de varios selectores de objetivo que hay que resolver desde el sistema atencional. Para dar respuesta a este tipo de situaciones, se ha propuesto un método de asignación del control que ajusta las frecuencias individuales de los selectores según las posibilidades de cada situación.

6. La fijación atencional se hace efectiva a través de movimientos de cámara que permiten centrar el objetivo en las dos imágenes del par estéreo. Para ello, se ha propuesto la descomposición de dicha fijación en dos movimientos independientes: un movimiento sacádico y de seguimiento en una de las cámaras y un movimiento asimétrico de vergencia en la otra. Esta separación permite que la programación del sacádico que proporciona la fijación atencional se lleve a cabo para una única cámara, no siendo así necesario mantener información binocular del objetivo. Para cada tipo de movimiento se ha presentado un método de control que proporciona el funcionamiento requerido en cada caso. Ambos métodos se basan en la estructura multi-escala de la imagen. El control de seguimiento del objetivo se lleva a cabo siguiendo un proceso de búsqueda ascendente en el espacio de escala. Esto permite descartar zonas de apariencia similar al objetivo. El control de vergencia sigue una estrategia de localización descendente en el espacio de escala, proporcionando un ajuste progresivo de la posición objetivo.

Con respecto a la implementación software, se ha empleado una metodología basada en componentes para construir el sistema propuesto. Se ha presentado la estructura lógica

de los componentes que intervienen en el sistema y se ha propuesto un método de sincronización entre componentes que asegura la consistencia de la información que fluye en el sistema en cada instante. La arquitectura software resultante presenta una estructura flexible que puede ser ampliada fácilmente añadiendo nuevos componentes sin modificar los ya existentes. La naturaleza distribuida de esta arquitectura permite obtener un rendimiento en tiempo real del sistema a través de la ejecución concurrente de sus componentes.

El sistema propuesto ha sido probado mediante una serie de experimentos reales conducidos por un robot móvil diseñado y construido en el laboratorio. Se han presentado pruebas de validación independiente de determinados módulos del sistema. Además, se han mostrado los resultados obtenidos en experimentos globales del sistema destinados a resolver distintas tareas de navegación.

## 8.2. Líneas futuras

Tal y como se expuso en la introducción de esta memoria, esta tesis pretende abrir nuevas vías de estudio sobre las relaciones entre la percepción visual y el control de la acción en robots. La propuesta que aquí se ha presentado muestra la posibilidad de relacionar ambos procesos utilizando la atención como intermediario. No obstante, para que este planteamiento sirva de punto de partida en futuras investigaciones, es necesario identificar varias líneas de continuación y mejora del trabajo desarrollado.

Con respecto a las mejoras que pueden ser incluidas en el sistema, cabe destacar las siguientes:

- El método de detección de regiones utilizado está limitado a zonas de la imagen que pueden interpretarse como esquinas en algún nivel del espacio de escala. Esto restringe el tipo de regiones que pueden ser detectadas a regiones circulares, lo que impide el tratamiento adecuado de aquellas zonas de la imagen que no pueden incluirse dentro de este grupo. Para mejorar este aspecto del sistema, se contempla la opción de sustituir el método Harris-Laplace original por su extensión como detector invariable a transformaciones afines. La aplicación de este método permitiría extraer regiones elípticas, aunque a costa de un aumento significativo del tiempo de procesamiento. Otra posibilidad es ampliar el tipo de primitivas, incluyendo, por ejemplo, líneas de imagen en el espacio de escala. La detección de este tipo de atributos de imagen en

la estructura multi-escala permitiría tratar cada línea como una región rectangular, consiguiendo así aumentar el tipo de regiones que pueden ser detectadas en el sistema, sin que esto suponga un incremento excesivo de la complejidad computacional del método.

- Cuando el sistema de control de seguimiento de un objetivo visual no consigue obtener la posición actual del objetivo, espera a recibir una nueva petición de seguimiento que le permita reanudar el control. Dentro de nuestro sistema, esta espera no supone un gran problema, ya que la selección de un foco de atención en cada ciclo del proceso se traduce en órdenes continuas de seguimiento de un objetivo. Sin embargo, como proceso aislado, el controlador de seguimiento podría mejorarse sustituyendo la espera por búsquedas más exhaustivas del objetivo. Esta búsqueda podría consistir, por ejemplo, en la aplicación del método propuesto a vistas previas del objetivo, lo que permitiría resolver situaciones en las que la pérdida del objetivo está provocada por una ocultación momentánea del mismo o por cambios bruscos de luminosidad.

