

# Manteniendo el Foco: Control de Seguimiento y Vergencia en un Sistema de Atención Visual

Pilar Bachiller y Pablo Bustos

**Resumen**—Durante los últimos años, la atención visual se ha convertido en una cuestión de gran interés en visión artificial. La mayoría de los modelos existentes plantean el mecanismo atencional como un medio de exploración visual y detección de objetos que se resuelve de forma estática, sin movimientos de cámara. En situaciones reales, la atención debe ir acompañada de mecanismos de fijación que permitan mantener el foco de atención aunque se produzcan variaciones provocadas por la dinámica del entorno o por las propias acciones del robot. Para tratar esta cuestión, en este artículo se presenta un sistema de seguimiento 3D de objetivos visuales que se encuentra integrado en un sistema de atención visual. Como principal aportación, se propone la descomposición de la fijación 3D en dos movimientos que forman dos subsistemas: controlador de seguimiento monocular y controlador de vergencia. Para validar esta propuesta, se muestran los resultados obtenidos en una serie de experimentos reales realizados en un robot móvil dotado de una torreta de visión estéreo.

**Index Terms**—Seguimiento 3D, atención visual, visión activa.

## I. INTRODUCCIÓN

LOS movimientos oculares son esenciales en muchos sistemas visuales de seres vivos porque permiten compensar la pérdida de resolución debida al pequeño tamaño de la fovea. Mediante dichos movimientos es posible dirigir la fovea hacia nuevos objetos de interés o compensar los movimientos de un objeto que está siendo focalizado [4]. Como en sus homólogos biológicos, los sistemas visuales de los robots con campo visual limitado deben incorporar capacidades motoras que permitan adaptar sus sensores de forma efectiva a su relación con un entorno dinámico.

En este artículo se presenta un sistema de control de seguimiento 3D de objetivos visuales. El sistema está integrado en una red de componentes que constituyen un modelo de atención visual para un robot móvil con visión estereoscópica [3]. Dentro de este modelo, la atención es dirigida hacia regiones de interés del entorno siguiendo una estrategia de control guiada por objetivo. Las regiones son extraídas en fases iniciales del proceso a partir de la detección de puntos de interés en el espacio de escala mediante el método Harris-Laplace [9]. La selección de una región como foco de atención se transforma en una respuesta motora que proporciona la fijación y el seguimiento binocular del nuevo foco hasta la selección del siguiente. Los componentes de control dedicados a esta última etapa de proceso forman el sistema de seguimiento 3D propuesto.

Existen numerosas propuestas de control de seguimiento visual [10]. Dentro de los trabajos de seguimiento con múltiples

vistas, el planteamiento mayoritario consiste en realizar un control de seguimiento homogéneo en las distintas cámaras basado en la búsqueda de correspondencias entre las diferentes vistas [5] [8]. La propuesta que aquí se presenta constituye un esquema de control cooperativo en el que cada cámara juega un papel diferente en el control global. Así, el sistema está formado por dos subsistemas de control: un subsistema de control monocular de seguimiento y un subsistema de control asimétrico de vergencia. Esta separación permite que el sacádico que proporciona la fijación inicial sobre el objetivo pueda programarse para una única cámara a la vez que se mantiene una focalización estable en ambas cámaras [6]. Además, este esquema proporciona una respuesta eficaz ante situaciones en las que no es posible obtener una correspondencia completa del objetivo en el par de imágenes por el cambio de perspectiva, la ocultación parcial en una de las vistas o incluso por la no visibilidad del objetivo desde una de las cámaras.

## II. CONTROL DE SEGUIMIENTO VISUAL

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 [7].

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 una representación multi-resolución de la imagen [1], este proceso es similar a una comparación en niveles de menor resolución manteniendo el tamaño de la ventana de correlación. Con esta idea, el método de seguimiento visual propuesto 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 cercana a 1, el grado de similitud obtenido no será suficiente para aceptar o rechazar dicha zona como nuevo punto de fijación. En este caso, será 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 repetiría para cada nivel hasta que sólo exista una zona de máxima similitud o se alcance el último nivel de la estructura.

