Detección probabilística de puertas con visión monocular activa

José María Cañas Plaza
Universidad Rey Juan Carlos

Reid Simmons
Carnegie Mellon University

María C. García-Alegre
Instituto de Automática Industrial (CSIC)

Febrero de 2001

Resumen

Este artículo describe un sistema de percepción activa que integra visión con sonars para detectar la existencia y la posición de puertas en el entorno de un agente móvil. El flujo visual se analiza buscando bordes verticales y esa información en imágenes tomadas desde distintas posiciones se mezcla con los datos sonar en un marco probabilístico. La dinámica de probabilidades que suben y bajan dependiendo de los datos sensoriales acaba confirmando la ubicación real de la puerta, desechando las ficticias, que son debidas a la ambigüedad de profundidad inherente a la visión monocular. El sistema se ha implantado con éxito en un robot de interiores. También se equipa al agente con un control autónomo que gobierna los movimientos del robot para facilitarle activamente la percepción de las puertas. Este control elige los puntos desde los cuales tomar sucesivas observaciones del entorno de manera que se maximice la discriminación entre las puertas candidatas que se tienen en cada momento.

1 Introducción

La incorporación de la visión como sensor a los agentes móviles se ha visto frenada por el alto precio del equipo (cámaras y tarjetas digitalizadoras) y sobre todo por la complejidad inherente en la extracción de información útil desde el flujo de imágenes. Así se han preferido otros sensores más baratos o con una interpretación directa cara a la navegación, como el sonar o el láser.

Si bien han aparecido recientemente enfoques que generan comportamientos desde las imágenes directamente (por ejemplo desde flujo óptico), el uso tradicional de la visión ha sido el de resconstrucción 3D del espacio (pares estéreo). En nuestro caso se utiliza la visión para enriquecer la representación del entorno con un estímulo concreto, las puertas. Este estímulo es muy frecuente en entornos de oficina, laboratorios, etc, por lo que la técnica descrita puede ser de utilidad para muchos investigadores de robótica móvil en interiores.

La motivación inicial para abordar este trabajo nació de la necesidad de que un robot móvil pasara por puertas relativamente estrechas. El robot móvil es cilíndrico y está equipado con diferentes sensores (sonars, encoders y tactiles). Las estrategias de navegación basadas en sonars resultan insuficientes para atravesar puertas estrechas (con las anchas no hay problema) debido a la alta incertidumbre angular asociada a las medidas sonar. Esa falta de resolución lleva a percibir los huecos de una anchura menor a la real, lo cual resulta crítico con puertas que ofrecen poco margen a parte de la anchura del robot. Para solventar esta situación se pensó incorporar visión al sistema perceptivo de modo que se pudiera detectar la anchura real de las puertas y con ello implementar más fiablemente el control para atravesarlas.

Este sistema perceptivo detector de puertas también resulta ventajoso en otras dos situaciones. Primero, como ayuda a la autolocalización inicial del robot, las puertas detectadas en el entorno tras arrancar el robot le ayudan a anclar mejor cual es su posición en cierto mapa global que ya pudiera tener. Segundo, como ajuste fino para llegar a una habitación concreta después de que la navegación global ha llegado aproximadamente al destino y deja al robot en los alrededores de la puerta dando por concluída su tarea (véase [Simmons95]). Estos dos ejemplos son de utilidad directa en Xavier (figura 1), el robot para el cual se programó este sistema perceptivo.

El problema se ha descompuesto en varias partes, que coinciden con las siguientes secciones del artículo. En la primera se extraen las jambas de las puertas que hay en la imagen, siguiendo técnicas de filtrado de bordes y detección de líneas verticales. En una segunda etapa se describe el marco probabilístico que permite

Figura 1: Robot Xavier

integrar información desde distintas imágenes y la fusión con otros sensores como los sonars y tactiles. Después se describe la estrategia que orienta el movimiento autónomo del robot y de la cámara en su búsqueda de las puertas, convirtiendo al sistema en un sistema de percepción activa. Finalmente se esbozan unas conclusiones y las líneas por las que puede avanzar el trabajo.

2 Detección de puertas en la imagen

Para percibir la puerta en profundidad primeramente hay que detectar sus jambas en cada imagen monocular tomada por la cámara. La cámara está montada paralela al suelo y sólo se permiten movimientos horizontales. De este modo las jambas aparecen siempre como líneas verticales en las imágenes de la cámara. El procedimiento seguido para detectarlas sigue tres pasos :

1. Las imágenes de entrada vienen en niveles de gris, no se utiliza información de color. Para minimizar el ruido se les pasa inicialmente por un suavizado que elimina pixels espúreos que sólo perturbarían los cálculos.

2. Se aplica un filtro de bordes siguiendo la técnica y el código de Stephen Smith descrito en [Smith97]. Este filtro clasifica cada pixel de la imagen como punto borde o no.