- La descomposición de movimientos que dan lugar a la fijación binocular provoca un cierto desfase en la localización del objetivo entre ambas cámaras. Así, el control de vergencia se inicia siempre con posterioridad al de seguimiento, dado que su único indicio de cambio viene dado por la disparidad entre el par de imágenes. Este aspecto podría mejorarse incluyendo algún tipo de comunicación entre los dos subsistemas que les permitiera sincronizarse. Una posibilidad sería el envío, previo al movimiento, de la información referente al nuevo objetivo desde el controlador de seguimiento al controlador de vergencia. Esto proporcionaría una anticipación interna en este último que daría lugar a la fijación simultánea del objetivo en ambas cámaras.

- Una cuestión abierta en el sistema propuesto es cómo determinar la frecuencia a la que debe funcionar cada selector de objetivo para obtener el comportamiento global deseado. En los ejemplos que se han presentado, la frecuencia es determinada por cada comportamiento, en función de la situación actual, siguiendo una relación preestablecida. La eficiencia en el control global depende de que esta relación se ajuste lo máximo posible a las necesidades de cada instante, por lo que este aspecto del sistema podría ser mejorado incluyendo un proceso de aprendizaje que permitiera a cada comportamiento adaptar sus funciones de frecuencia de acuerdo con la experiencia ante nuevas situaciones.

Además de estas líneas de actuación que permitirían mejorar diferentes aspectos del

sistema, se pueden establecer otras líneas de continuación que supongan nuevas exploraciones del enlace entre atención y acción en el control de robots. En este sentido, creemos que sería de especial interés añadir nuevos elementos al cuerpo del robot que amplíen sus capacidades de actuación y permitan incluir diferentes dinámicas de atención-acción. Tal y como ya se expuso, una posibilidad sería dotar al robot de un manipulador. En relación a este nuevo elemento, la atención debería centrarse en zonas del entorno acordes con las nuevas posibilidades de actuación. Para este caso concreto, el sistema debería dar respuesta a situaciones de mayor complejidad que las planteadas en esta tesis, donde, por ejemplo, ciertas zonas del manipulador podrían formar parte de las regiones visibles del entorno. Una prueba de este tipo supondría un paso más en la validación de nuestra propuesta.

# Bibliografía


Agre, P. y Chapman, D.: 1987, Pengi: an implementation of a theory of activity, *Proceedings of 6th AAAI National Conference on Artificial Intelligence*, pp. 268–272.

Albus, J.: 1993, A reference model architecture for intelligent systems design, *in* P. Antsaklis y K. Passino (eds), *Introduction to Intelligent and Autonomous Control*, Kluwer Academic Press, pp. 27–56.

Albus, J., Lumia, R., Fiala, J. y Wavering, A.: 1989, Nasrem: The nasa/nbs standard reference model for telerobot control system architecture, *Proceedings of 20th International Symposium on Industrial Robots*.

Allport, A.: 1987, Selection for action: Some behavioral and neurophysiological considerations of attention and action, *in* H. Heuer y A. Sanders (eds), *Perspectives on perception and action*, Erlbaum.

Anderson, C. y Van Essen, D.: 1987, Shifter circuits: A computational strategy for dynamic aspects of visual processing, *Proceedings of the National Academy of Sciences, USA* **84**, 6297–6301.

Arbib, M.: 1981, *Perceptual structures and distributed motor control*, American Physiological Society, Bethesda, MD, pp. 1449–1480.

Arkin, R.: 1987, Motor schema based navigation for a mobile robot: An approach to programming by behavior, Vol. 4, pp. 264–271.

Arkin, R.: 1989, Towards the unification of navigational planning and reactive control, *AAAI Spring Symposium on Robot Navigation*.

Arkin, R.: 1993, Modeling neural function at the schema level: implications and results for robotic control, *Proceedings of the workshop on "Locomotion Control in Legged Invertebrates" on Biological neural networks in invertebrate neuroethology and robotics*, Academic Press Professional, Inc., pp. 383–410.

Arkin, R.: 1998, *Behaviour-based Robotics*, MIT Press.