Algoritmo 1 Seguimiento del foco de atención	
$prismaR$	= 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
$(xTrack, yTrack)$	= Posición de la región objetivo en el <i>frame</i> actual
$tracked$	= Indicador de éxito en el proceso de localización

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

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 I. 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 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

### Algorithm 1 Control de seguimiento visual

```

iCorrel  $\leftarrow$  cCorrel(prismaR[0], wE/2, hE/2, wE, hE,
  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
      xC  $\leftarrow$  (listMax[p].x) * sl
      yC  $\leftarrow$  (listMax[p].y) * sl
      correl  $\leftarrow$  scCorrel(prismaR[l], wE/2, hE/2,
        wE, hE, piramide[l], xC, yC, wE, hE)
      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

```

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 mantenimiento de otra región objetivo.

### III. CONTROL DE VERGENCIA

La función de los movimientos de vergencia en un sistema estéreo es la foveatización de un mismo objetivo visual en am-

bas cámaras. A partir de dichos movimientos, es posible fijar regiones de interés del espacio, proporcionando un mecanismo de seguimiento de objetivos visuales 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 el comportamiento del sistema. 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, 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 representación multi-escala de las imágenes del par estéreo. 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 [2].

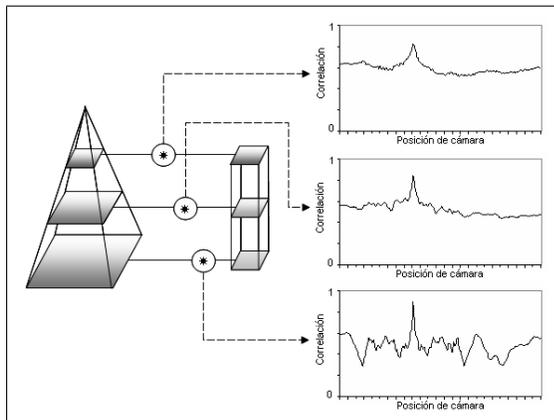


Fig. 1: Función de correlación en distintos niveles de la estructura multi-escala

En función de su posición en la estructura, cada nivel tiene asociado una función de correlación que se caracteriza por lo siguiente (figura 1):

- 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 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 fovea 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 [11]. 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.

Algoritmo 2 Control de vergencia	
$prismaF$	= Prisma central de la cámara fija de tamaño $r \times r$
$piramideV$	= 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
$(xVerg, yVerg)$	= Posición de vergencia de la cámara móvil
$maxCorrel$	= Máximo valor de correlación en toda la estructura

