Reconstrucción 3D visual atentiva para la navegación de un robot móvil

José María Cañas, Olmo León y Roberto Calvo

***Abstract*—El artículo describe un nuevo algoritmo de reconstrucción 3D visual que permite navegar autónomamente a un robot dotado de un par estéreo móvil. Para solventar el alcance limitado de las cámaras el algoritmo utiliza un sistema atentivo que determina continuamente hacia dónde dirigir su mirada. Además, va integrando la información visual de los objetos relevantes en una memoria 3D de corto plazo que le sirve para navegar de manera segura y representar el entorno local más allá del alcance de las cámaras. Se han realizado numerosos experimentos y se ha validado el sistema en un robot real tipo Pioneer.**

***Index Terms*—computer vision, visual attention, autonomous navigation, 3D map building**

I. Introducción

La visión es el sentido principal en muchos animales, incluidos los seres humanos. No es por casualidad, la visión es capaz de proporcionar a un animal mucha información sobre su entorno. Y también a un robot. De hecho, en los últimos años las cámaras se han incorporado al equipamiento sensorial de los robots móviles.

Hoy día los robots más avanzados utilizan cámaras como sensor principal. Por ejemplo los humanoides Asimo de Honda, Wakamaru de Mitsubishi, Qrio de Sony la incorporan. Entre otras cosas para la detección de caras y la interacción hombre-máquina [BS99]. En el entorno de la RoboCup, el perrito Aibo de Sony y el robot Nao de Aldebaran tienen una cámara como sensor principal para determinar su posición en el campo, percibir la pelota, las porterías, etc..

Son numerosos los trabajos que presentan robots guiados por visión, que usan cámaras para navegar, tanto en interiores como en exteriores [H97], [H97b], [V03], [BCE2007]. También son frecuentes los trabajos que emplean visión para que el robot se autolocalice. Por ejemplo, el robot Minerva de Sebastian Thrun utilizaba una cámara mirando al techo para estimar su posición en un museo.

En parte debido a las economías de escala (hay cámaras en los teléfonos móviles, en los ordenadores, etc.) y en parte debido a la implantación de la barata tecnología CMOS, las cámaras han bajado de precio y ello ha favorecido su utilización en robótica. Una de sus características como sensores es que producen un caudal desbordante de datos y en general es difícil extraer desde ese caudal ingente de píxeles información útil para la tarea del robot.

Los algoritmos de atención visual se muestran útiles para ayudar a digerir la gran la cantidad de datos en el flujo de imágenes y para utilizar de manera eficiente las cámaras en la percepción del entorno. La atención visual es un fenómeno complejo y tema activo de investigación. Para avanzar un paso en este campo, en este artículo presentamos un algoritmo atentivo para que un robot perciba su entorno local eficientemente a partir de dos cámaras y navegue de manera segura por él sin utilizar sensores de distancia. El algoritmo mantiene una memoria visual de corto plazo y gobierna hacia donde miran las cámaras en todo momento.

En la sección siguiente se repasa de manera breve el estado del arte en técnicas de atención visual. En la sección III se presenta el algoritmo propuesto, desgranándolo en sus diferentes componentes de análisis bidimensional de las imágenes, paso a tres dimensiones, dinámicas de atención y navegación. En la sección IV se describen varios de los experimentos realizados. Finalmente se recapitulan las conclusiones obtenidas a lo largo de este trabajo.

II. Atención visual

Uno de los elementos más importantes dentro de la generación de comportamiento en robots móviles es el control de la atención. Los mecanismos de atención visual permiten a los robots responder con la rapidez necesaria a estímulos visuales y a objetivos internos aún disponiendo de una capacidad limitada de proceso. El interés en estos mecanismos ha sido muy alto en los últimos años y siempre ha ido de la mano de las investigaciones psicológicas experimentales en visión humana y animal. Por esto, los modelos más extendidos tienen una fuerte inspiración biológica [T95], [ZRT04].

Un robot capaz de explorar su entorno y de interaccionar con él debe disponer de mecanismos de atención que seleccionen regiones visuales de interés, sitúen al robot en posiciones favorables para la captación sensorial y, al mismo tiempo, mantengan la capacidad de reacción ante imprevistos y cambios bruscos del entorno[OMS05]. En general, el problema de la atención, tanto en seres vivos como en robots, se plantea en torno a dos ejes principales: la realización o no de movimientos del sensor; y la criterios de selección de la zona de interés.

En el primer caso tenemos dos tipos básicos: atención interna (*covert* en inglés) en la que se selecciona una región

José María y Roberto trabajan en la Universidad Rey Juan Carlos.
E-mail: jmplaza,rocapal@gsyc.es

Olmo León trabaja en Telefónica I+D.
E-mail: superolmin@gmail.com

Este trabajo ha sido parcialmente financiado por el Ministerio de educación (DPI2007-66556-C03) y por la Comunidad de Madrid (S-0505/DPI/0176 y CCG07-URJC/DPI-1694).

dentro del campo visual captado por el sensor para su procesamiento; y externa (*overt*) donde la región seleccionada implica un movimiento de reorientación o sacádico que sitúa a esa región en el centro del sensor [C05]. Este centro coincide en muchos ojos biológicos y algunas retinas artificiales con una zona reducida de muy alta resolución llamada fóvea. En ambos casos se selecciona una región de interés a la que se pueden dedicar momentáneamente todos los recursos de proceso. La secuencia temporal resultado de estas selecciones constituye un inteligente compromiso entre análisis detallado y extensión del entorno explorado, entre recursos y tiempo.