Arkin, R. y MacKenzie, D.: 1994, Temporal coordination of perceptual algorithms for mobile robot navigation, *IEEE Transactions on Robotics and Automation* **10**, 276–286.

Bachiller, P., Bustos, P., Cañas, J. M. y Royo, R.: 2007, An experiment in distributed visual attention, *3rd Int. Workshop on Multi-Agent Robotic Systems*.

Bachiller, P., Monasterio, F., Bustos, P. y Vicente, J.: 2003, Fuzzy controller for dynamic vergence in a stereo head, *11th International Conference on Advanced Robotics*, pp. 1653–1658.

Bajcsy, R.: 1988, Active perception, *Proceedings of the IEEE* **76**, 996–1005.

Ballard, D.: 1991, Animate vision, *Artificial intelligence* **48**, 57–86.

Bandera, C. y Scott, P.: 1989, Foveal machine vision systems, *IEEE International Conference on Systems, Man and Cybernetics*, pp. 596–599.

Bridgeman, B., Vander Heijde, A. H. C. y Velichkovsky, D. M.: 1994, A theory of visual stability across saccadic eye movements, *Behavioral and Brain Sciences* **1782**, 247–292.

Broadbent, D. E.: 1958, *Perception and communication*, Pergamon Press, New York, NY.

Brooks, A., Kaupp, T., Makarenko, A., Williams, S. B. y Oreback, A.: 2005, Towards component-based robotics, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'05)*.

Brooks, A., Kaupp, T., Makarenko, A., Williams, S. y Oreback, A.: 2007, Orca: A component model and repository, *Software Engineering for Experimental Robotics*, Springer Berlin/Heidelberg, pp. 231–251.

Brooks, R.: 1986, A robust layered control system for a mobile robot, *IEEE Journal of Robotics and Automation* **2**, 14–23.

Brooks, R.: 1991a, Intelligence without reason, *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pp. 569–595.

Brooks, R.: 1991b, Intelligence without representation, *Artificial Intelligence* **47**(1-3).

Brown, R., Pham, B. y Aidman, E.: 2000, Efficient image rendering using a fuzzy logic model of visual attention, *Proceedings Conference: Advances in Intelligent Systems: Theory and Applications (AISTA)*, pp. 314–319.

Brown, R., Pham, B. y Maeder, A. J.: 2001, A fuzzy logic model of visual importance for efficient image synthesis, *FUZZ-IEEE*, pp. 1400–1403.

Cave, K.: 1999, The featuregate model of visual selection, *Psychological Research* **62**, 180–194.

Crowley, J.: 1981, A representation for visual information, *PhD thesis, Carnegie Mellon University* .

Crowley, J. y Parker, A.: 1984, A representation for shape based on peaks and ridges in the difference of low pass transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Deutsch, J. y Deutsch, D.: 1963, Attention: some theoretical considerations, *Psychological Review* **70**, 80–90.

Duncan, J.: 1996, Coordinated brain systems in selective perception and action, *Attention and Performance XVI. Information Integration in Perception and Communication*, Cambridge, MA: MIT Press, pp. 549–578.

Enright, J.: 1998, Monocularly programmed human saccades during vergence changes?, *Journal of Physiology* **512**, 235–250.

Eriksen, C. W.: 1990, Attentional search of the visual field, *Visual search*, London: Taylor and Francis, pp. 3–19.

Fikes, R. y Nilsson, N.: 1971, Strips: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* **2**, 189–208.

Frintrop, S.: 2006, *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search*, Vol. 3899 of *Lecture Notes in Computer Science*, Springer.

Frintrop, S., Backer, G. y Rome, E.: 2005, Goal-directed search with a top-down modulated computational attention system., *in* W. G. Kropatsch, R. Sablatnig y A. Hanbury (eds), *DAGM-Symposium*, Vol. 3663 of *Lecture Notes in Computer Science*, Springer, pp. 117–124.

Gat, E.: 1992, Integrating planning and reaction in a heterogenous asynchronous architecture for controlling real-world mobile robots, *Proc. 10th National Conf. on Artificial Intelligence, AAAI-92.*

Georgeff, M. y Lansky, A.: 1987, Reactive reasoning and planning, *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI'87)*, pp. 677–682.