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

---

#### Algorithm 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
 $xVerg \leftarrow xS * s$ 
 $yVerg \leftarrow yS * s$ 

```

---

El algoritmo 2 muestra los pasos de este método de control de vergencia. En la tabla II 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 fovea de la cámara fija. Esta búsqueda consiste en la localización de la posición que 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 una posición central ( $xS, yS$ ), un alto ( $hS$ ) y un ancho ( $wS$ ). El alto se mantiene constante ( $hWin$ ) para conservar la proporción entre las dimensiones de la fovea 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.

## IV. EXPERIMENTOS

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.

### A. 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 2 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.

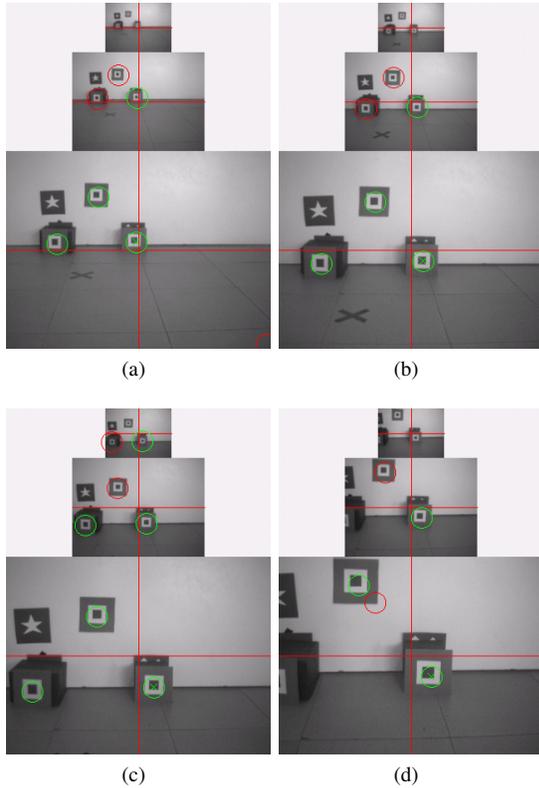


Fig. 2: Seguimiento de una región en una secuencia de avance

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 2. 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 2. 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.

### B. Control de vergencia

El segundo bloque de experimentos 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.

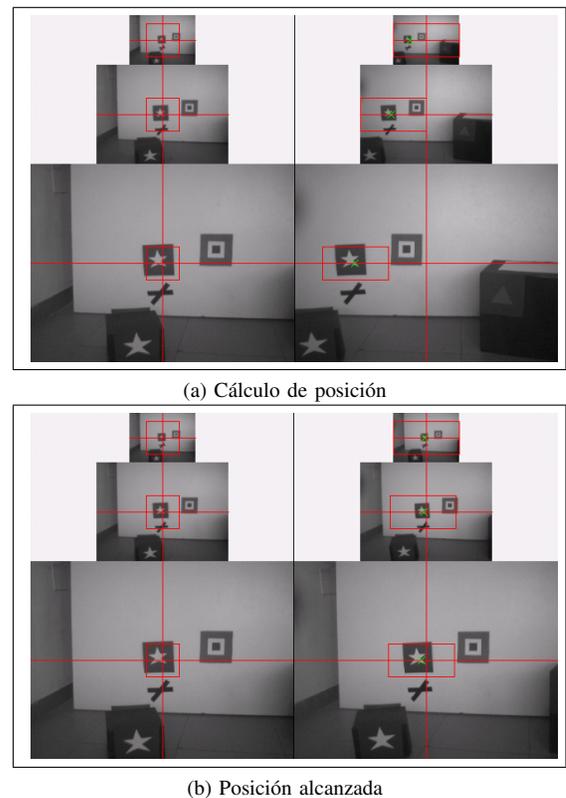
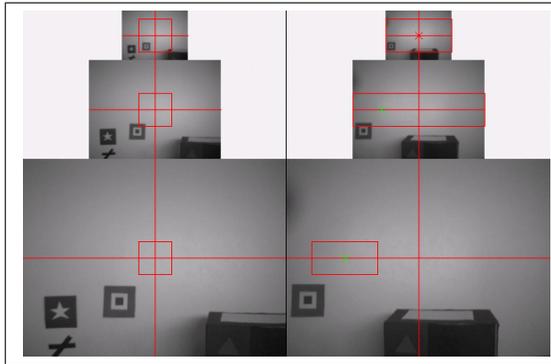


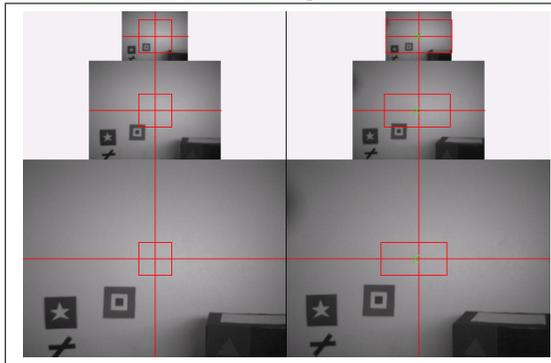
Fig. 3: Vergencia en una situación favorable.

En la primera situación (figura 3), 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 3 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 3).



(a) Cálculo de posición

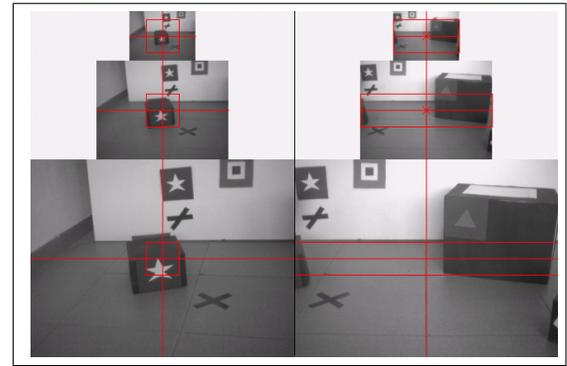


(b) Posición alcanzada

Fig. 4: 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 4 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.

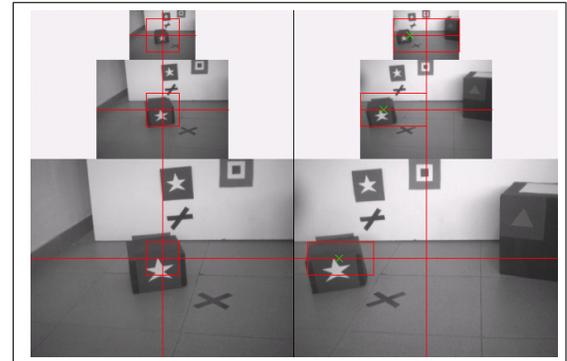
Para finalizar, se muestra una última situación en la que la zona de vergencia especificada por la cámara fija se encuentra fuera del campo visual de la cámara sobre la que se realiza el control. Esta situación es la indicada en la figura 5. En la imagen (a) de esta figura se muestra la aplicación del 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