3. Sobre la imagen de bordes se buscan líneas verticales largas, esto es, columnas que acumulen más de $umbral_1 = 80\%$ de puntos borde. Como la verticalidad estricta es difícil de conseguir en la práctica también se tienen en cuenta las $n = 4$ columnas adyacentes. Todas las columnas que superen ese umbral de puntos borde representan una línea vertical pronunciada, posiblemente la jamba de una puerta.

Este detector funciona en tiempo real sin suponer demasiada sobrecarga al procesador y con él los bordes de las puertas aparecen resaltados. También resalta los límites de armarios o los bordes verticales de pizarras que contrasten con el color de la pared. Sin embargo no hay problema con esos falsos positivos, está previsto desecharlos posteriormente. Lo importante en este primer paso es no quedarse ninguna jamba real sin detectar, puesto que lo que no se vea en esta etapa no existe de ningún modo para las restantes partes del sistema perceptivo, y por ello se pierde.

Para que esta técnica sea válida se necesita que haya contraste entre el marco y la pared, o entre el marco y la hoja de la puerta, que suele ser habitual. Como todos los sensores, tiene asociado un rango de alcance: sólo detecta jambas próximas puesto que las puertas lejanas ocupan pocos pixels y por ello son ignoradas. No obstante en entornos de oficina las distancias no suelen ser muy grandes.

La figura 2 muestra una imagen típica, de 320*240 pixels, obtenida con una cámara de 55° de apertura. En ella se pueden observar los puntos borde superpuestos a la imagen suavizada, y la salida final del detector de jambas, con tres columnas resaltadas, las dos jambas reales y un tercer borde vertical muy pronunciado.

Figura 2: Detección de la puerta en la imagen. Obsérvense los bordes resaltados y el falso positivo al lado de la jamba izquierda.

3 Mezcla probabilística de información desde varias imágenes

El objetivo último del sistema descrito es percibir la existencia de puertas y ubicar perfectamente tanto su posición respecto del robot como su anchura real. Para extraer profundidad usando visión se necesita al menos un par de imágenes desde distintas posiciones para resolver la ambigüedad inherente a la visión monocular. Dicho de otro modo cada una de las imágenes sólo proporciona información parcial acerca de la puerta, información no concluyente. La idea de este trabajo es utilizar el marco probabilístico para integrar esa información de varias imágenes tomadas por el robot desde posiciones distintas, así como de los sensores ultrasónicos, con el objetivo de inferir la existencia y localización de las puertas (figura 3).

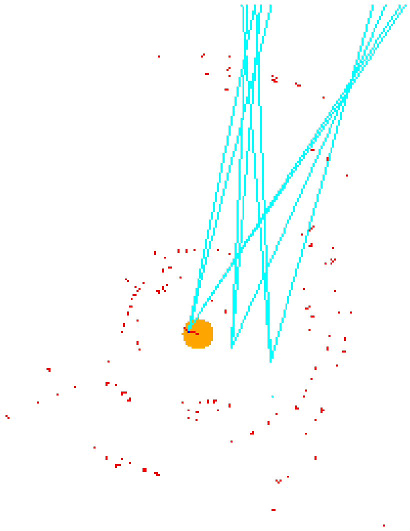

Figura 3: Rayos visuales procedentes de imágenes obtenidas desde 3 posiciones distintas.

Para representar la creencia que tiene el robot sobre la existencia o no de puertas en su entorno, éste mantiene un **grid de probabilidad de jamba**. Este grid se va actualizando a medida que el robot recibe nueva información sensorial, bien desde la cámara, bien desde los sonars y es la base sobre la que se decide interpretar lo visto como puerta o no.

Es un grid bidimensional, basado en la representación propuesta por Elfes [Elfes90] y situado en un plano paralelo al suelo a la altura de los sonars ($\approx 1m$). Cada celda almacena la probabilidad de que en esa posición haya una jamba de puerta, y en los experimentos concretos se eligió un tamaño de $10cm * 10cm$. Esta representación horizontal resulta suficiente para representar las puertas, su orientación, posición y anchura. Las jambas reales son verticales y en este plano aparecen como puntos. Con un par de ellos se representa una puerta. Puesto que esta representación esta dirigida exclusivamente a la identificación del estímulo puerta no aparecen en ella objetos por debajo de la altura de los sonars como las papeleras.

Inicialmente todas las casillas del grid tienen valor 0.5, reflejando el desconocimiento total. La probabilidad de cada una de las casillas del grid se va actualizando siguiendo la *regla de Bayes* a medida que se incorpora información. En cada momento t la estimación que se tiene de la probabilidad de jamba en el punto (x, y), $p_{jamba}(x, y, t)$, se define como la probabilidad de jamba en ese punto condicionada a las observaciones que se han obtenido hasta ese momento. Así lo muestra la ecuación 1, donde $data(t-1)$ supone el conjunto de observaciones acumuladas hasta el instante $t-1$ e $imagen(t)$ la observación actual.

$$\begin{aligned} p_{jamba}(x,y,t) &= p(jamba/imagen(t), data(t-1)...) && (1) \\ &= \frac{p(jamba/data(t-1)) * p(imagen(t)/jamba, data(t-1))}{p(imagen(t)/data(t-1))} && (2) \end{aligned}$$

