

Modelado 3D dinámico del entorno por un robot móvil con visión estereoscópica

Bachiller P. Vicente J. Bustos P.

Laboratorio de Robótica y Visión Artificial
Departamento de Informática
Universidad de Extremadura
Avda de la Universidad sn 10071 Cáceres SPAIN
pilarb@unex.es

Resumen

En este trabajo se introduce al robot móvil Tornasol, como una evolución del robot Murphy presentado en la anterior edición de esta conferencia. Sobre esta nueva plataforma se han continuado las investigaciones en torno a la generación dinámica de comportamientos y los procesos de visión estereoscópica. La propuesta teórica y los resultados que presentamos se enmarcan dentro de lo que se conoce como aproximación dinámica al comportamiento inteligente. En líneas generales, el robot Tornasol y su entorno configuran un sistema dinámico en el cuál se alternan situaciones de equilibrio y desequilibrio. Esta situación cambiante genera simultáneamente actividad externa - movimientos sacádicos y de traslación de la base- y actividad interna - representación del espacio 3D y refinamiento de esta representación. A diferencia de otros enfoques, en esta propuesta no se asume un conocimiento simbólico explícito sobre objetivos y objetos en el mundo, sino una relación dinámica entre los procesos visuales y motores a través de una representación interna de la estructura 3D del entorno. Esta propuesta se está desarrollando sobre el robot Tornasol y presentamos aquí algunos de los resultados obtenidos.

1. Introducción

Durante los últimos años se ha estudiado en profundidad la relación existente entre la estructura 3D y la estructura 2D captada por varias cámaras que realizan una transformación proyectiva del espacio. La conclusión más relevante de estos estudios es que, desde el punto de vista geométrico, es posible invertir el proceso de proyección recuperando la estructura original a partir de las señales de luminancia de dos o más vistas en distintas posiciones del espacio. Además, ésta reconstrucción es posible hasta un factor de escala sin el conocimiento previo de los parámetros intrínsecos o de calibración de las cámaras y sin conocimiento previo de las posiciones en el espacio donde se sitúan. Por medio de la geometría proyectiva se pueden determinar qué movimientos son adecuados para la calibración y cuáles permiten una estimación de la estructura tridimensional. Analizando el caso menos restrictivo, donde no existe conocimiento sobre la estructura 3D ni sobre los parámetros intrínsecos e extrínsecos de las cámaras, se puede concluir que la autocalibración con la propia escena necesita de una componente de rotación en los movimientos de los sensores junto con una mínima suposición de área sensible cuadrada. Por el contrario, la estimación de puntos 3D necesita de una componente de traslación entre las vistas [1,2].

Pese a la generalidad de estos resultados es importante analizar las limitaciones de la teoría y del paso de la teoría a la ingeniería. Al plantear como objetivo práctico de este trabajo la construcción de un robot móvil que se desplace por un entorno real utilizando sus sensores de visión, es posible definir a priori cuáles serán las hipótesis de partida y qué aspectos debe resolver el robot sobre la marcha. Una alternativa posible es evitar la necesidad de una representación espacial. 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. Existen, incluso, robots de este tipo que utilizan esencialmente información visual [3]. Estos logros se han debido, fundamentalmente, a un cambio conceptual en la forma de diseñar las arquitecturas de control inteligente promovido, entre otros, por R. Brooks durante los años ochenta [4]. Estos resultados cuestionan la necesidad de una representación explícita del espacio y sugieren una interpretación directa más en la línea de los *affordances* de Gibson [5]. 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: ¿cómo se explica que a pesar de que los ojos y el cuerpo estén en movimiento, percibimos un mundo estático?. Y extendiendo esta idea, 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 [6]. Además, a pesar de que se puedan establecer multitud de interacciones útiles entre el entorno y el robot, 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 modelo que pueda ser propagado internamente.

En este trabajo mantendremos la hipótesis de que es necesario algún tipo de representación del espacio, aunque esta representación debe tener una naturaleza dinámica. En este sentido nos apartamos de formas de representación espacial monolíticas y centrales, ligadas a la acción por un sistema de manipulación simbólica. En su lugar, planteamos una representación del espacio como una memoria dinámica directamente acoplada e inseparable de los sistemas de acción del robot. Este acoplamiento implicará que lo que se almacene en la memoria afectará funcionalmente al movimiento y, a su vez, el movimiento afectará a los contenidos de la memoria.

Esta propuesta es un paso dentro de un modelo más general sobre los agentes inteligentes que no desarrollaremos aquí. Algunos aspectos de este modelo están relacionados con la caracterización de objetos en esta memoria activa, la memoria a largo plazo o la extensión de la capacidad predictiva e imaginativa. La importancia de lo que hemos denominado memoria activa es fundamental en la validación y desarrollo de las diferentes hipótesis que articulan este nuevo modelo.

Una vez definida la posición que adoptaremos, pasamos a analizar en que medida la teoría de la visión geométrica nos va a permitir desarrollar nuestro objetivo, y cómo complementaremos aquellos aspectos que queden fuera de esta teoría.

En general, la base de la reconstrucción 3D con métodos geométricos está en encontrar parejas de puntos homólogos. Estos puntos permiten llevar a cabo una triangulación sencilla para recuperar la profundidad perdida en el proceso de proyección. Si imponemos unos requerimientos mínimos de tiempo real para las aplicaciones de robótica, la búsqueda de homólogos debe restringirse a aquellos puntos de las imágenes que, normalmente, se denominan esquinas. Estos puntos tienen la particularidad de estar rodeados en la imagen de un alto contenido en textura. Aunque es posible establecer correspondencias entre segmentos y otras formas de orden superior, para que el proceso sea fiable se requiere normalmente un seguimiento continuo sobre una secuencia de imágenes, en lugar de un par estéreo.