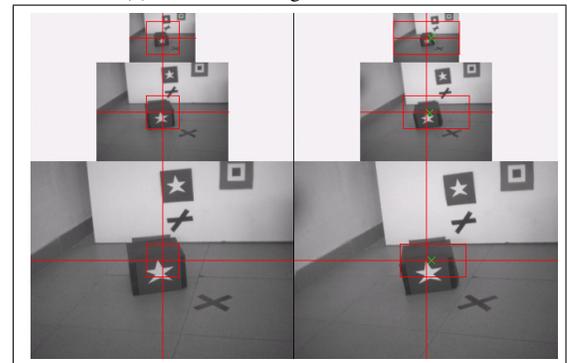
(a) Sin posición de vergencia. Se inicia el proceso de búsqueda dinámica

Fig. 5: Zona de vergencia fuera del campo de visión

la imagen (b) de la figura 5. 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

Fig. 5: Zona de vergencia fuera del campo de visión (continuación)

### C. Control conjunto de seguimiento y vergencia

Para concluir la serie de experimentos, 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 6 muestra varios

instantes de la secuencia de control obtenida en esta prueba. La imagen superior de cada instante se corresponde con las dos vistas de la escena obtenidas por las cámaras del par estéreo. La inferior muestra una vista general de la escena capturada por una cámara situada en el techo. 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 instantáneamente cambios 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.

## V. CONCLUSIONES

Se ha descrito un sistema de seguimiento 3D de objetivos visuales precatagóricos compuesto por dos subsistemas de control de los movimientos de seguimiento de la cámara dominante y de vergencia de la cámara secundaria. Cada controlador hace un uso específico del prisma multiescala visual. Ambos elementos funcionan como componentes separados de una arquitectura distribuida y se sincronizan a través de la información visual. Dos aspectos que mejorarían el comportamiento del sistema en términos de robustez y fiabilidad son: la interacción con un sistema conceptual de categorización que mantenga una referencia a qué se está siguiendo; y un mecanismo de señalización interna o copia eferente que permita al controlador de vergencia anticipar la posición final del sacádico para comenzar su propio movimiento antes de que éste finalice.