Aplicando a (1) la regla de Bayes obtenemos (2). Si asumimos que las observaciones son *independientes en sentido markoviano* (3) y utilizamos nuevamente la regla de Bayes (4) llegamos a la expresión (5). La independencia markoviana supone que dado el hecho de la existencia de la jamba, la $imagen(t)$ no depende de las observaciones anteriores $data(t-1)$, según expresa (3). En general el mundo no es markoviano, pero en la práctica asumimos esa independencia entre observaciones incorporando al grid exclusivamente las imágenes tomadas desde posiciones relativamente separadas. Esta suposición facilita enormemente el tratamiento probabilístico de la información y posibilita el desarrollo teórico siguiente, paralelo al de Margaritis [Margaritis98].

$$\begin{aligned} p(imagen(t)/jamba, data(t-1)) &= p(imagen(t)/jamba) && (3) \\ p(imagen(t)/jamba) &= \frac{p(jamba/imagen(t)) * p(imagen(t))}{p(jamba)} && (4) \\ p_{jamba}(x,y,t) &= \frac{p(jamba/data(t-1)) * p(jamba/imagen(t)) * p(imagen(t))}{p(jamba) * p(imagen(t)/data(t-1))} && (5) \end{aligned}$$

Normalmente en lugar de manejar directamente la probabilidad de jamba p_{jamba} se maneja el *ratio de probabilidad*, definido como $\rho_{jamba} = p_{jamba}/p_{\overline{jamba}}$. Con esta simplificación se elimina la dependencia de $p(imagen(t))$ y de $p(imagen(t)/data(t-1))$ sin perder información, puesto que afectan por igual a p_{jamba} y a $p_{\overline{jamba}}$, según muestra (8).

$$\begin{aligned} \rho_{jamba} &= p_{jamba}/(1-p_{jamba}) && (6) \\ p_{jamba} &= \rho_{jamba}/(1+\rho_{jamba}) && (7) \\ p_{\overline{jamba}}(x,y,t) &= \frac{p(\overline{jamba}/data(t-1)) * p(\overline{jamba}/imagen(t)) * p(imagen(t))}{p(\overline{jamba}) * p(imagen(t)/data(t-1))} && (8) \end{aligned}$$

Sustituyendo (5) y (8) en (6) llegamos a (9), que posibilita un tratamiento incremental de la información. Con cada nueva observación la probabilidad acumulada se multiplica por un factor $\frac{p(jamba/imagen(t))}{1-p(jamba/imagen(t))}$ que es precisamente el valor dado por el modelo de sensor, y una constante $\frac{1-p_{jamba}}{p_{jamba}}$ determinada por la probabilidad a priori de jamba, sin ningún otro conocimiento. Trabajando con logaritmos los productos se convierten en rápidas sumas. Si en algún momento se quiere obtener la probabilidad basta con deshacer los logaritmos y aplicar (7).

$$\rho_{jamba}(x,y,t) = \frac{p(jamba/imagen(t))}{1-p(jamba/imagen(t))} * \frac{1-p_{jamba}}{p_{jamba}} * \rho_{jamba}(x,y,t-1) \quad (9)$$

Por lo tanto, si una imagen proporciona información sobre el estado de determinada celdilla (x, y), es el valor del modelo de sensor $p(jamba/imagen(t))$ en esa posición el que determina si allí la probabilidad de jamba sube o baja después de la nueva observación. La información proveniente de las lecturas sonar se trata del mismo modo, pero con otro modelo de sensor distinto. En el siguiente apartado se describen con detalle ambos modelos.

3.1 Modelo de visión

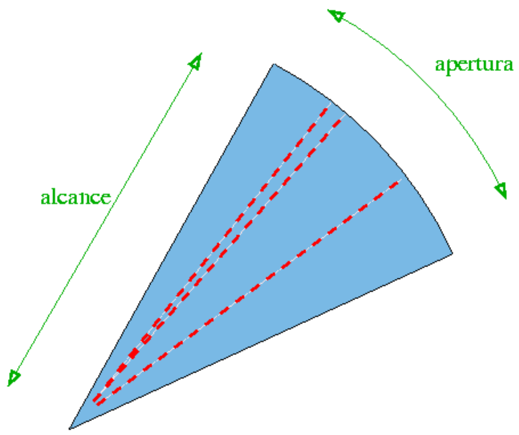

Figura 4: Ejemplo de cómo actualiza una imagen el grid con probabilidad de jamba.

Cada vez que se recibe una imagen *independiente* se actualiza la probabilidad asociada a las celdillas del grid que caen dentro del campo visual. El campo visual viene determinado por un *alcance* y una *apertura* angular. Si en la imagen se observa un borde vertical pronunciado, es indicio de que hay realmente una jamba *en algún lugar* del plano perpendicular al suelo que proyecta en los pixels de ese borde. En el grid horizontal ese plano proyectante se convierte en una línea, y en cualquier punto de esa línea puede estar físicamente la jamba. La probabilidad de jamba asociada a esas celdillas debe aumentar.