El problema con la extracción de esquinas es que éstas representan sólo un muestreo muy pequeño del contenido total de la imagen. 200 esquinas de una imagen de 16K pixels suponen en torno al 1 por ciento del total de pixels. El problema es, por tanto, recuperar el resto de la estructura sin disponer de medidas adicionales provenientes del sistema estéreo.

Recientemente Keriven y Faugeras en [7] han propuesto un marco matemático para recuperar la estructura 3D sin necesidad de obtener previamente esquinas u otros elementos estructurales 2D. Conceptualmente la idea es sencilla si se piensa el problema al revés de como se hace normalmente. Se parte de una hipótesis de estructura 3D inicial y se va deformando progresivamente guiada por un criterio de optimización global. Este criterio no es más que el criterio de correspondencia entre homólogos. Así, un punto de esta estructura hipotética satisfará parcialmente este criterio si su proyección sobre las cámaras del par estéreo genera dos puntos cuyos entornos están altamente correlados. Técnicamente, el desarrollo matemático de esta idea y su adaptación a un algoritmo estable distan mucho de ser sencillos. La función de optimización global se interpreta como una función de energía que se transforma en su ecuación diferencial asociada. Algunos de los problemas más relevantes tienen que ver con la discretización y estabilidad de los métodos numéricos de integración, los tiempos de convergencia o la poca información que aportan las zonas poco texturadas, donde la función de correlación poco puede aportar.

Aunque la propuesta es extremadamente elegante y general, su utilización en un robot móvil no es viable actualmente y, quizás, no lo sea nunca. Nuestro objetivo es recoger la idea general y plantear un método que pueda funcionar de forma mucho más eficiente. Un requerimiento compartido es que la reconstrucción sea progresiva, de tal forma, que el robot pueda optar por seguir mejorando su representación o desplazarse a otro lugar con la información disponible. Además, este proceso debe complementarse y acelerarse con el reposicionamiento de las cámaras -movimientos sacádicos- y la translación de la base.

El modelo que proponemos se puede estructurar como un bucle continuo que comienza con la obtención de la estructura 3D, continua re proyectando esta estructura obtenida sobre las propias cámaras para estimar la calidad del modelo y, finalmente, se generan movimientos de las cámaras y la base dirigidos a disminuir este error. Después de realizar el movimiento el bucle comienza de nuevo. Este bucle enlaza percepción a acción para obtener y mejorar el modelo 3D del entorno. Actualmente, estamos trabajando en otros bucles que funcionarían

simultáneamente pero con otras constantes temporales, que enlazarían percepción y acción para desplazar al robot a ciertos lugares o para interactuar con ciertas formas del entorno mediante un manipulador.

La superficie matemática que utilizamos como modelo 3D del entorno es continua y definida sobre la semiesfera situada delante del robot. Al cubrir un espacio mayor que el del campo visual de las cámaras actúa como memoria espacial. En cada punto de esta semiesfera se almacena información de profundidad, nivel de gris y orientación local, lo que hace que pueda verse como un campo vectorial de varias variable y dependiente de dos parámetros (u,v) que codifican la dirección de cada punto en la esfera. Este campo está sujeto a una ecuación dinámica interna que le hace difundirse según la ecuación del calor y sujeto a unas condiciones de contorno fijadas por las medidas 3D del sistema estereoscópico. Además está conectado al sistema motriz bidireccionalmente, de tal forma que ciertas activaciones en el campo producen movimientos en el robot y, cada movimiento del robot actualiza la posición de todo el campo respecto a un sistema de referencia centrado en el centro de la cabeza.

En el resto del trabajo introducimos al robot Tornasol, abordamos en detalle la arquitectura del modelo y presentamos los resultados experimentales obtenidos.

2. El Robot Tornasol

El robot Tornasol ha sido construido en el laboratorio de robótica y visión artificial de la Escuela Politécnica de la Universidad de Extremadura. La construcción ha abarcado un diseño integrado de los aspectos mecánico, electrónico y de software de los diferentes procesadores de la arquitectura.

La estructura del robot consta de una plataforma móvil y una torreta de visión estereoscópica construidas en aluminio. Sobre la plataforma hay colocado un doble rack. La parte superior contiene las tarjetas de control de los motores y las etapas de potencia. En la parte inferior se ha situado una placa base de PC.

La torreta contiene dos cámaras digitales que utilizan el protocolo serie IEEE1394 con una velocidad de transferencia de 400 Mbps. El dispositivo CCD contiene $692 \times 504 \approx 350K$ elementos sensibles de area cuadrada $5,6\mu \times 5,6\mu$ y lectura progresiva. El modelo del sensor es el ICX098BQ de Sony. A través del enlace IEEE1394 se pueden programar los distintos modos de captura de la cámara, así como parámetros de funcionamiento como ganancia, brillo, saturación, balance de blancos, obturador electrónico, etc.

El proceso de las imágenes se realiza en una placa con un procesador AMD Athlon a 1.4 GHz. En este proceso se emplean los registros y operaciones MMX para enteros y 3DNow en coma flotante que están disponibles en este microprocesador. Con esta tecnología el microprocesador realiza por hardware dos operaciones en coma flotante en cada ciclo de reloj. Con el fin de optimizar al máximo este proceso, durante los tres últimos años se ha desarrollado una librería con todas la funciones que se realizan sobre las imágenes. El número real de operaciones que el micro es capaz de realizar con este software es del orden de 1 Gflops.



Figura 1: Robot Tornasol.