El segundo eje recoge las criterios de selección de las regiones de interés o, en otras palabras, "hacia donde miraremos después". Aquí, los extremos del eje se definen como atención ascendente y descendente. En el lado ascendente los criterios de selección del próximo punto de atención provienen directamente de las propiedades de la imagen de la escena captada por el sensor. La solución más frecuente es obtener un mapa de saliencia en el que a cada pixel se le asigna un valor relacionado con cuanto destaca visualmente la zona de la imagen de la que él es el centro [I02]. Los criterios para obtener este valor son complejos y recogen no linealmente la contribución de diferentes atributos calculados en la imagen, como la luminancia, la orientación del gradiente, diferentes canales de color, componentes del flujo óptico y disparidad estereoscópica, normalmente. El otro extremo de este eje recoge criterios generados a partir de los objetivos actuales del robot [NI05], [B08]. Así, si la misión que está realizando incluye localizar un elemento rojo en la escena, esta información llega hasta el sistema de atención para combinarse con la selección ascendente. Aunque existen varias propuestas, el caso más sencillo de combinación sería integrar en el mapa de saliencia la petición descendente y así obtener la zona de mayor interés según ambos criterios. Una vez que se ha seleccionado el máximo del mapa de saliencia, se ejecuta el movimiento de orientación de la cámara o sacádico para hacer coincidir la zona de interés con el centro de la imagen. Finalmente, incorporando una mecanismo que inhiba la zona visitada se permite que la siguiente región de mayor interés, según el mapa, sea la nueva candidata para la ejecución de otro sacádico.

Este sencillo sistema de cuatro pasos: saliencia, selección, sacádico e inhibición es un sistema dinámico que mantiene a la cámara realizando movimientos de orientación y fijando, cada vez, una zona del entorno en el centro de su campo visual. Este muestreo modulable del entorno es el mecanismo básico para multiplicar la capacidad del robot de reaccionar ante cambios en la escena y, al mismo tiempo, buscar, localizar y seguir ciertos elementos que sean de interés para la tarea en curso.

III. Reconstrucción 3D visual atentiva

El algoritmo propuesto busca en las imágenes características interesantes para la navegación y las sitúa en 3D utilizando el par estéreo. Esas características interesantes dependen de la aplicación concreta y pueden variar a lo largo del tiempo dependiendo de la actividad del robot. En nuestro caso son el color de determinadas balizas y el final del suelo, que también se estima por color.

El robot tiene una memoria de corto plazo que almacena la representación interna del entorno local. Esta memoria es un conjunto de objetos (balizas y bordes del suelo) que tienen asociados una saliencia, una *vida* y datos espaciales como la posición o la longitud.

La memoria de corto plazo se va actualizando a medida que se obtienen nuevas imágenes. Por ejemplo afinando la posición de objetos ya existentes observados de nuevo, consolidando la creencia en ellos (su vida) e incorporando nuevos objetos cuando se descubren. La memoria visual se utiliza para dos propósitos independientes: determinar el siguiente punto hacia el que se mira con las cámaras y calcular la orden de movimiento adecuada para el siguiente instante.

El algoritmo se articula entorno a cuatro bloques funcionales, que se detallan a continuación: (1) un análisis bidimensional de las imágenes, (2) extracción de información tridimensional, (3) dinámicas de atención visual y (4) navegación.

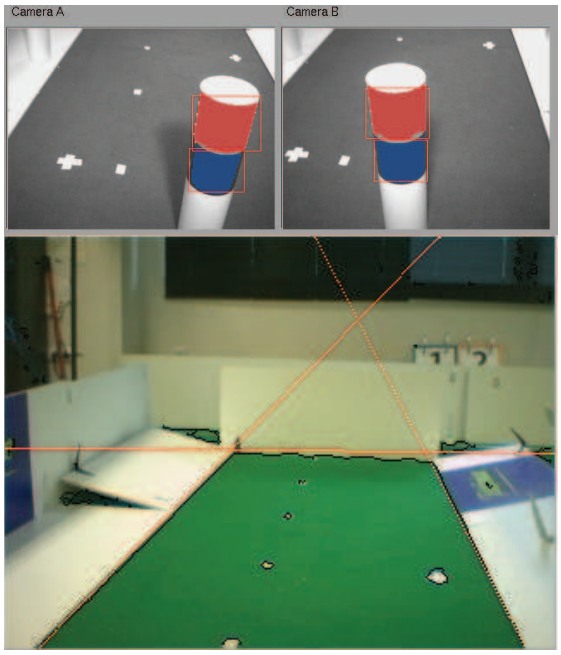

Fig. 1. Filtro de color para balizas (arriba) y para color de suelo, que además segmenta los bordes del suelo (abajo).