Por el contrario si en la imagen no se aprecia ningún borde vertical pronunciado eso es síntoma de que en el campo de visión abarcado por la cámara no hay ninguna puerta y la estimación de probabilidad de jamba debe bajar.

La interpretación de las imágenes sigue el modelo de la figura 4, que corresponde a la proyección en profundidad de la figura 2. Siguiendo este modelo se aumenta la probabilidad de jamba en las celdillas sobre las líneas discontinuas porque todas ellas proyectan en pixels donde se aprecia una jamba en la imagen. También se baja la probabilidad en las celdillas oscuras que caen dentro del campo visual, porque en su proyección no se ha apreciado un borde vertical pronunciado. Empíricamente ajustamos los valores concretos a $p(jamba_{x,y}/imagen(t)) = 0.72$ si proyecta en una jamba de imagen o $p(jamba_{x,y}/imagen(t)) = 0.42$ si no.

En otras palabras, se aumenta la probabilidad de jamba en todas aquellas posiciones que podrían ser la causa de que se haya apreciado una jamba en la imagen y se disminuye en el resto. La idea subyacente es que después de unas cuantas observaciones sólo la posición real de la jamba recibirá la realimentación positiva de la mayoría de las imágenes. El resto de localizaciones se verán desmentidas con alguna observación y su probabilidad de jamba no podrá ser tan alta como en la posición auténtica. Los valores concretos del modelo suponen un compromiso entre el peso de cada observación individual y la tolerancia a algún error en la percepción.

3.2 Modelo de sonars

Se puede suponer que un sonar mide la distancia al obstáculo más cercano en la dirección perpendicular al sensor. Por ejemplo si un sonar detecta el primer obstáculo a 3 metros eso quiere decir que entre el sensor y esos 3 metros hay espacio vacío y desde el punto de vista de la puerta, difícilmente hay una jamba en ese área.

El modelo de sonar utilizado consiste en un cono que se abre desde la posición del sensor hasta el radio medido, con el eje en la perpendicular del sensor y siguiendo una apertura que refleja precisamente la de la onda ultrasónica. Todas las celdas del grid que caigan dentro de ese cono disminuyen su probabilidad de ser jamba, puesto que el sonar ha sensado espacio vacío. De este modo los sonars se utilizan para descartar hipotéticas localizaciones de las jambas.

3.3 Probabilidad de puerta

El grid de probabilidad de jamba tiene la información sensorial necesaria para determinar si el robot ha visto una jamba real o no. En nuestro sistema una puerta viene definida por un par sus extremos, $punto1$ y $punto2$, que son precisamente sus jambas. A cada posible puerta le asociamos una **probabilidad de puerta** dada por la fórmula:

$$p_{puerta}(punto1, punto2, t) = p_{jamba}(x1, y1, t) * p_{jamba}(x2, y2, t) * condiciones\_de\_puerta \tag{10}$$

Es decir el producto de la probabilidad de cada una de sus jambas por un factor que indica en qué medida esa posible puerta satisface las condiciones de puerta. La condición principal evalúa si su anchura casa con la esperada en las puertas del entorno. En la implementación realizada ese factor es binario: 1 si la distancia entre jambas caen entre un mínimo y un máximo parametrizables, y 0 fuera de ese intervalo. También se anula si las dos jambas están contenidas en el mismo rayo óptico, ya que en ese caso no puede tratarse de una puerta.

Para examinar la existencia o no de puertas primero se explora el grid de probabilidad de jamba, y sólamente teniendo en cuenta aquellas celdillas más probables, se elabora un **conjunto de puertas candidatas**, formado por todos los posibles pares. Seguidamente se evalúa la probabilidad de cada una de estas puertas candidatas y el robot asume que ha detectado realmente una puerta cuando su probabilidad está por encima de cierto *umbral de aceptación*. Es decir cuando sus jambas están confirmadas por las observaciones sensoriales y la distancia entre ellas es la adecuada.

3.4 Experimentación

El esquema probabilístico propuesto para integrar información desde varias imágenes y sensores sonar se ha implementado con éxito en el robot real Xavier (figura 1) de la Universidad Carnegie Mellon operando en una planta llena de despachos y laboratorios.