3. Obtención de la Estructura 3D

La obtención de la estructura 3D se realiza en tres pasos:

1. Extracción de esquinas y cálculo de pares de puntos homólogos.

2. Detección de regiones planas en la escena.
3. Difusión de la estructura.

Estos tres pasos conducen a una primera aproximación del modelo 3D. Para conseguir un sistema eficiente, cada uno de los pasos puede verse como un refinamiento del anterior. El primero, extrae un conjunto pequeño de medidas 3D desde los puntos esquinas que se pueden poner en correspondencia. Con estas medidas se inicializa la superficie 3D que modela el entorno. El segundo actúa sobre este primer modelo haciendo hipótesis sobre la planaridad de ciertas regiones. Finalmente, se hace variar al modelo en el tiempo siguiendo la ecuación de difusión del calor con las condiciones de contorno fijadas en los dos pasos anteriores.

3.1. Extracción de esquinas y cálculo de correspondencias

Para poder recuperar la información tridimensional a partir de las imágenes capturadas por las cámaras de un sistema estéreo, es necesario realizar una correspondencia entre puntos de las dos imágenes. Este proceso puede simplificarse considerablemente si previamente se extrae un conjunto de puntos relevantes de cada imagen. A esta fase previa a la correspondencia se la conoce como detección de esquinas. El método empleado para llevarla a cabo es el filtro de Harris, cuyo funcionamiento fue presentado para el robot Murphy en la pasada edición de esta conferencia [8].

Los conjuntos de puntos obtenidos en el proceso de extracción de esquinas constituyen la entrada a la fase de correspondencia. 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 imágenes [9].

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. El conjunto de puntos 3D, que se estiman a partir de las correspondencias en el par estéreo, constituye la estructura inicial del espacio. La representación de esta estructura inicial se realiza a través de una cámara virtual (o cámara central) situada en el sistema de referencia del robot, entre las dos cámaras de la torreta. La figura siguiente muestra estos elementos.

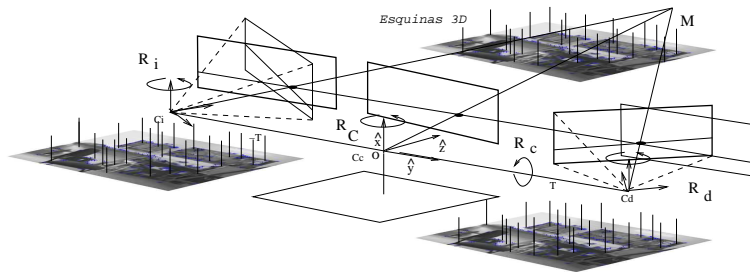


Figura 2: Fusión de las imágenes de esquinas de cada cámara en una sola estructura 2D. Las posiciones activadas se enlazan formando una correspondencia. Las posiciones del par enlazado se combinan por medio de las orientaciones de las cámara obteniéndose el punto 3D. Este se proyecta sobre una cámara virtual situada entre las dos cámaras reales, resultando en la activación de una posición sobre la nueva imagen central. Para que la imagen central sea capaz de integrar las esquinas enlazadas, la cámara del centro se orienta para que abarque el campo de visión común entre las dos cámaras reales. Esta condición se consigue cuando el eje óptico de la cámara central coincide con la bisectriz del ángulo de vergencia.

Esto permite integrar la información extraída por el par de cámaras en una única representación que mantiene la estructura inicial de los puntos del espacio. El proceso que construye esta representación consiste en la proyección de cada punto 3D en la cámara virtual. Esta proyección permite activar una posición de la nueva imagen en la que se almacena la información correspondiente a cada punto de la escena.

Para que esta imagen sea capaz de integrar las esquinas enlazadas, todos los puntos 3D estimados tienen que ser visibles por la cámara virtual. El único requisito para conseguirlo es que ésta pueda orientarse correctamente para abarcar el campo visual común a las dos cámaras. Al igual que las cámaras laterales de la torreta, los movimientos de la cámara virtual son rotaciones puras sobre su centro óptico. Cualquier movimiento en los motores de la torreta induce un movimiento de rotación pura en la cámara virtual que viene determinado por la rotación de las cámaras laterales. La nueva orientación se calcula imponiendo que el eje óptico de la

cámara central coincide con la bisectriz del ángulo de vergencia determinado por las cámaras laterales. Con esta restricción se garantiza que el ángulo de visión coincide con el común a las dos cámaras.

Tras la fase de correspondencia entre esquinas y a través de la proyección en la cámara virtual de los puntos de la escena calculados, es posible obtener una representación de una primera aproximación a la estructura 3D visible en el campo perceptivo de la torreta. Para refinar esta representación inicial, es necesario inferir información adicional a partir de las posiciones de las esquinas en el espacio y de las imágenes capturadas por las cámaras.

3.2. Detección de Segmentos y Planos

El conjunto de puntos obtenidos a partir de la correspondencia entre esquinas no es suficiente para que la reconstrucción sea útil al robot. Es necesario obtener información adicional para completar la representación de la escena.

En las zonas donde la señal mantiene una orientación bien definida, la señal de luminancia proporciona información añadida susceptible de ser interpretada en la estructura 3D. Cualitativamente, se corresponden con segmentos 2D que actúan de frontera entre dos valores de luminancia diferente. La importancia de estos segmentos es que se corresponden con rectas 3D en el espacio que unen superficies con colores diferentes. La detección de estas zonas en las imágenes se lleva a cabo aplicando un filtro Laplaciano a cada una de ellas. De esta forma, se obtienen dos imágenes de bordes que pueden ser estudiadas para extraer información de segmentos en la escena.