Gerkey, B. P., Vaughan, R. T., Stoy, K., Howard, A., Sukhatme, G. S. y Mataric, M. J.: 2001, Most valuable player: a robot device server for distributed control, *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, Vol. 3, pp. 1226–1231.

Gerkey, B., Vaughan, R. y Howard, A.: 2003, The player/stage project: Tools for multi-robot and distributed sensor systems, *Proceedings of the International Conference on Advanced Robotics (ICAR 2003), Coimbra, Portugal, June 30 - July 3, 2003*, pp. 317–323.

Gibson, J. J.: 1979, *The ecological approach to visual perception*, Lawrence Erlbaum Associates.

Giralt, G., Chatila, R. y Vaisset, M.: 1984, An integrated navigation and motion control system for autonomous multisensory mobile robots, *First International Symposium on Robotics Research*, pp. 191–214.

Grimes, R.: 1997, *Professional Dcom programming*, Wrox Press Ltd., Birmingham, UK, UK.

Guerra, C.: 2002, *Contribuciones al seguimiento visual precategórico*, PhD thesis, Universidad de Las Palmas de Gran Canaria.

Hanss, M.: 1999, Identification of enhanced fuzzy models with special membership functions and fuzzy rule bases.
**URL:** *citeseer.ist.psu.edu/hanss99identification.html*

Harris, C. y Stephens, M.: 1988, A combined corner and edge detector, *Proceedings of The Fourth Alvey Vision Conference*, pp. 147–151.

Hartley, R. y Zisserman, A.: 2004, *Multiple View Geometry in Computer Vision*, Cambridge University Press.

He, J., Li, X. y Liu, Z.: 2005, Component-based software engineering: the need to link methods and their theories, *Proceedings of ICTAC 2005, Lecture Notes in Computer Science 3722*, pp. 70–95.

Heinke, D. y Humphreys, G.: 1997, Saim: A model of visual attention and neglect, *ICANN '97: Proceedings of the 7th International Conference on Artificial Neural Networks*, Springer-Verlag, London, UK, pp. 913–918.

Henning, M. y Vinoski, S.: 1999, *Advanced CORBA programming with C++*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Hoff, J. y Bekey, G.: 1995, An architecture for behavior coordination learning, *IEEE International Conference on Neural Networks*.

Humphreys, G. y Forde, M.: 1998, Disorder action schema and action dysorganization syndrome, *Cognitive Neuropshychology* **15**, 771–811.

Itti, L. y Koch, C.: 2000, A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research* **40**, 1489–1506.

James, W.: 1890, *The Principles of Psychology*, Dover Publications.

Johnson, A. y Hebert, M.: 1999, Using spin images for efficient object recognition in cluttered 3d scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 433–449.

Johnson, J. S., Spencer, J. P. y Schöner, G.: 2007, Moving to higher ground: The dynamic field theory and the dynamics of visual cognition, *New Ideas in Psychology* .

Koch, C. y Ullman, S.: 1985, Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology* **4**, 219–227.

Košecká, J. y Bajcsy, R.: 1993, Discrete event systems for autonomous mobile agents, *Proceedings Intelligent Robotic Systems '93 Zakopane*, pp. 21–31.

LaBerge, D.: 1995, *Attentional Processing*, Harvard University Press.

Laird, J., Newell, A. y Rosenbloom, P.: 1987, Soar: an architecture for general intelligence, *Artificial Intelligence* **33**, 1–64.

Lazebnik, S., Schmid, C. y Ponce, J.: 2005, A sparse texture representation using local affine regions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, 1265–1278.

Lindeberg, T.: 1998, Feature detection with automatic scale selection, *International Journal of Computer Vision* **30**(2), 77–116.

Lowe, D.: 1999, Object recognition from local scale-invariant features, *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, Vol. 2, pp. 1150–1157.

Lowe, D.: 2004, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**, 91–110.

Lyons, D. y Hendriks, A.: 1995, Planning as incremental adaptation of a reactive system., *Robotics and Autonomous Systems* **14**, 255–288.

Maes, P.: 1989, How to do the right thing, *Technical report*, Massachusetts Institute of Technology, Cambridge, MA, USA.

Mamdani, E. H. y Assilian, S.: 1975, An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal on Man-Machine Studies* **7**, 1–13.

Matarić, M.: 1992, Behavior-based control: main properties and implications, *Proceedings of Workshop on Intelligent Control Systems, International Conference on Robotics and Automation*, pp. 46–54.