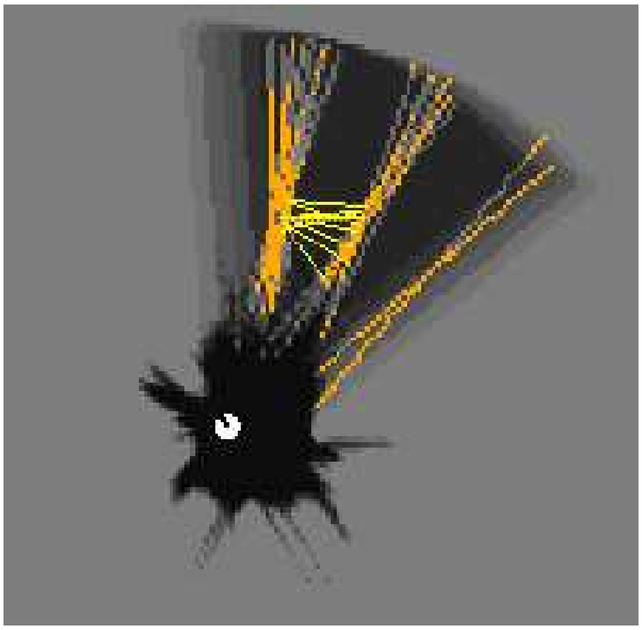

Figura 5: Mezcla en un grid probabilístico de información obtenida desde varias posiciones. Obsérvese el espacio entre rayos con probabilidad menor que 0.5, si no se ha visto jamba se disminuye su probabilidad.

En la figura 5 se puede observar el estado del grid de probabilidad de jamba después de la observación del entorno desde cinco posiciones distintas, separadas entre sí unos 30cm. Alrededor del robot la probabilidad de jamba es muy baja (color oscuro), gracias a los sonars que han ayudado a descartar las hipótesis en ese área. En la parte superior se aprecian tres haces de alta probabilidad correspondientes a los distintos rayos visuales de los bordes verticales detectados en las sucesivas imágenes. Las zonas oscuras intermedias corresponden a la proyección de la zonas de imagen con ausencia de bordes verticales.

Las partes intermedias de los dos haces centrales son las que reciben más probabilidad porque es donde intersectan los "rayos de los haces". En color claro, más o menos horizontales entre esos dos haces aparece superpuesto el conjunto de puertas candidatas obtenido hasta ese momento. Obsérvese que no hay ninguna puerta candidata con un extremo en el haz de rayos de la derecha, porque no han acumulado suficiente probabilidad. La longitud de todas las puertas candidatas generadas está dentro de los márgenes aceptables.

En estas pruebas el movimiento del robot y la orientación de la cámara venían teleoperados por el humano. Por lo tanto no se evaluaba la autonomía del agente percibiendo puertas, sino que la aproximación descrita fuera capaz de concluir la existencia y ubicación de la puerta desde unas observaciones razonables obtenidas desde posiciones distintas. En la siguiente sección describimos el enfoque seguido para dotar de autonomía agente y eliminar la participación humana en todo el proceso.

4 Percepción activa: movimiento para percibir la puerta

Debido al campo visual limitado, cobra importancia el lugar desde donde se observa el entorno. Dependiendo de ese lugar se percibirán unos estímulos y se ignorarán otros. En cierto modo se puede interpretar la propia posición del robot y la orientación de su cámara como el foco sobre el cual tiene puesta su atención sensorial. Ese foco se deberá desplazar a lo largo del espacio para percibir todo el entorno y encontrar las puertas que en él pueda haber.

Las partes descritas hasta el momento son pasivas, en el sentido de que procesan lo que caiga delante de la cámara. Ahora describiremos la estrategia que controla los motores y guía su movimiento para facilitar al agente la percepción de la puerta. Con ella se completa el sistema de percepción *activa* que dota al robot de la capacidad de detectar las puertas de modo completamente autónomo, sin intervención humana alguna.

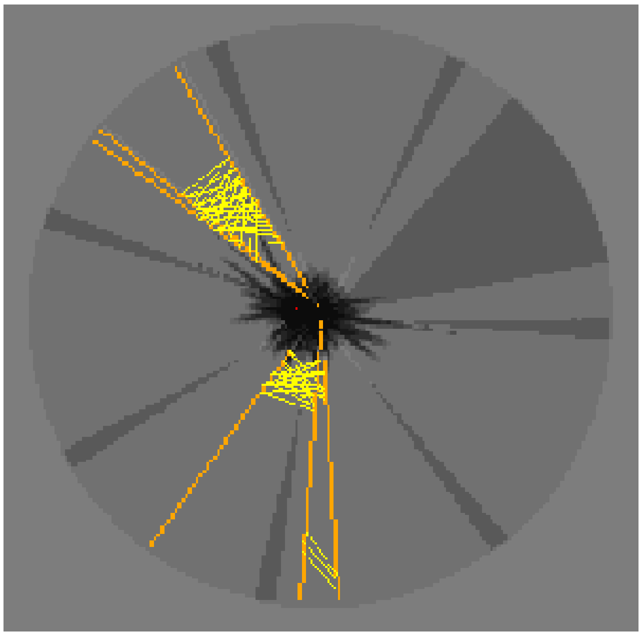

Figura 6: Aspecto del grid después de la vuelta inicial, hay un conjunto de puertas candidatas. Entre ellas hay que discernir la puerta real con nuevas observaciones.

Inicialmente el robot da una vuelta completa sobre sí mismo mirando en todas direcciones, buscando todas las posibles jambas que pueda haber en sus cercanías. Después de ese primer vistazo el grid de probabilidad de jamba tiene un aspecto como el mostrado en la figura 6, con algunas jambas candidatas.