Una vez que se han establecido los segmentos en las imágenes del par estereó, es necesario abordar el problema de la correspondencia entre ellos para poder así determinar la estructura de las rectas 3D. Sin embargo, la búsqueda de segmentos homólogos es muy inestable debido fundamentalmente a la falta de coincidencia entre los extremos de los segmentos de las dos imágenes. Para intentar solucionar este problema, el detector propuesto impone la restricción de que los extremos de los segmentos coincidan con esquinas enlazadas en la fase de búsqueda de puntos homólogos. Esto proporciona dos claras ventajas. Por un lado, resuelve el problema de inestabilidad en la búsqueda de segmentos homólogos. Por otro lado, la detección de segmentos en las imágenes permite conocer directamente la estructura del segmento en la escena, al estar éste definido por puntos 3D determinados en la fase anterior.

Para establecer el conjunto de segmentos candidatos que serán comprobados en las imágenes de bordes, se realiza una triangulación de Delaunay [10] sobre los puntos fijados en la imagen central durante la fase de correspondencia entre esquinas. Este método permite interpolar una superficie mediante triángulos imponiendo criterios de optimización con respecto al tamaño, los ángulos y la longitud de los lados de cada triángulo. La condición de que exista un segmento en la escena se comprueba por medio de los lados de cada triángulo resultante y las imágenes de bordes. Esta comprobación se lleva a cabo siguiendo el criterio de proximidad de cada segmento obtenido en la triangulación a un borde de imagen. Adicionalmente, se considera segmento en la escena a todo aquel que se corresponde con una zona sin textura en la imagen. El resultado de este proceso es un conjunto de segmentos 3D que pueden ser proyectados y fijados en la cámara central.

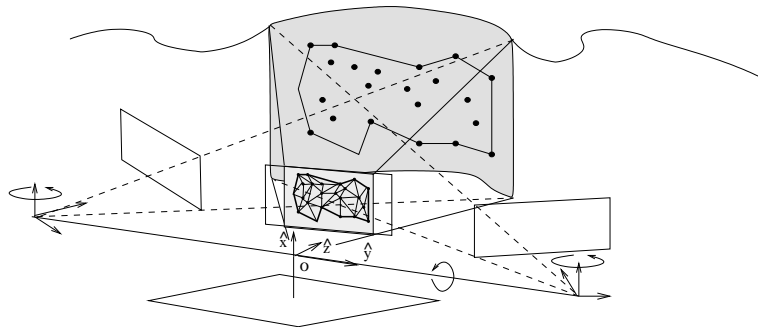


Figura 3: El campo de visión común entre las dos cámaras laterales se proyecta como un franga en la cámara virtual, el ancho de esta franja se determina a partir del ángulo de vergencia entre las cámaras. Esta zona se corresponde con el máximo dominio, sobre la cual el algoritmo de triangulación puede realizar la partición 2D. El dominio real quedará finalmente definido por el conjunto de posiciones en las que existen esquinas activas.

Junto con la detección de segmentos, la triangulación de la superficie 3D permite además la detección de superficies planas en la escena. Así cada triángulo obtenido se considerará plano en la escena si sus lados han sido detectados como segmentos.

Tras la extracción de segmentos y planos y su posterior proyección en la cámara central, se obtiene una representación de la estructura 3D que, aunque no es completa, permite aproximar el espacio percibido a una superficie 3D que proporciona información suficiente para que el robot puede llevar a cabo las acciones correspondiente.

3.3. Difusión de la Estructura

La aproximación de la estructura obtenida por esquinas, segmentos y planos a una superficie 3D se realiza a través de una difusión sobre la representación resultante de la proyección de dicha estructura en la cámara central. El objetivo de esta difusión es interpolar la señal de distancia en aquellas zonas que contienen solo componentes de baja frecuencia. En este proceso, la señal de distancia $\rho(u, v)$ en el instante de muestreo en curso, constituye el estado inicial de una función dependiente de tiempo $\rho(u, v, t)$, cuya evolución entre periodos de muestreo está controlada por la siguiente ecuación de difusión.

$$\frac{\partial \rho(u, v, t)}{\partial t} = \text{div}(\nabla_{u,v} \rho(u, v, t)) = \left(\frac{\partial^2 \rho(u, v, t)}{\partial u^2} + \frac{\partial^2 \rho(u, v, t)}{\partial v^2} \right) = \Delta_{u,v} \rho(u, v, t) \quad (1)$$

$$\rho(u, v, t = 0) = \rho(u, v) \quad (2)$$

$$\rho(u_f, v_f, t) = \rho(u, v) \quad (3)$$

En la difusión de la estructura, se establecen un conjunto de puntos que permanecen fijos $\rho(u_f, v_f)$, sobre los que la ecuación anterior no tiene efecto. Estos puntos fijos están constituidos por la estructura inicial de alta frecuencia formada por los puntos, segmentos y triangulos que se establecieron anteriormente. Utilizando un simil con la ecuación de difusión del calor, estos puntos se corresponden con puntos de temperatura constante que ceden o absorben energía de los puntos vecinos, determinando de esta forma la evolución final de la temperatura en los puntos de la imagen.

4. Representación del Espacio

Una vez que se ha difundido la estructura, se obtiene una representación 3D completa de los puntos que caen dentro del campo de visión de la cámara central. No obstante, a lo largo del tiempo, las cámaras de la torreta modificarán su orientación induciendo un movimiento de rotación en la cámara virtual. Este movimiento provoca que se procese la estructura 3D de otra zona del espacio y, como consecuencia, que se pierda la información que se obtuvo en instantes anteriores.

Para mantener la estructura espacial correspondiente a intervalos de muestreo pasados, se recurre a otra representación que permita integrar la estructura actual con la obtenida en instantes anteriores. Esta representación tiene que soportar las posiciones de aquellos puntos del espacio que no están dentro del campo de visión en curso. A medida que se producen movimientos en el robot, esta representación se completa con las zonas del espacio que la cámara virtual recorre a lo largo del tiempo.