Medvidovic, N. y Taylor, R.Ñ.: 2000, A classification and comparison framework for software architecture description languages, *IEEE Transactions on Software Engineering* **26**(1), 70–93.

Mikolajczyk, K.: 2002, *Interest point detection invariant to affine transformation*, PhD thesis, Institut National Polytechnique de Grenoble.

Mikolajczyk, K. y Schmid, C.: 2001, Indexing based on scale invariant interest points, *Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 1, pp. 525–531.

Mikolajczyk, K. y Schmid, C.: 2004, Scale & affine invariant interest point detectors, *International Journal of Computer Vision* **60**(1), 63–86.

Mikolajczyk, K. y Schmid, C.: 2005, A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, 1615–1630.

Milner, A. D. y Goodale, M. A.: 1995, *The visual brain in action*, Oxford University Press.

Montemerlo, M., Roy, N. y Thrun, S.: 2003, Perspectives on standardization in mobile robot programming: The carnegie mellon navigation (carmen) toolkit, *Proceedings of*

*the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, Vol. 3, Las Vegas, NV, pp. 2436–2441.

Moravec, H.: 1981, Rover visual obstacle avoidance, *Proceedings of the seventh International Joint Conference on Artificial Intelligence*, pp. 785–790.

Navalpakkam, V. y Itti, L.: 2005, Modeling the influence of task on attention., *Vision Research* **45**, 205–231.

Navalpakkam, V. y Itti, L.: 2006, An integrated model of top-down and bottom-up attention for optimizing detection speed, *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 2049–2056.

Neumann, O., van der Heijden, A. H. C. y Allport, A.: 1986, Visual selective attention: Introductory remarks, *Psychological Research* **48**, 185–188.

Nilsson, N.: 1969, A mobile automaton: An application of artificial intelligence techniques, *IJCAI*, pp. 509–520.

Nilsson, N.: 1984, Shakey the robot, *Technical note no. 323*, Artificial Intelligence Center, SRI International, Menlo Park, CA.

Olshausen, B., Anderson, C. y Van Essen, D.: 1993, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *The Journal of Neuroscience* **13**(11), 4700–4719.

O'Regan, J. K., Rensink, R. A. y Clark, J. J.: 1999, Change-blindness as a result of 'mudsplashes', *Nature* **398**(6722).

Perry, R. y Hodges, J.: 1999, Attention and execution deficits in aliheimer's disease: a critical review, *Brain* **122**, 383–404.

Pirjanian, P.: 1998, *Multiple objective action selection and behavior fusion using voting*, PhD thesis, Aalvorg University.

Posner, M. I., Snyder, C. R. R. y Davidson, B. J.: 1980, Attention and the detection of signals, *Journal of Experimental Psychology: General* **109**, 160–174.

Posner, M. y Dehaene, S.: 1994, Attentional networks, *Trends in Neuroscience* **17**, 75–79.

Postma, E., van den Herik, H. y Hudson, P.: 1997, SCAN: A scalable model of attentional selection, *Neural Networks* **10**(6), 993–1015.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., Lamantia, A., Mcnamara, J. y Williams, S. (eds): 2004, *Neuroscience 3rd edition*, Sinauer Associates.

Riddoch, M., Humphreys, G. y Edwards, M.: 2000a, Neuropshychological evidence distinguishing object selection from action (effector) selection, *Cognitive Neuropshychology* **17**, 547–562.

Riddoch, M., Humphreys, G. y Edwards, M.: 2000b, Visual affordances and object selection, *Attention and Performance XVIII*, Cambridge, MA: MIT Press.

Riekki, J. y Röning, J.: 1997, Reactive task execution by combining action maps., *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'97)*, pp. 224–230.

Rizzolatti, G., Riggio, L., Dascola, I. y Umilta, C.: 1987, Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention, *Neuropsychologia* **25**, 31–40.

Rohrer, T.: 2006, The body in space: embodiment, experientialism and linguistic conceptualization, *Body, Language and Mind*, Vol. 2.

Roselló, J., Munar, E. y Garrido, M.: 2001, La naturaleza de la atención visual: ¿monarquía, oligarquía o anarquía?, *Revista de Psicología General y Aplicada* **54**, 31–46.

