

# DETECCIÓN DE SUELO Y PAREDES CON VISIÓN MONOCULAR PARA NAVEGACIÓN POR INTERIORES

J. Omedes, G. López-Nicolás y J. J. Guerrero  
Instituto de Investigación en Ingeniería de Aragón. Universidad de Zaragoza, España  
jason.omedes@gmail.com, gonlopez@unizar.es, jguerrer@unizar.es

## Resumen

*En este trabajo se estudia el problema de detectar el suelo y las paredes verticales de una escena a partir de una única imagen tomada en el interior de un edificio. Aunque existen métodos equivalentes que utilizan imágenes convencionales, las imágenes omnidireccionales resultan particularmente útiles para esta tarea debido a su amplio campo de vista. No obstante, debido a la mayor complejidad geométrica de las imágenes omnidireccionales es necesario el diseño de algoritmos específicos.*

*Presentamos un nuevo método para la estimación de puntos de fuga (VPs) y la clasificación de las líneas extraídas sobre imágenes catadióptricas. La principal contribución es un nuevo método para obtener la estructura de la escena basado en las líneas y puntos de fuga extraídos que, combinados con un conjunto de restricciones geométricas, nos permiten segmentar en la imagen las regiones que forman parte del suelo y las paredes. Este método comienza haciendo la hipótesis de que la habitación está compuesta por 4 paredes, para posteriormente buscar condiciones que permitan expandir esta primera hipótesis con el objetivo de encontrar el contorno en el suelo que mejor encaja con la escena real y posteriormente se mejora la estimación mediante el uso de homografías que permiten propagar las hipótesis resultantes secuencialmente eliminando posibles errores en la clasificación. Se muestra cómo nuestro algoritmo es capaz de encontrar los contornos que delimitan el suelo y las paredes en el interior de un edificio, incluso cuando las paredes están parcialmente ocultas por objetos.*

**Palabras clave:** Vision Computador, visión omnidireccional, direcciones principales, segmentación de regiones.

## 1 Introducción

La obtención de la distribución estructural de la escena a partir de una imagen es una tarea sencilla para cualquier persona, sin embargo, no es tan fácil para los ordenadores. Al mismo tiempo, es una herramienta poderosa, ya que conocer los límites entre suelo y paredes proporciona información valiosa en tareas de

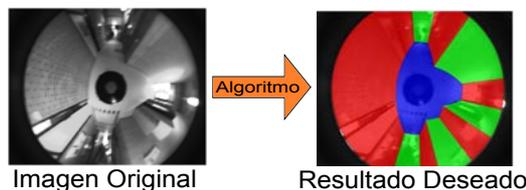


Figura 1: Ejemplo de una imagen tomada con un sistema hyper catadióptrico y el resultado deseado después de aplicar el algoritmo. En el resultado el color azul representa el suelo, el color rojo representa paredes paralelas en una dirección dominante, y el color verde paredes paralelas en una dirección dominante ortogonal a la anterior.

navegación, movimiento, detección de obstáculos o reconstrucción 3D.

Este problema ha sido estudiado en distintas ocasiones y continúa atrayendo los esfuerzos de muchos investigadores que tratan de implementar cada vez mejores algoritmos. La mayoría de las contribuciones funcionan bajo la hipótesis de un mundo tipo Manhattan [5], que asume que la escena tiene 3 direcciones principales ortogonales entre sí. Los entornos de interior normalmente satisfacen esta condición por lo que se entiende la aplicación extensiva de esta hipótesis. Algunos ejemplos son [10], que utiliza líneas extraídas a partir de cámaras perspectivas y a las que aplica condiciones geométricas para buscar el mejor ajuste al entorno, o [9] que representando las habitaciones como cubos en 3D intenta reconocer los límites entre pared y suelo en habitaciones llenas de objetos. También se encuentran otros trabajos como [15] que usa filtros bayesianos sobre un conjunto de hipótesis pared-suelo sin la asunción de un mundo tipo Manhattan, en este caso sigue habiendo 3 direcciones principales pero no se les impone la condición de ortogonalidad.