La estructura 2D que soporta este esquema es la superficie de una esfera de radio unidad, cuyo centro coincide con el sistema de referencia del robot (figura 4). Cada posición sobre la superficie de la esfera representa la dirección de un rayo que parte de su centro y atraviesa dicha posición. Por medio del rayo asociado a cada posición, se puede establecer una correspondencia entre los puntos de la esfera y una superficie de puntos que rodea al robot. De esta forma, podemos considerar la esfera como una señal bidimensional definida por la activación de cada posición junto con una propiedad cuantitativa (distancia, color, etc.) relacionada con el punto correspondiente del espacio.

Las posiciones sobre la esfera se activan con las señales provenientes de la cámara central en el instante actual. El proceso de activación se lleva a cabo por medio de una proyección de cada posición de la imagen central sobre la esfera. El resultado de esta proyección es un área sobre la superficie de la esfera correspondiente al ángulo sólido del campo de visión del robot.

Para poder representar la superficie de la esfera en la memoria de un computador es necesario realizar una discretización de la misma que permita obtener un array de puntos asociados con puntos de la esfera. Con el objetivo de mantener constante la relación entre elementos de área del espacio definido por la superficie de la esfera y el resultante de la discretización, proponemos lo que hemos denominado transformación por longitudes de arco sobre paralelos. Esta transformación relaciona cada punto $p_s(x_s, y_s, z_s)$ de la superficie de la esfera con un punto en el dominio de los parámetros $p_d(i, j)$ cuyas coordenadas representan longitudes de arco. Estas longitudes de arco se miden sobre las circunferencias (c_i, c_j) que se obtienen de la intersección de

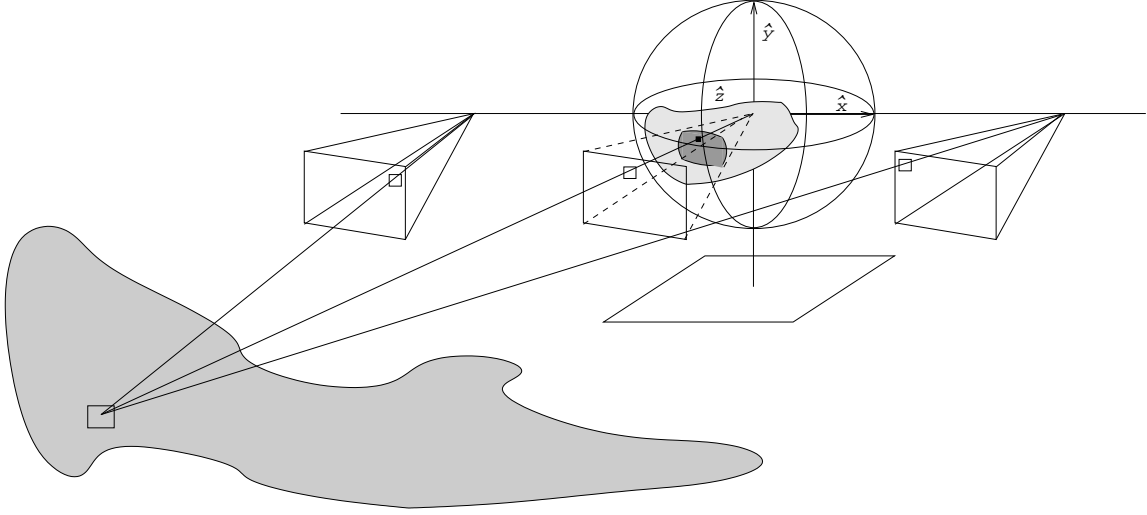


Figura 4: La orientación de las cámaras laterales de la torreta, definen una rotación en la cámara virtual. Esta rotación define a su vez, una zona del espacio donde se determina la estructura 3D en el periodo de muestreo actual. De esta forma, se establece una correspondencia entre los puntos de la esfera y la superficie 3D que rodea al robot.

la esfera con los planos $\pi_i : (y = y_s)$ y $\pi_j : (x = x_s)$. Considerando la división habitual de la esfera terrestre en paralelos y meridianos, los paralelos c_i coinciden con los paralelos terrestres. Análogamente, los paralelos c_j provienen del mismo concepto extendido a la dirección perpendicular. Los índices (i, j) , resultantes de aplicar esta transformación sobre la superficie de la esfera, se corresponden con las longitudes de arco sobre estos paralelos medidas desde los puntos $x = 0$ de c_i e $y = 0$ de c_j , respectivamente.

Según se puede observar en la figura 5, los paralelos horizontales y verticales, se corresponden con circunferencias de radio $r_i = \sqrt{1 - y_s^2}$ y $r_j = \sqrt{1 - x_s^2}$, respectivamente. Los ángulos se calculan desde el origen de longitudes hasta el punto de la superficie, por medio de las ecuaciones: $\alpha_i = \arcsin\left(\frac{x_s}{\sqrt{1 - y_s^2}}\right)$ y

$$\alpha_j = \arcsin\left(\frac{y_s}{\sqrt{1 - x_s^2}}\right).$$

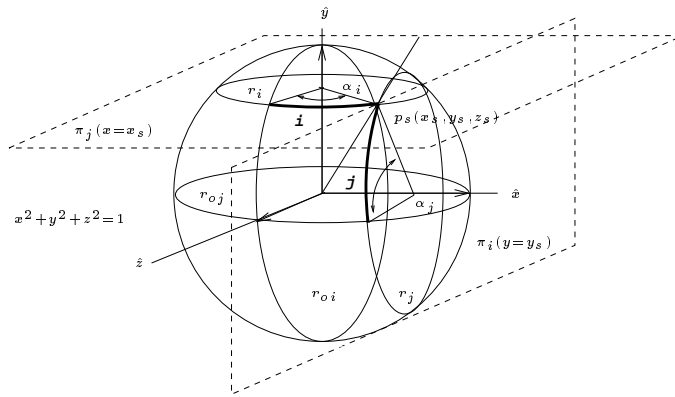


Figura 5: Interpretación geométrica de la transformación: longitudes sobre paralelos.