A partir de ese momento la estrategia de movimiento autónomo es iterativa: se calcula el punto óptimo para tomar la siguiente imagen y se desplaza el robot hasta él mediante navegación local. Una vez allí se vuelve a analizar la imagen de la cámara y se concluye si se ha detectado alguna puerta, si hace falta seguir explorando o si es mejor abandonar la búsqueda. En caso de necesitar más evidencia sensorial se calcula el nuevo punto óptimo y así sucesivamente.

4.1 Cálculo del punto óptimo

Después del primer vistazo, se tiene ya una cierta información sobre dónde pueden estar las puertas, en caso de existir alguna en el entorno. Las jambas que se hayan apreciado en la imagen se proyectan en el grid, elevando la probabilidad de todas las posiciones que explican esas observaciones. Debido a la ambigüedad de profundidad tanto la posición real como otras hipotéticas resultan resaltadas.

Desde el grid de probabilidad de jamba se extrae el *conjunto de puertas candidatas* formando pares con las celdillas de mayor probabilidad de jamba según se describió en la sección anterior. Para cada uno de esos pares se evalúa su probabilidad de puerta y el conjunto global queda ordenado de mayor a menor probabilidad. Tras la vuelta inicial ningún par habrá conseguido acumular evidencia suficiente como para sobrepasar el *umbral de aceptación* y dar la búsqueda por concluída, sin embargo ese conjunto representa las mejores hipótesis que el robot tiene para explicar las imágenes que ha visto. Previsiblemente la ubicación real estará contenida en ellas y será conveniente dirigir las nuevas observaciones para que confirmen o desmientan estas candidaturas.

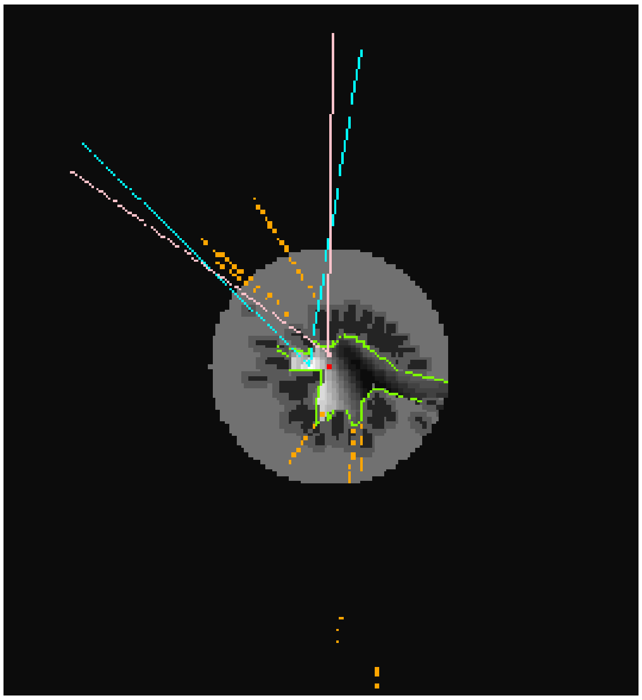

Figura 7: Cálculo del punto desde el cual tomar la siguiente imagen. Para cada celdilla dentro de la zona aceptable se representa la utilidad de la mejor orientación de la cámara.

El lugar y orientación de la cámara desde el cual tomar la siguiente imagen del entorno se calculan maximizando una *función de utilidad*, definida en la ecuación (11). Por un lado cuantas más jambas candidatas se observen desde una posición su utilidad asociada es mayor, porque se pueden constrastar más candidaturas. En este sentido el sumatorio de (11) se extiende a todas las jambas candidatas que caigan dentro del campo visual desde (x, y, θ). Por otro lado la utilidad es mayor cuanto menor es el solapamiento en la imagen entre las candidatas que se observen. Por ejemplo si desde una posición tres jambas candidatas aparecen en los mismos pixels entonces desde esa posición no se puede discriminar cual de ellas es la verdadera. Ese punto de observación es menos recomendable que otro desde el cual esas mismas candidatas aparecen en pixels distintos y por lo tanto se puede comprobar mejor la validez de cada una de ellas. En este sentido el sustraendo de (12) resta el solapamiento medio a la información que aporta esa jamba candidata observada desde ese punto. $solapamiento(jamba_i, jamba_j)$ es una función triangular que se hace 1 si ambas jambas candidatas proyectan en el mismo pixel y disminuye hasta 0 a medida que crece la distancia entre esos pixels de proyección.

$$utilidad(x,y,\theta) = \sum_{jamba_i \in campo\_visual} informacion(jamba_i) \quad (11)$$

$$informacion(jamba_i) = 1 - \frac{\sum_{i \neq j} solapamiento(jamba_i, jamba_j)}{\#jambas \in campo\_visual} \quad (12)$$

Al maximizar la utilidad (11) se encuentra tanto la mejor posición (x, y) como la orientación apropiada θ de la cámara para tomar la siguiente imagen. Para agilizar estos cálculos la función de utilidad sólo se calcula en posiciones aceptables: que no estén muy alejadas de la posición actual del robot (a menos de 3 m), que no estén cerca de ningún obstáculo y que sea directamente accesible con navegación local (un giro y una traslación). Estas condiciones reducen considerablemente el espacio de búsqueda.