En los últimos años, la investigación en visión omnidireccional está tomando más importancia debido a su amplio campo de vista, lo que ayuda a una mejor detección de los VPs y permite observar una mayor longitud de las rectas de la escena. Por otro lado, en los sistemas catadióptricos centrales, las líneas rectas de la escena se proyectan como cónicas incrementando la complejidad geométrica, de forma que la mayoría de los algoritmos existentes para cámaras convencionales no puedan aplicarse. Por ello es necesario el desarrollo de nuevos métodos que tengan en cuenta las car-

acterísticas de este nuevo tipo de imágenes.

Partiendo del trabajo [13], proponemos un nuevo método más robusto, ya que no depende de encontrar esquinas (que son difíciles de detectar), y mucho más rápido, debido a que los métodos de clasificación o combinación de esquinas para formar la hipótesis de suelo consumían mucho tiempo.

Nuestra propuesta para la estimación estructural de imágenes utiliza de partida una única imagen omnidireccional, de la que se extraen líneas características, y se clasifican en tres direcciones principales dependiendo de su orientación. Estas líneas contienen información redundante por lo que seleccionamos únicamente un conjunto de puntos que posteriormente permiten generar una primera hipótesis de la forma de la escena considerando que ésta tiene 4 paredes, para después expandir (o no) esta habitación hipótesis de acuerdo a la distribución de los datos. Finalmente, se aplica un proceso secuencial basado en homografías de forma que las distintas hipótesis de imágenes consecutivas compartan información entre sí consiguiendo resultados más homogéneos y precisos. Este trabajo ha sido parcialmente publicado en [12].

## 2 Estimación de puntos de fuga a través de extracción de líneas

El primer paso de nuestra propuesta comienza con la extracción de líneas de la imagen. En lo relativo a extracción de líneas para sistemas catadióptricos dos métodos son [2] [3]. Ambos comienzan usando el detector de líneas Canny y uniendo cadenas de píxeles. La principal diferencia es que [3] trabaja sobre la imagen catadióptrica, donde las líneas y los VPs se extraen mediante RANSAC. Mientras que [2] hace uso del modelo de la esfera unitaria propuesta en [6] donde puntos de la imagen  $\mathbf{p}_I = (X_o, Y_o, 1)$  se proyectan como  $\mathbf{p}_S = (X_S, Y_S, Z_S)$ . Haciendo esto, cada cadena de píxeles que formaba una línea en la imagen define un *gran círculo* en la esfera que se puede representar por su vector normal  $\mathbf{n} = (n_x, n_y, n_z)$ .

En nuestro trabajo usaremos la toolbox de Matlab de Bazin [2], pero adaptando las ecuaciones de sistema para-catadióptrico ( $\xi = 1$ ) a hyper-catadióptrico ( $0 < \xi < 1$ ) para generalizar el método. A partir de este punto, [2] propone probar cada posible combinación entre parejas de vectores normales para identificar las direcciones principales (1 vertical, 2 horizontales) al mismo tiempo que sus respectivos puntos de fuga. Sin embargo, este método consume mucho tiempo y algunas veces produce clasificaciones erróneas.

Nosotros proponemos un nuevo método robusto y rápido que clasifica las líneas paralelas de acuerdo a 3 direcciones principales a partir de dos consideraciones: a) Suposición de mundo de Manhattan [5] que

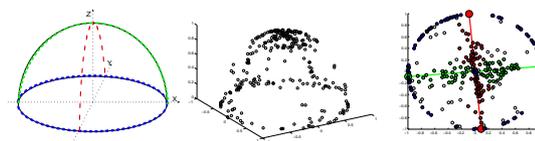


Figura 2: En el modelo de la esfera, se proyecta en un gran círculo en la esfera, que a su vez es representado por su vector normal. La figura representa a la esfera, cada punto corresponde a un vector normal. (Código de Colores: X=Rojo, Y=Verde, Z=Azul). De izquierda a derecha: Esfera con datos ideales; Esfera con datos de una imagen real; clasificación de los datos de la imagen central usando nuestro algoritmo (esfera vista desde arriba). Los puntos grandes representan VPs.