Finalmente, la transformación que relaciona los puntos de la esfera con las longitudes (i, j) o posiciones del array, se expresan mediante las siguientes ecuaciones, en las que se ha añadido el factor de escala $\sqrt{2} \left(\frac{2}{\pi}\right)$, cuyo objetivo es mantener una relación de unidad entre las áreas globales de la esfera y del círculo que define

el dominio de longitudes.

$$i = \sqrt{2} \left(\frac{2}{\pi} \right) \sqrt{1 - y^2} * \arcsin \frac{x}{\sqrt{1 - y^2}} \quad (4)$$

$$j = \sqrt{2} \left(\frac{2}{\pi} \right) \sqrt{1 - x^2} * \arcsin \frac{y}{\sqrt{1 - x^2}} \quad (5)$$

En la siguiente figura se muestra la estabilidad de la transformación propuesta en cuanto a la relación entre las áreas de las superficies original y discreta. Como puede observarse, esta transformación permite mantener constante la relación entre elementos de área del espacio original y del resultante de la discretización en gran parte de la superficie de la esfera, lo que proporciona grandes ventajas a la hora de almacenar y procesar la información representada.

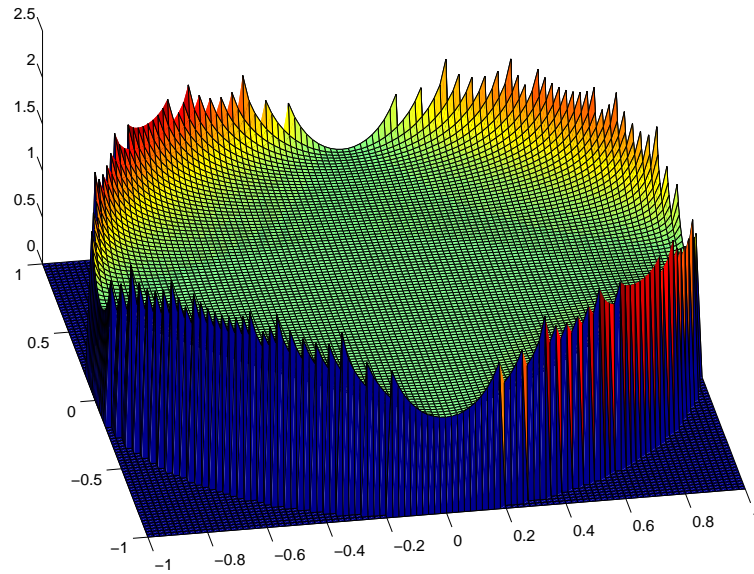


Figura 6: Relación entre elementos de área de las superficies original y discreta.

5. Experimentos

En esta sección presentamos algunos resultados experimentales realizados con el robot Tornasol. Las figuras 7 a 9 muestran el modelo 3D creado por el robot a partir de un par estéreo.

El siguiente grupo de figuras muestran el estado de la memoria visual después de realizar varias rotaciones puras.

El último grupo (figura 12) muestra una secuencia de movimientos sacádicos generados por el robot hacia zonas de alto error de reproyección.

La Figura 11 muestra el error de reproyección durante la secuencia anterior.

6. Conclusiones

En este trabajo se ha presentado al robot Tornasol como una plataforma de investigación sobre la generación dinámica de comportamientos y los procesos de la visión estereoscópica. Partiendo de la hipótesis de que es necesario mantener algún tipo de representación del espacio, la propuesta realizada se basa en considerar dicha representación como una memoria dinámica directamente ligada a los sistemas de acción del robot. Este acoplamiento permitirá que sea posible crear en el robot diferentes comportamientos visuales y motrices dirigidos a mantener, mejorar y ampliar su propia representación.

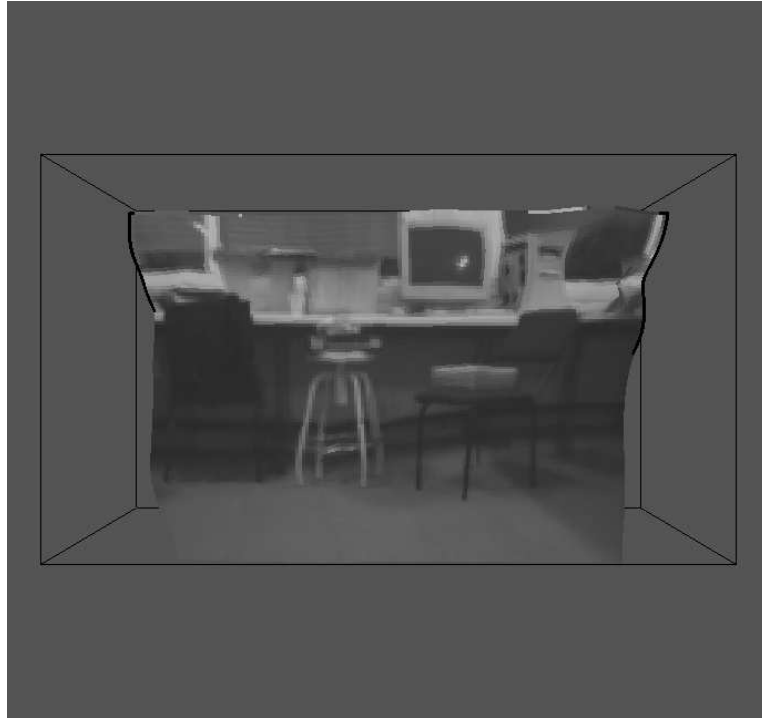


Figura 7: Vista frontal.