En la figura 7 se puede apreciar la función de utilidad evaluada en las cercanías del robot después de la vuelta inicial reflejada en la figura 6. El punto central es el robot, y en él tiene el vértice el sector que representa la orientación actual de la cámara. Desde el punto central se abre también un contorno de celdillas directamente accesibles con navegación local sin chocar con los obstáculos, que son las manchas oscuras. Dentro de ese contorno se evalúa la función de utilidad, generando en ese caso un patrón que favorece a los puntos situados a la izquierda. Se hace máxima en el vértice del segundo sector, que es el punto y orientación óptimos recomendados. Los puntos claros sueltos son las jambas candidatas que necesitan confirmación.

4.2 Navegación local

Una vez calculado el punto óptimo para situar la cámara se hace uso de arquitectura de control desarrollada para Xavier [Simmons94] encargando a la capa de navegación local la tarea de llevar al agente hasta aquella posición. No es necesario acudir a técnicas de navegación más compleja porque en el cálculo ya exigimos al punto destino que fuera alcanzable con un giro y una traslación.

Los navegadores reactivos toman control del robot hasta que finalizan su tarea. Entonces el sistema perceptivo vuelve a tomar y analizar una imagen, actualizando el grid de probabilidad de jamba y determinando si ha percibido alguna puerta o aún se necesita acumular más observaciones. Aunque la arquitectura lo permite se ha preferido no procesar imágenes mientras el agente se está moviendo.

5 Discusión

El sistema perceptivo presentado combina información visual, obtenida desde distintas imágenes, con datos ultrasónicos en un marco probabilístico para detectar la existencia y la posición de las puertas en el entorno. El enfoque permite integrar medidas que sólo aportan información parcial concluyendo la presencia o no de puertas en el entorno del agente móvil.

Cada vez que se obtienen nuevas imágenes y lecturas sonar se proyectan en un grid bidimensional siguiendo sendos modelos sensoriales. El grid contiene la probabilidad de que realmente exista una jamba en cada una de sus celdillas. A medida que se reciben nuevas observaciones debido a la dinámica de probabilidades se van descartando ubicaciones ficticias de las jambas y sólo las posiciones verdaderas acumulan suficiente evidencia sensorial. Emparejando distintas celdillas se obtiene el conjunto de puertas candidatas y su probabilidad asociada. El agente asume que ha visto una puerta cuando alguna de las candidatas sobrepasa el umbral de aceptación.

El sistema se ha materializado con éxito en un robot de interiores y se le ha dotado de un control autónomo que dirige los movimientos del agente y de su cámara para facilitar la percepción. El robot se mueve siempre hacia el punto que más discrimina entre la puertas candidatas existentes en cada momento, iterando hasta que finaliza la búsqueda. No obstante esta parte activa necesita aún más experimentación para conseguir más robustez en los resultados.

5.1 Trabajos relacionados

Además del trabajo realizado con pares estéreo la idea de combinar varias imágenes para inferir características 3D ha sido explorada profusamente. Por ejemplo Margaritis [Margaritis98] utiliza también la probabilidad como marco para localizar objetos en 3D a partir de visión. Su desarrollo teórico es más completo que el expuesto aquí: es realmente tridimensional e incluye modelos probabilísticos para reflejar el error inducido por el movimiento no determinista del robot. Sin embargo no aborda el problema de cómo dirigir ese movimiento para percibir mejor el estímulo buscado.

Un enfoque similar es el de Collins [Collins96][Collins97] que extrae estructura 3D del entorno, por ejemplo los edificios de un campus universitario, desde varias imágenes aéreas. Como primitiva visual también utiliza bordes y los retroproyecta en el espacio 3D desde todas las imágenes disponibles. Las celdillas con los bordes reales quedan resaltadas porque recibirán más retroproyecciones que el resto. Una distribución estadística determina si el número de retroproyecciones que caen en una celdilla se debe al azar o a la existencia de un borde 3D real.

5.2 Ventajas e inconvenientes

El marco probabilístico utilizado ofrece muchas ventajas. La primera es que extrae la posición de la puerta sin necesidad de buscar homólogos. Para extraer mapas de profundidad, los pares estéreo emparejan puntos homólogos entre la imágenes derecha e izquierda y aplican triangulación. El problema de asignación de homólogos puede ser intratable si las imágenes han sido tomadas desde posiciones completamente distintas. Esa es precisamente nuestra situación con el robot, que puede observar la puerta desde lugares distantes más de un metro. En lugar de resolver explícitamente esta asignación de homólogos el grid de probabilidad permite representar la ambigüedad de cada imagen y solucionar el problema de modo implícito.