## AGRADECIMIENTOS

Este trabajo ha sido realizado gracias a la financiación proporcionada por la Consejería de Infraestructuras y Desarrollo Tecnológico de la Junta de Extremadura a través del proyecto: "Mecanismos de Atención Visual en Robots Móviles" (ref. 3PR05A044)

## REFERENCIAS

[1] J. Arbeiter and R. Bessler, *A two-dimensional real-time video pyramid*, RCA Review, 1: 3-31, 1986.

[2] P. Bachiller, F. Monasterio, P. Bustos y J. Vicente, *Fuzzy controller for dynamic vergence in a stereo head*, 11th International Conference on Advanced Robotics, 2003.

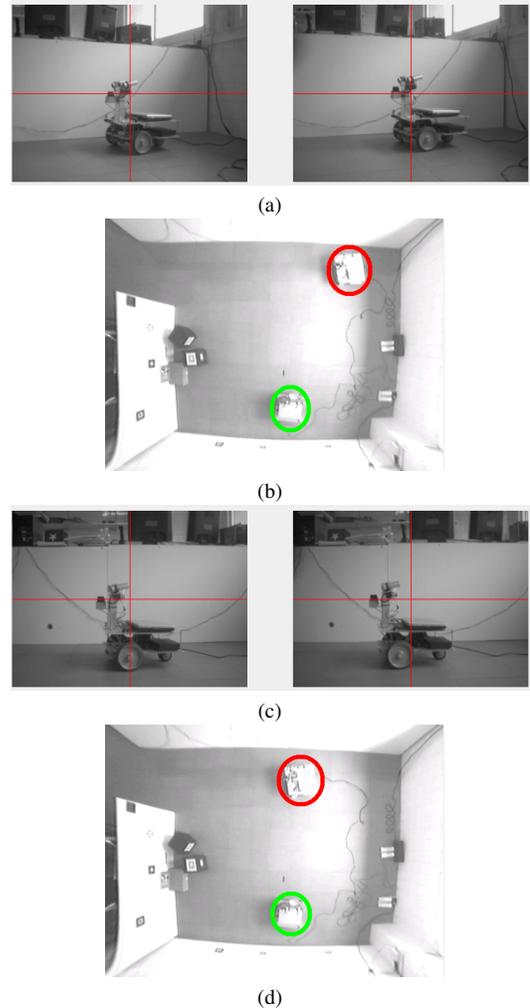


Fig. 6: Seguimiento y vergencia de un objetivo en movimiento (primera parte)

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

[4] R. Carpenter, *Eyes movements*, The Macmillan Press., 1991.

[5] Z. Chen and S. T. Birchfield, *Person following with a mobile robot using binocular feature-based tracking*, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007.

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

[7] C. Guerra, *Contribuciones al seguimiento visual precatagórico*, Tesis Doctoral, Universidad de Las Palmas de Gran Canaria, 2002.

[8] S. Khan and M. Shah, *Consistent labeling of tracked objects in multiple cameras with overlapping fields of view*, IEEE Trans. PAMI, 25(10):1355-1360, 2003.

[9] K. Mikolajczyk and C. Schmid, *Indexing based on scale invariant interest points*, Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, 1:525-531, 2001.

[10] A. Yilmaz, O. Javed, and M. Shah, *Object tracking: A survey*, ACM Comput. Surv. 38(4), 2006.

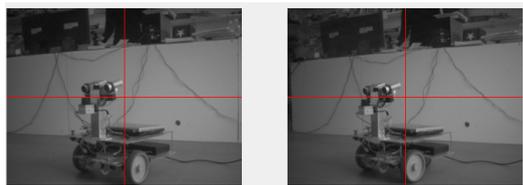
[11] J. Z. Zhang and Q. M. J. Wu, *A pyramid approach to motion tracking*, Real-Time Imaging, 7:529-544, 2001.



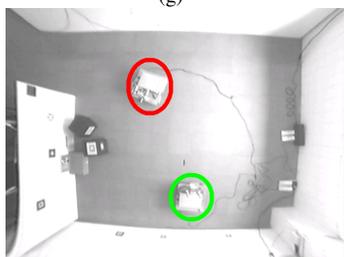
(e)



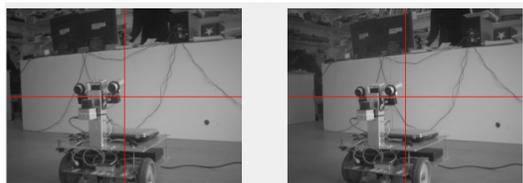
(f)



(g)



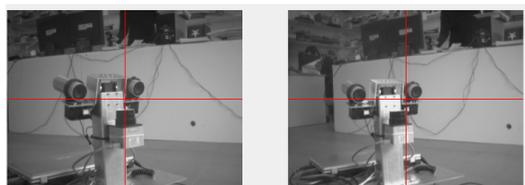
(h)



(i)



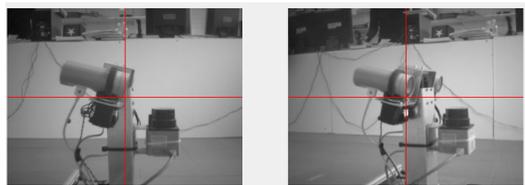
(j)



(k)



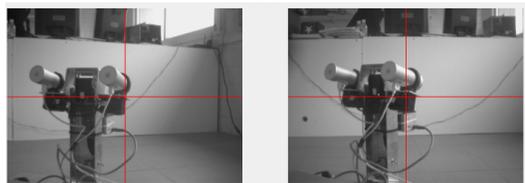
(l)



(m)



(n)



(o)



(p)

Fig. 6: Seguimiento y vergencia de un objetivo en movimiento (segunda parte)

Fig. 6: Seguimiento y vergencia de un objetivo en movimiento (tercera parte)