A. Análisis bidimensional

En primer lugar las imágenes de ambas cámaras se analizan filtrando por colores significativos, que son los posibles colores de todas las balizas y el color del suelo. Todos ellos se han aprendido previamente y caracterizado en el espacio HSV. Una vez extraídos los píxeles con colores significativos se agrupan en segmentos o ventanas rectangulares. Para ello se usa un algoritmo basado en doble umbral sobre el histograma de color [C05], que es rápido y razonablemente preciso.

La figura 1 (arriba) muestra las balizas coloreadas, cómo el filtro de color las ha identificado correctamente en ambas imágenes y las ventanas rectangulares correspondientes.

El procesamiento bidimensional para detectar los bordes del suelo es algo más complejo. Una vez que se han filtrado las imágenes por el color del suelo se calculan los bordes: allá donde acaba el suelo empiezan los obstáculos. Se ha implementado un filtro de bordes sencillo pero muy rápido, que recorre cada columna buscando discontinuidades de color con cierta tolerancia. El resultado es un conjunto de píxeles que marcan la frontera del suelo.

Seguidamente se utiliza la transformada de Hough para detectar segmentos rectos de puntos frontera. En la figura 1 se muestran los segmentos rectos detectados, que después se recortan adecuadamente.

B. Visión 3D

Una vez detectadas en las imágenes las balizas y los segmentos rectos que marcan la frontera del suelo, se utiliza geometría proyectiva y triangulación para situar en 3D ambas características.

En primer lugar, las balizas detectadas en la imagen izquierda se emparejan con sus homólogas en la imagen derecha. Cada baliza se identifica por una combinación de dos colores, y caso de existir varias similares en la escena, se emplean la restricción epipolar para emparejarlas correctamente. La figura 2 (arriba) muestra una triangulación situando en 3D el punto central de la baliza. Esa misma figura, abajo, muestra cómo la restricción epipolar ayuda a discriminar correctamente pares homólogos.

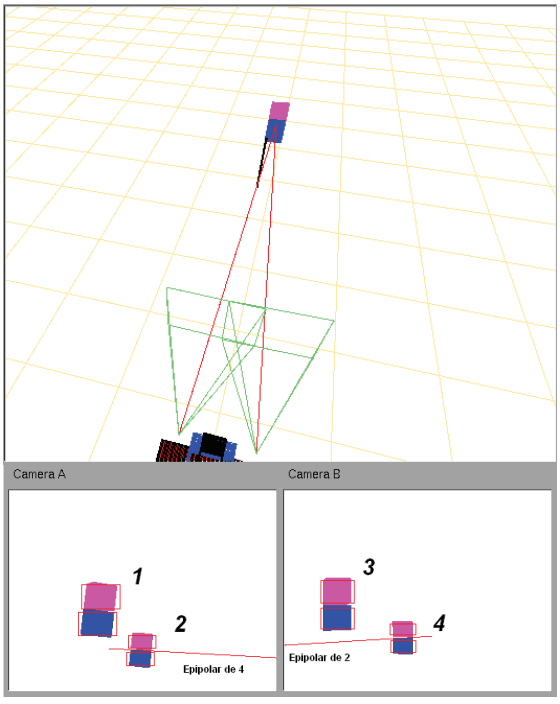

Fig. 2. Emparejamiento (abajo) y triangulación (arriba) de balizas.

En segundo lugar, la triangulación de los segmentos frontera del suelo se realiza asumiendo que pertenecen al plano del piso. Desde la imagen izquierda se proyecta en el plano de altura nula y da lugar a un segmento horizontal. Desde la imagen derecha se hace lo mismo, y si ambos segmentos horizontales solapan significativamente, se considera válido en 3D. La figura 3 (arriba) se muestra un emparejamiento correcto, que da lugar a un segmento válido en el suelo. Esa misma figura, abajo, muestra un emparejamiento incorrecto.

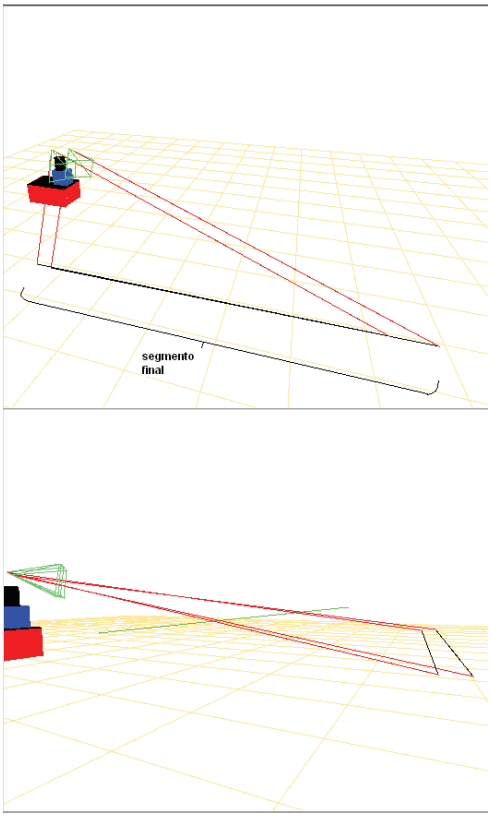

Fig. 3. Emparejamiento y triangulación de segmentos: válidos (arriba) en inválidos (abajo).