Rosenblatt, J.: 1995, Damn: A distributed architecture for mobile navigation, *Proc. of the AAAI Spring Symp. on Lessons Learned from Implememted Software Architectures for Physical Agents.*

Roshandel, R., Schmerl, B., Medvidovic, N., Garlan, D. y Zhang, D.: 2004, Understanding tradeoffs among different architectural modeling approaches, *Software Architecture, 2004. WICSA 2004. Proceedings. Fourth Working IEEE/IFIP Conference on*, pp. 47–56.

Saffiotti, A.: 1997, The uses of fuzzy logic for autonomous robot navigation: a catalogue raisonn'e, *Soft Computing Research journal* **1**, 180–197.

Saffiotti, A., Konolige, K. y Ruspini, E.: 1995, A multivalued-logic approach to integrating planning and control, *Artificial Intelligence* **76**, 481–526.

Simons, D. y Levin, D.: 1998, Failure to detect changes to attended objects, *Psychonomic Bulletin and Review* **5**, 644–649.

Sun, Y. y Fisher, R.: 2003, Object-based visual attention for computer vision, *Artificial Intelligence* **146**(1), 77–123.

Szyperski, C.: 1998, *Component software: beyond object-oriented programming*, ACM Press/Addison-Wesley Publishing Co.

Takagi, T. y Sugeno, M.: 1985, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man and Cybernetics* **SMC-15**, 116–132.

Torr, P.: 1995, *Outlier detection and motion segmentation*, PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.

Torralba, A.: 2003a, Modeling global scene factors in attention, *Journal of Optical Society of America* **20(7)**, 1407–1418.

Torralba, A.: 2003b, Contextual priming for object detection, *Internation Journal of Compututer Vision* **53**, 169–191.

Torralba, A., Oliva, A., Castelhano, M. S. y Henderson, J. M.: 2006, Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search., *Psychol Rev* **113**, 766–786.

Treisman, A.: 1988, Features and objects: the fourteenth bartlett memorial lecture, *Quarterly Journal of Experimental Psychology A* **40**, 201–237.

Treisman, A. M. y Gelade, G.: 1980, A feature-integration theory of attention, *Cognitive Psychology* **12**(1), 97–136.

Trucco, E. y Verri, A.: 1998, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall.

Tsotsos, J.: n.d., The selective tuning model for visual attention.
**URL:** *citeseer.ist.psu.edu/702709.html*

Tsotsos, J., Culhane, S. y Cutzu, F.: 2001, From theoretical foundations to a hierarchical circuit for selective attention, *Visual Attention and Cortical Circuits*, MIT Press, Cambridge MA, pp. 285–306.

Tsotsos, J. K., Culhane, S. M., Winky, W. Y. K., Lai, Y., Davis, N. y Nuflo, F.: 1995, Modeling visual attention via selective tuning, *Artificial Intelligence* **78**(1-2), 507–545.

Ungerleider, L. G. y Mishkin, M.: 1982, Two cortical visual systems, *in* D. J. Ingle, M. A. Goodale y R. J. W. Mansfield (eds), *Analysis of Visual Behavior*, Cambridge, MA: MIT Press, pp. 549–586.

Wolfe, J. M.: 1994, Guided search 2.0: A revised model of visual search, *Psychonomic Bulletin & Review* **1**(2), 202–238.

Yantis, S.: 1998, Control of visual attention, *in* H. Pashler (ed.), *Attention*, Hove: Psychology Press, pp. 223–256.

Yarbus, A.: 1967, *Eyes movements and vision*, Plenum Press.

Yen, J. y Pfluger, N.: 1995, A fuzzy logic based extension to payton and rosenblatt's command fusion method for mobile robot navigation, *IEEE Trans. on Systems, Man, and Cybernetics* **25**, 971–978.

Zadeh, L.: 1965, Fuzzy sets, *Information and Control* (8), 338–353.

Zeki, S.: 1993, *A vision of the brain*, Oxfor: Blackwell.

ZeroC: 2007, Internet communication engine.
**URL:** *http://www.zeroc.com/ice.html*

Zhang, J. Z. y Wu, Q. M. J.: 2001, A pyramid approach to motion tracking, *Artificial Intelligence* **7**, 529–544.

Zhang, Z., Deriche, R., Faugeras, O. y Luong, Q.: 1995, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence* **78**(1-2), 87–119.