El enfoque probabilístico resulta válido para cualquier estímulo, no sólo puertas. Únicamente se necesitan los modelos sensoriales adecuados que traduzcan a probabilidad la información proporcionada por los sensores sobre el estímulo en cuestión.

Otra ventaja de este marco es que permite fusión sensorial de modo sencillo. Para ello se utilizan modelos distintos adecuados a los sensores diferentes pero que afectan a la misma probabilidad. También admite la incorporación de conocimiento a priori como valor inicial de la probabilidad de jamba. Por ejemplo si se sabe que la puerta estás más o menos por un área, esas celdillas tendrán un valor inicial mayor que el resto.

Esta aproximación facilita la tolerancia al ruido, a los falsos positivos y a fallos de observación. Por ejemplo si temporalmente deja de observarse una jamba por la presencia espúrea de un obstáculo interpuesto, el sistema funcionará bien si obtiene nuevas imágenes buenas con las que compensar ese fallo, la probabilidad acabará subiendo. Para materializar esa robustez los modelos de sensor deben diseñarse cuidadosamente.

Uno de los puntos débiles de este enfoque es el ruido en los encoders. Para poder mezclar distintas imágenes su información se ancla espacialmente a la posición del robot cuando tomó cada una de ellas. Si la estimación de posición es errónea la mezcla de probabilidades no es válida. Hemos utilizado encoders para estimar la posición y es de sobra conocido en la literatura el error acumulativo que padecen. No obstante las posiciones del robot sólo necesitan ser relativamente válidas, es decir precisas respecto de la posición actual del agente, no en términos absolutos. Ese tipo de precisión se tiene con los encoders si sólo se tienen en cuenta medidas recientes, dentro de una ventana temporal en la que no da tiempo a acumular demasiado error relativo de posición. Por otro lado el propio grid asume bien errores de localización pequeños.

Otra desventaja es el actual tamaño de las celdillas, 10 cm, que pone en duda la precisión obtenible con este sistema. Sin embargo el enfoque se muestra eficiente para comprobar la propia existencia de puertas, y su posición dentro de la resolución del grid. Si se quiere más precisión se puede complementar con otros algoritmos, que resultan mucho más sencillos una vez que la existencia del estímulo puerta ha sido verificada.

5.3 Líneas futuras

Está previsto continuar el trabajo mejorando la parte activa, que no está suficientemente contrastada. En concreto la extracción en cada momento de las puertas candidatas es clave puesto que ellas determinan implícitamente el siguiente punto de observación. Umbrales fijos para esa extracción resultan inadecuados porque al principio la probabilidad acumulada es baja pero va creciendo con nuevas observaciones. El esquema de selección debe mantener en todo caso cierta discriminación sin excluir buenas candidatas fuera. Se está pensando en umbrales dinámicos que tengan en cuenta esa acumulación.

También se tiene pensado portar el sistema al robot B21-Hermes del Instituto de Automática Industrial e implementar el comportamiento *atraviesa-puerta* que motivó inicialmente este sistema y que debe sacar beneficio de todo el trabajo perceptivo realizado.

Agradecimientos

Este trabajo ha sido promocionado por la Comunidad Autónoma de Madrid a través de su programa de becas de Formación de Personal Investigador y realizado en los laboratorios del Computer Science Department de Carnegie Mellon University.

Referencias

[Elfes90] Alberto Elfes, Occupancy grids: a stochastic spatial representation for active robot perception. En *actas de la Sixth AAAI Conference on Uncertainty in AI*, Julio 1990.

[Simmons94] Reid Simmons, Structured Control for Autonomous Robots. En *IEEE Transactions on Robotics and Automation*, 10:1, páginas 34-43, Febrero 1994.

[Simmons95] R. Simmons y S. Koening, Probabilistic navigation in partially observable environments. En *actas de International Joint Conference on Artificial Intelligence*, Montreal, Julio 1995.

[Simmons97] R. Simmons, R. Goodwin, K. Zita Haigh, S. Koening y J. O'Sullivan, A layered architeture for office delivery robots. En *actas de la ACM International Conference on Autonomous Agents*, páginas 245-252, Marina del Rey (USA), Febrero 1997.

[Collins96] R. Collins, A space-sweep approach to true multi-image matching. En *actas de la IEEE Conference on Computer Vision and Pattern Recognition*, páginas 358-363, San Francisco, Junio 1996.

[Collins97] R. Collins, Multi-image focus of attention for rapid site model construction. En *actas de la IEEE Conference on Computer Vision and Pattern Recognition*, San Juan (Puerto Rico), Junio 1997.

[Margaritis98] D. Margaritis y S. Thrun, Learning to locate an object in 3D space from a sequence of images. En *actas de la International Conference on Machine Learning*, páginas 332-340, 1998.

[Smith97] S. Smith y J.M. Brady, SUSAN-A new approach to low level image processing. En *International Journal of Computer Vision*, 23(1):45-78, Mayo 1997.