C. Control de atención

Con las dos secciones anteriores el sistema sabe ubicar en 3D las balizas y los segmentos del suelo que tiene delante de las cámaras. Ahora se desea que el robot represente la escena que le rodea, que es más amplia que su campo visual. Para conseguirlo se necesita mover las cámaras con un cuello mecánico.

El algoritmo de atención desarrollado busca nuevos objetos de la escena (balizas o segmentos suelo) mediante el movimiento de este cuello. Cuando se haya encontrado alguno se memorizará su posición 3D para ser capaz de atenderlo periódicamente. Estos objetos tendrán asociada una determinada vida, de forma que si hace tiempo que no se ve, su vida se reduce.

La forma que tiene el algoritmo de elegir los puntos a los que mirar, es mediante la elección de un punto de atención 3D. Estos puntos de atención pueden ser las balizas, puntos de un segmento del suelo o *puntos de exploración*. Estos puntos de

atención tendrán asignada una saliencia, que indicará las ganas que se tiene de mirar a dicho punto.

El sistema de atención hace uso de dos dinámicas concurrentes, la dinámica de saliencia y la dinámica de vida, sobre los objetos y los focos de atención. La dinámica de saliencia permite alternar la mirada entre diferentes puntos y saber a cuál de ellos hay que mirar en cada momento. La dinámica de vida permite saber cuántos objetos hay en la escena y si ha desaparecido alguno.

1) Dinámica de Saliencia: *Saliencia* es todo aquello que sobresale en una situación determinada. Cada foco de atención tiene asignada una saliencia, para que el cuello pueda moverse y apuntar al más saliente. Si se tiene un foco de atención con una saliencia muy alta, éste será visitado próximamente, ya que es un punto que llama la atención. En contra, si ésta fuera baja no será visitado.

Una forma de decidir la saliencia que posee cada foco de atención, es en función del tiempo que hace que no se visita. Un punto que hace tiempo que no se ha visitado causará mayor atracción, tendrá mayor saliencia, que uno que se ha atendido recientemente. La implementación de esta dinámica hace que cuando se visita un punto, su saliencia disminuya drásticamente y las de los demás focos no atendidos se incrementarán. Esto se representa en las ecuaciones (1), que se aplica cuando no se visita el punto, y (2), cuando se visita. Con ellas el punto que es atendido irá variando, puesto que cada vez habrá un punto diferente con mayor saliencia.

$$saliencia(punto, t) = saliencia(punto, t-1) + \Delta S_{time} \quad (1)$$

$$saliencia(punto, t) = 0 \quad (2)$$

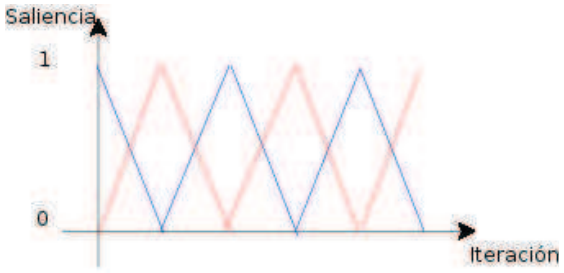

Fig. 4. La dinámica de saliencia con dos objetos, sin exploración, lleva a la alternancia.

Cuando un foco de atención es elegido se calcula el movimiento necesario del cuello mecánico para mirar a ese punto 3D absoluto. Para ello se tienen en cuenta la posición y orientación del robot, la posición relativa del cuello respecto al robot y la posición de las cámaras respecto al cuello. Con todo ello se calcula la rotación y traslación que deben efectuarse sobre las cámaras para que el punto 3D al que se debe atender cuadre dentro del campo visual central de ambas cámaras.

En el caso de las balizas es sencillo elegir su foco de atención 3D, ya que ella misma es un punto. Para elegir los foco de atención de los segmentos, se ha elegido la intersección suya con las direcciones significativas para navegar: por delante del robot y a 50^o por cada lado [1] (Figura 5). Si resultase que no cortan al segmento se elige el punto medio de éste.

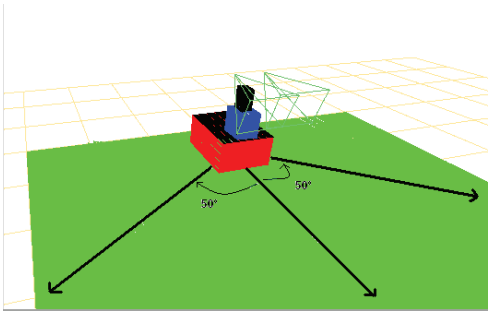

Fig. 5. Focos de atención generados por los segmentos suelo.