7. Bibliografía

1. ¹ (Artificial Intelligence Laboratory. MIT)
2. Agapito de L. Hayman E. Reid I. Self-Calibration of Rotating and Zooming Cameras. Proceedings of the Computer Vision and Pattern Recognition Conference, June 1999.
3. Eric Hayman, Lourdes de Agapito, Ian D. Reid and David W. Murray. The Role of Self-Calibration in Euclidean Reconstruction from Two Rotating and Zooming Cameras. Proceedings of the European Conference on Computer Vision, 26th June - 1st July 2000
4. Brooks R. Intelligence Without Representation, Artificial Intelligence, Vol 47, No.1-3, 1991
5. Gibson.J.J. The Ecological Approach to Visual Perception. LEM 1982
6. Bridgeman B., van der Heijden A.H.C, Velichkovsky D.M., A Theory of Visual Stability across saccadic eye movements. Behavioral and Brain Sciences 1782:247-292, 1994
7. Faugeras O., Keriven R., Variational Principles, Surface Evolution, PDE's, level set methods and the Stereo Problem, Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis of the IEEE Transactions on Image Processing, Vol. 7, No. 3, pages 336-344, March 1998
8. Bustos P., Bachiller P., Vicente J. Broncano M, Fernández C. Murphy, hacia un robot con visión estereoscópica. II Workshop Hispano-Luso en Agentes Físicos, Madrid 2000.
9. Falkenhagen, L. Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints. International workshop on SNHC and 3D Imaging, September 1997.
10. Shewchuk, J.R. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. First Workshop on Applied Computational Geometry:124-133. May 1996.

¹<http://www.ai.mit.edu>

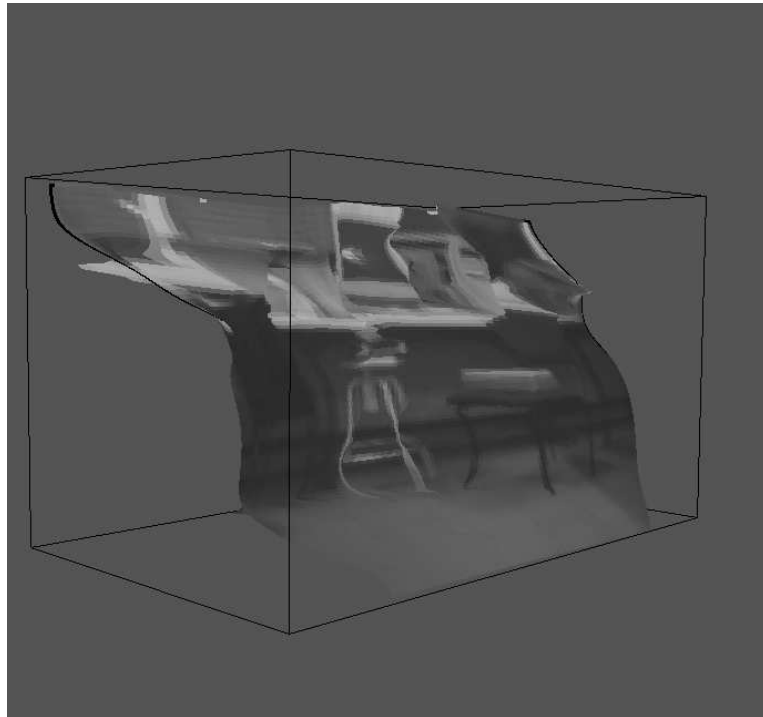


Figura 8: Vista lateral izquierda.

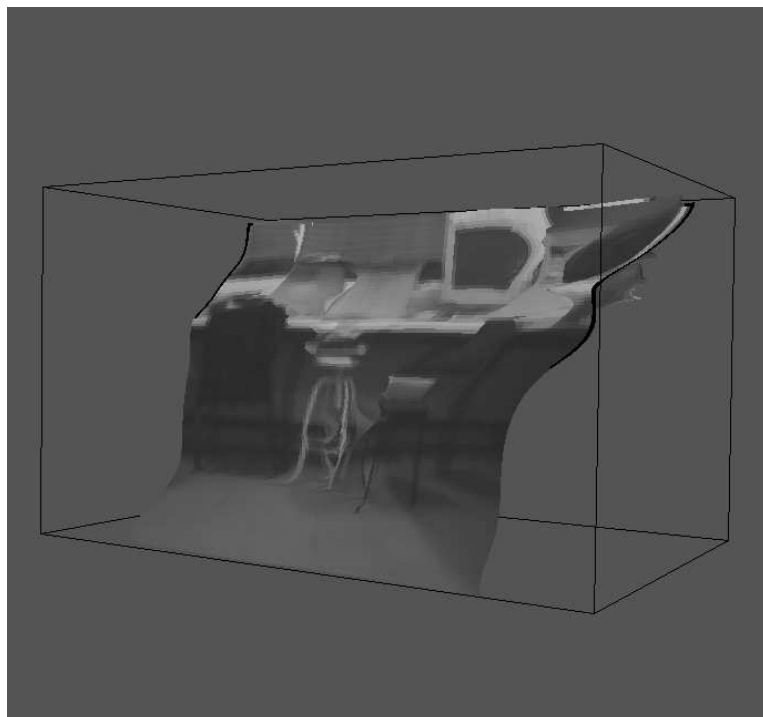


Figura 9: Vista lateral derecha.