2) Dinámica de Vida: Con la dinámica de vida se pretende saber cuándo un objeto ha salido de la escena y si aún sigue en ella. Para ello si la vida de un objeto es superior a un cierto umbral, es que todavía sigue en la escena, pero si está por debajo es que ha desaparecido. Su funcionamiento es inverso al de la saliencia. Un objeto frecuentemente observado en las imágenes tendrá mayor vida que uno que apenas se visita. Si la vida de un objeto es inferior a un determinado umbral, este "morirá" y será, por tanto, olvidado y no se volverá a visitar.

La implementación de esta dinámica consiste en que, cada vez que se visita un objeto se incrementa su vida (hasta un determinado umbral) y la de los objetos no observados en esa iteración disminuye. Esto se representa en las fórmulas (3), cuando no se ve, y (4) cuando se ha visto.

$$vida(objeto, t) = vida(objeto, t-1) - \Delta V_{time} \quad (3)$$

$$vida(objeto, t) = vida(objeto, t-1) + \Delta V_{observacion} \quad (4)$$

El algoritmo permite asignar prioridades distintas a los diferentes objetos, ya sean segmentos del suelo o balizas. Concretamente, se ha optado por dar a los segmentos una vida mayor cada vez que se ven, ya que se consideran elementos fijos. No así las balizas, que en un determinado momento podrían moverse en tiempo de ejecución.

3) Exploración de la escena: Para buscar nuevos objetos relevantes en la escena se insertan de vez en cuando puntos de exploración a los que mirar con las cámaras, por si en esa zona hubiera objetos interesantes. Ésta búsqueda puede interesar, por ejemplo, al principio de la ejecución, momento en el que aún se desconocen las zonas de la escena donde hay objetos a atender. Estos nuevos puntos de exploración pueden ser de dos clases: *puntos de barrido* o *puntos aleatorios*.

Por un lado, los puntos de barrido van siguiendo una trayectoria definida para garantizar un recorrido completo de la escena. Concretamente, se irán dando puntos de atención alrededor del robot, siguiendo una trayectoria de derecha a

[1] En la solución final se ha contemplado la posibilidad de modificar esos ángulos en función de la velocidad del robot. Así, si el robot fuera deprisa sería interesante mirar más hacia delante y esto se consigue reduciendo el ángulo de 50^o.

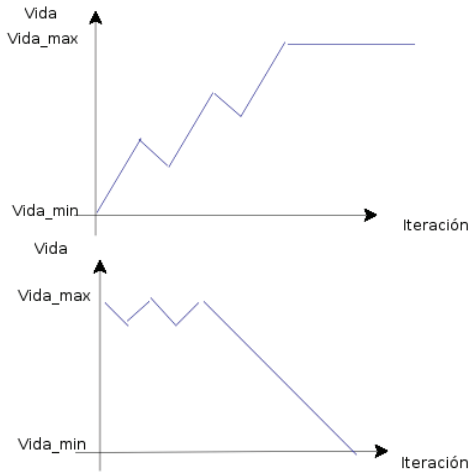

Fig. 6. Dinámica de vida con observaciones periódicas (arriba) y tras la desaparición del objeto (abajo).

izquierda, y de abajo a arriba. Por otro lado, los puntos aleatorios son generados aleatoriamente y sirven para encontrar nuevos objetos que pudieran aparecer de repente en la escena.

Ambos puntos de atención tienen una saliencia inicial alta para que sean visitados lo más pronto posible y de ese modo comprobar si en esa zona explorada existe alguna baliza o borde del suelo. Si como resultado de mirar a alguno de estos puntos de exploración se encuentra alguna baliza o borde del suelo, éste se insertará como un nuevo objeto que formará parte de la memoria local y del sistema de atención.

Para gobernar la inserción de estos puntos, cada tipo tiene asociada una probabilidad de inserción, en cada iteración.

4) Fusión con la memoria local: Cuando un segmento suelo o una baliza descubierta con anterioridad es de nuevo visitada, puede que se haya movido, o puede que la posición estimada por la triangulación no sea exactamente la misma. Por estos motivos, es necesario llevar a cabo un proceso de emparejamiento en el tiempo.

Para las balizas el emparejamiento es sencillo, si la posición nueva está en un determinado radio 3D alrededor de la anterior, se considerará que se trata de la misma. Como posición final se respeta la última medición.

Los segmentos se emparejan a lo largo del tiempo comprobando si solapan o son prolongación uno de otro. Si se obtiene un resultado positivo, se pasa a realizar la unión de ambos, ya que la meta es reconstruir el mayor trozo de pared posible.

Cuando hay una observación parcial de una baliza, que se ve en una cámara y en la otra no, o incompleta, no es posible calcular la posición 3D de la baliza. Cuando esto ocurre se inserta un punto de exploración futuro, con el fin de poder captar la baliza en las dos cámaras la próxima vez. El punto elegido está en la recta de retroproyección y a una distancia que sea la mínima donde se pueda ver la próxima vez desde ambas cámaras.

D. Navegación autónoma

Con los bordes del suelo detectados y almacenados en la memoria local se tiene la distancia a los obstáculos más próximos alrededor del robot. Sobre esa representación hemos usado un algoritmo reactivo de navegación local para gobernar el movimiento del robot.

En concreto hemos usado el algoritmo VFF (*Virtual Field Force*) [BK89], que atribuye fuerzas virtuales repulsivas a los obstáculos, inversamente proporcionales a la distancia, y fuerzas virtuales atractivas al destino local. El robot se mueve siguiendo la fuerza virtual resultante que combina ponderadamente las repulsivas con la atractiva.

IV. Experimentos

Las diferentes partes constituyentes del algoritmo se han implementado en lenguaje C como componentes en la infraestructura software *jdec* [CDBM06]. Para probarlas y validar el algoritmo atentivo de reconstrucción 3D visual hemos realizado un conjunto de experimentos, tanto sobre un simulador de imágenes propio como sobre un robot real. En concreto con el robot pioneer de la figura 7, que está equipado con un cuello mecánico Directed Perception y dos cámaras Isight formando un par estéreo móvil. Las cámaras tienen unos 55 grados de apertura.

En la web http://gsyc.es/jmplaza/research.html#attention se puede acceder a vídeos describiendo varios experimentos realizados con el algoritmo propuesto.

A. Repartiendo la mirada entre dos balizas separadas

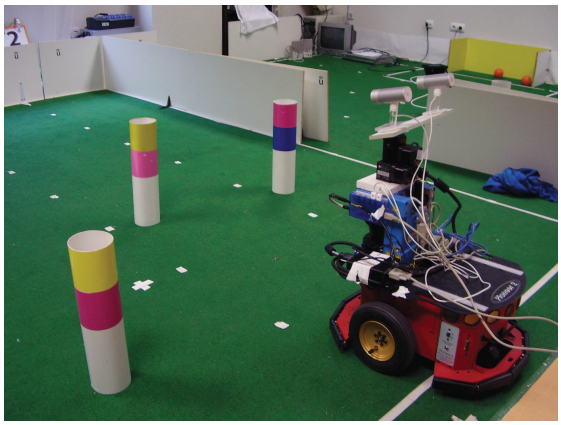

Fig. 7. Dos balizas interesantes no caben en un sólo vistazo de las cámaras.

En un primer experimento los únicos objetos de interés aceptados fueron las balizas rosa-amarillo. En la figura 7 se puede apreciar al robot frente a tres balizas, que no caben en un único vistazo de las cámaras debido a su ángulo de apertura limitado. En ese momento el robot tenía activos exclusivamente los filtros de color para el rosa y para el amarillo, por lo que ignora completamente a la baliza rosa-azul.

Tal y como muestra la figura 8, el algoritmo atentivo empieza a insertar puntos de barrido que hacen mover las cámaras explorando el entorno. A medida que va descubriendo balizas rosa-amarillo las va incorporando a la memoria local, y entran en el juego de la doble dinámica de saliencia y vida.

A partir de entonces los objetos existentes pugnan porque las cámaras los vuelvan a visitar, para subir su vida, y también se sigue explorando la escena con nuevos puntos de exploración.

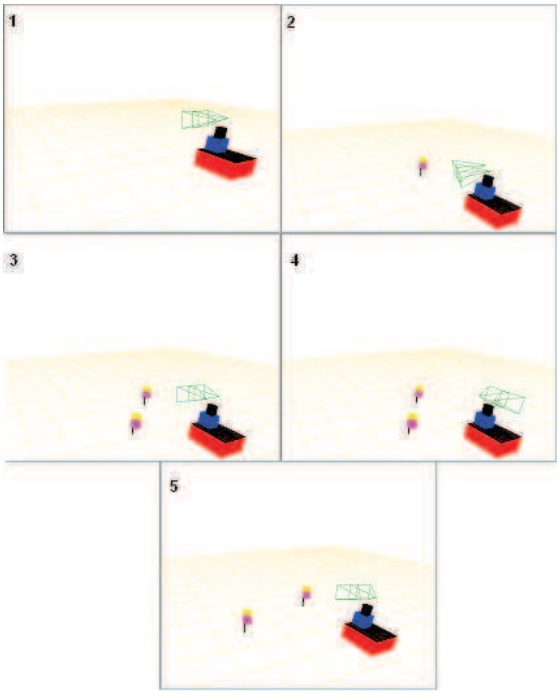

Fig. 8. Representación interna y reparto de la mirada entre dos balizas.