Figura 10: Representación espacial realizada por el robot tras varios movimientos de rotación. Las imágenes muestran las propiedades de color y profundidad percibidas

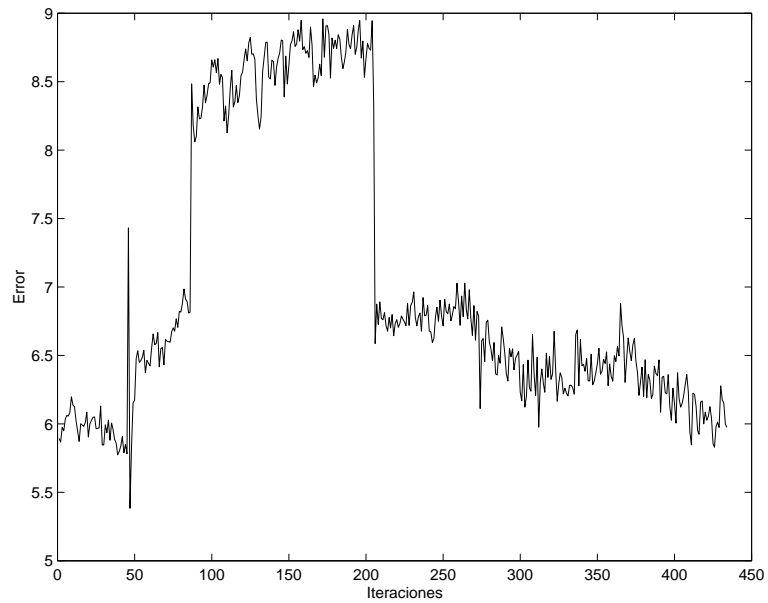


Figura 11: Evolución del error de reproyección del 3D estimado por el robot a lo largo del tiempo. Esta gráfica de error se corresponde con la secuencia de sacádicos de la figura 12, realizando un movimiento sacádico cada 75 iteraciones.

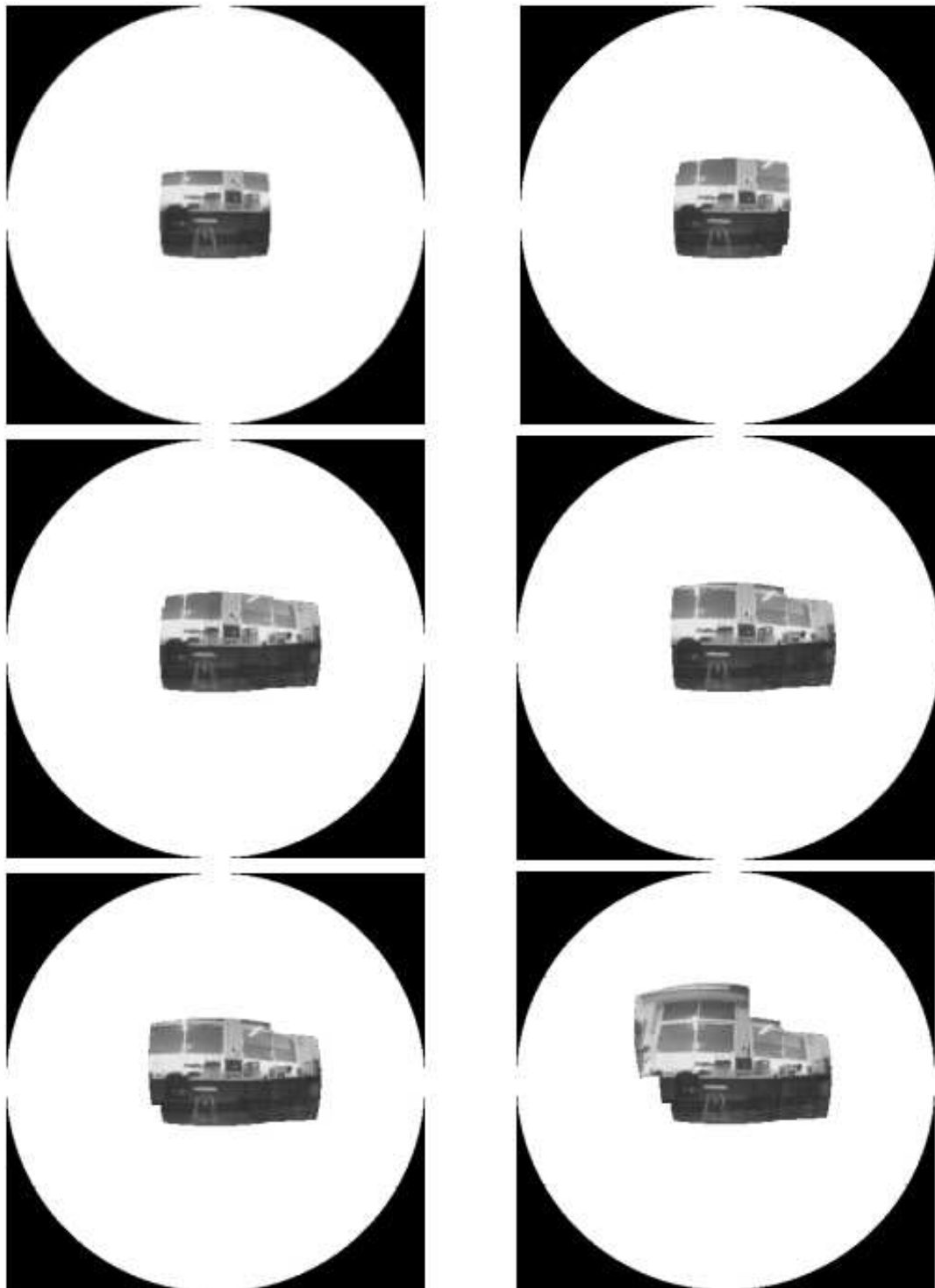


Figura 12: Secuencia de sacádicos generados por el robot hacia puntos del espacio con máximo error de reproyección