B. Balizas y obstáculos

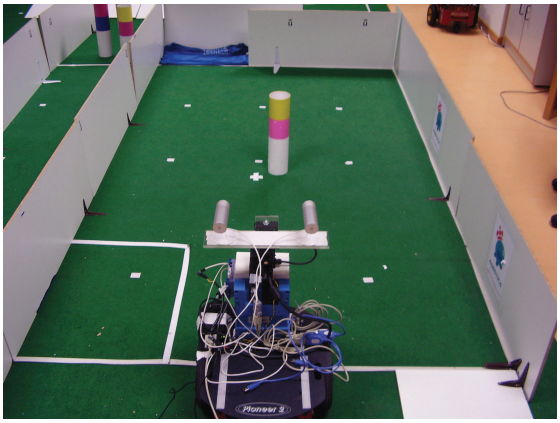

Fig. 9. Escena con baliza y bordes de suelo.

En un segundo experimento se incluyó también el suelo como objeto interesante, activando el filtro de color y cálculo de segmentos de suelo respectivo. Para la escena de la figura 9 se obtuvo la secuencia de movimientos de cámaras de la figura 10, en la que se observa la representación de la memoria local.

Se puede apreciar que inicialmente la representación interna está vacía y el algoritmo atentivo mueve las cámaras para explorar la escena, insertando puntos de barrido. A medida que se va barriendo la escena se detecta la baliza y las distintas porciones del límite del suelo. Los distintos segmentos del suelo se van incorporando y fusionando con los ya existentes según de observan nuevas zonas. Finalmente, tras un periodo de exploración, se completa una representación interna muy ajustada a la escena real, y el cuello mecánico alterna su movimiento apuntando a los distintos objetos de la memoria interna para manterlos frescos.

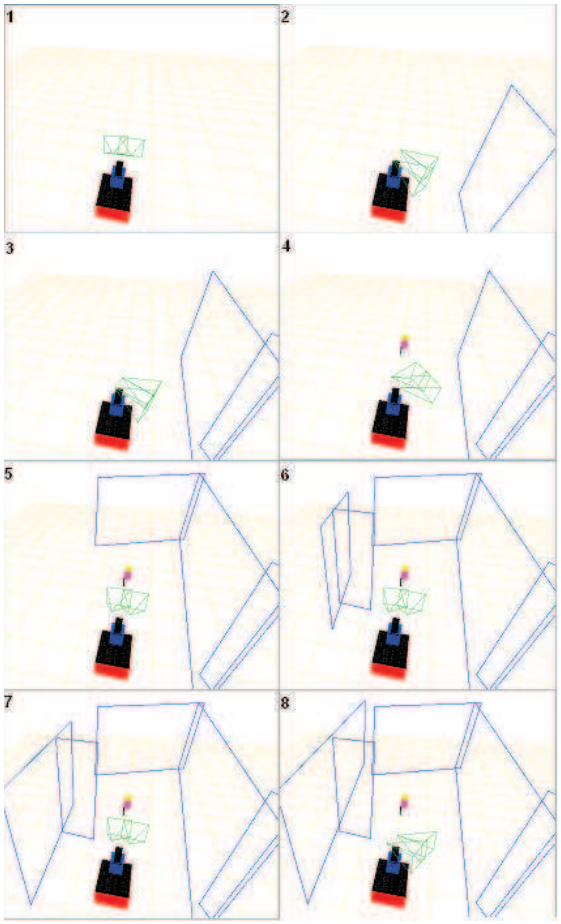

Fig. 10. Representación interna de escena con baliza y bordes de suelo.

C. Navegación sobre representación 3D atentiva

Un tercer experimento involucra el movimiento del robot usando exclusivamente la memoria visual local, rellenada y mantenida de manera atentiva. El plan de movimiento para el robot es: busca la baliza rosa-azul, acércate a ella, búsca la baliza amarillo-rosa y acércate a ella.

Para llevar a cabo esa secuencia el robot arranca explorando con su mirada el entorno, nuevamente siguiendo los puntos de barrido que inserta el algoritmo atentivo (puntos 1 a 9 de la figura 11). Durante ese barrido el robot ha ido almacenando en su memoria visual la posición de la baliza rosa-azul y de los bordes del pasillo. Cuando el robot ha barrido toda la escena,

y sólo entonces, se mueve siguiendo VFF sobre la información de obstáculos (límites del suelo) de la memoria visual (puntos 10, 11).

Una vez allí vuelve a explorar el entorno, esta vez buscando la baliza amarillo-rosa, y activando exclusivamente esos filtros de color (puntos 12,13,14). También sigue percibiendo los bordes del pasillo para refrescar y afinar la posición de los segmentos suelo respectivos en la memoria visual.

Finalmente, una vez que ha encontrado la baliza de destino y su posición tridimensional, inicia una nueva navegación VFF teniendo en cuenta la información recogida en la memoria visual (puntos 15 a 18).

Fig. 11. Representación interna de escena con baliza y bordes de suelo.

Es destacable que en todo este experimento no se utilizó ningún sensor de distancia a obstáculos. El robot navegó de manera segura utilizando exclusivamente sus cámaras y las lecturas de su odometría (tanto de las ruedas motrices como del cuello mecánico).

V. Conclusiones

En este artículo hemos presentado un algoritmo atentivo que gobierna el movimiento voluntario de un par estéreo de cámaras para que un robot perciba eficientemente su entorno. Por un lado explora la escena y por otro visita periódicamente los objetos interesantes descubiertos, para mantener su representación interna al día. Esa memoria visual almacena más información de la que cabe en un solo vistazo de las cámaras, debido a su limitada apertura angular.

El algoritmo utiliza dos dinámicas, una de saliencia y otra de vida, para construir y mantener actualizada la memoria visual, y para determinar hacia dónde debe mirar en todo momento. La saliencia de un punto aumenta con el tiempo y se anula cuando se apuntan las cámaras hacia él. Se elige como siguiente punto a mirar el de mayor saliencia y se ordena un movimiento sacádico de las cámaras hacia él. La vida de un objeto disminuye cuando no aparece en las imágenes y sube cuando sí lo hace.

El algoritmo atentivo se ha implementado y validado con varios experimentos sobre un robot real, como prueba de concepto. Los objetos de interés eran balizas de colores y los límites del suelo, que estaban orientados a tareas sencillas de navegación. Sin pérdida de generalidad se pueden sustituir por otros para otras tareas diferentes. Definir qué objetos son interesantes supone un primer filtro atentivo, que hace que el robot se fije sólo en las cosas que merecen la pena de las imágenes.

En los experimentos se aprecia cómo el algoritmo reparte alternativamente la mirada entre los distintos objetos interesantes de la escena y cómo la va interiorizando poco a poco.

En cuanto a las líneas futuras de continuación, estamos trabajando en la incorporación de objetos altamente dinámicos en el entorno y en el uso de movimiento como característica saliente, más allá de las descripciones de color. También estamos preparando un montaje con cámaras de amplia apertura angular y con transformaciones logpolar para simular lo que serían la periferia y la fovea de un ojo humano.

References

[H97] Ian Horswill. Real-time control of attention and behavior in a logical framework. Proceedings of the ACM International Conference on Autonomous Agents, 130–137, 1997.

[H97b] Ian Horswill. Visual architecture and cognitive architecture. Journal of Experimental and Theoretical AI, 9(2-3):277–292, 1997.

[V03] Scott Lenser y Manuela Veloso. Visual Sonar: Fast Obstacle Avoidance Using Monocular Vision. Proceeding of the IEEE Intelligent Robots and Systems, pp 886 - 891, 2003.

[B08] Pilar Bachiller. Percepción dinámica del entorno en un robot móvil. Tesis Doctoral, Universidad de Extremadura, 2008.

[C05] J.M.Cañas, M.Martínez-de-la-casa, P.Bustos, P.Bachiller. Overt visual attention inside JDE control architecture. IROBOT workshop inside Portuguese Conference on Artifical Intelligence EPIA'2005 pp 226-229, 2005.

[BBCR07] P.Bachiller, P. Bustos, J. M. Cañas, R. Royo. An experiment in distributed visual attention. 3rd Int. Workshop on Multi-Agent Robotic Systems (MARS-2006).

[CDBM06] J.M.Cañas, Pedro M. Díaz, Pablo Barrera, Víctor M. Gómez. Visual memory for robot navigation using JDE architecture. Proceedings of VII Workshop de Agentes Físicos, pp 91-98, 2006.

[BCE2007] B. Bonev, M. Cazorla, F. Escolano. Robot Navigation Behaviors based on Omnidirectional Vision and Information Theory. Proceedings of VIII Workshop de Agentes Físicos (WAF'2007), 2007.

[L06] O. León. Navegación de un robot con un sistema de atención visual 3D. Proyecto Fin de Carrera, Ingeniería Informática, Universidad Rey Juan Carlos, 2006.

[I02] L. Itti. Models of Bottom-Up and Top-Down Visual Attention Tesis doctoral, California Institute of Technology, 2002.

[IKB00] L. Itti, C. Koch, J. Braun. Revisiting Spatial Vision: toward a unifying model. Journal of the Optical Society of America, 17(11), 2000.

[NI05] V. Navalpakkam, L. Itti. Modeling the influence of task on attention. Vision Research, 45:205-231, 2005.

[T95] J.K. Tsotsos, et.al. Modeling visual attention via selective tuning. Artifical Intelligence 78:507-545, 1995.

[BS99] C. Breazeal, B. Scassellati. A context-dependent attention system for a social robot. Proc. of Int.J.Conf. on Artificial Intelligence, 1146-1151, 1999.

[OMS05] F. Orabona, G. Metta, G. Sandini. Object-based visual attention: a model for a behaving robot. Proc. of Int. Workshop on Attention and Performance in Computational Vision WAPCV-2005

[ZRT04] A. Zaharescu, A.L. Rothenstein, J.K. Tsotsos. Towards a biologically plausible active visual search model. Proc. of Int. Workshop on Attention and Performance in Computational Vision WAPCV-2004, Springer LNCS 3368:133-147, 2005.

[BK89] J. Borenstein y Y. Koren. Real-time obstacle avoidance for fast mobile robots. IEEE Transactions on Systems, Man, and Cybernetics, 19(5):1179-1187, 1989.