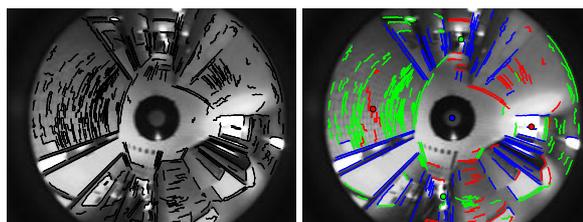


Figura 3: Izda: Líneas extraídas con el detector Canny y píxeles ya encadenados. Dcha: Mismas líneas agrupadas en 3 direcciones dominantes según nuestro algoritmo de clasificación. Los puntos gordos representan VPs.

asume que el mundo está construido en una cuadrícula cartesiana, y b) el eje  $Z$  de la cámara está alineado con el eje  $Z$  de referencia absoluto y, debido a que los sistemas catadióptricos se usan principalmente en robots con ruedas, se puede asumir movimiento plano y verticalidad en los ejes. Es fácilmente demostrable que bajo estas asunciones y con datos perfectos los vectores normales  $\mathbf{n} = (n_x, n_y, n_z)$  correspondientes a las tres direcciones principales  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$  no contienen su propia componente (i.e. líneas que pertenecen a la dirección  $\mathbf{X}$  en el mundo real tienen vectores normales  $\mathbf{n}$  cuya componente  $n_x = 0$ ), Fig. 2. Un problema es que habitualmente los datos no son perfectos por lo que existen desviaciones con respecto a la configuración ideal. El proceso de clasificación puede ser consultado en [12].

## 3 Generación de hipótesis jerárquica

Debido al ruido y las imperfecciones de las imágenes reales, habitualmente es difícil definir claramente donde está el borde entre pared y suelo, por lo que con la información extraída y un conjunto de condiciones geométricas debemos buscar dónde se encuentran estos límites. Para hacer esto vamos a generar cónicas (representan los posibles bordes) a partir de un conjunto de puntos pertenecientes a las rectas anteriormente clasificadas.

### 3.1 Selección de puntos clave

El trabajo propuesto en [13] presta especial atención a encontrar esquinas que delimiten la zona del suelo de la imagen estudiada. Estas esquinas vienen definidas por puntos donde intersectan segmentos verticales/horizontales entre sí, las cuales son difíciles de detectar en muchos casos.

En este trabajo se va a hacer uso de otro tipo de puntos para plantear futuras hipótesis sobre la localización del suelo de la escena. Haciendo un estudio de diversas imágenes (ver Fig. 11) a las cuales les ha sido aplicado el algoritmo de clasificación de la sección 2, se puede observar cómo las líneas pertenecientes a la dirección Z (representadas en azul), son más robustas que las líneas pertenecientes a las direcciones horizontales. Además en la mayoría de los casos todas estas líneas nacen desde el suelo y se extienden de forma radial hacia el exterior de la imagen. Por otro lado, las líneas horizontales son más susceptibles al ruido o a ser clasificadas de forma incorrecta, de forma que suelos y paredes con gran concentración de segmentos (véase muchos tipos de baldosas o ladrillos) harán aparecer líneas innecesarias sobre la imagen, resultando en una dificultad añadida.

En [4] se demuestra cómo, si se dispone de la calibración del sistema catadióptrico, es posible definir una cónica sobre la imagen a partir de tan solo dos puntos. Si además aplicamos la condición de que toda línea de la imagen ha de pasar por uno de los puntos de fuga, solo se necesita un punto más que pertenezca a la intersección de las paredes con el suelo para definir la cónica que actúa como límite separador de estos elementos estructurales.

Ahora bien, obviamente si el objetivo es encontrar dónde está situado el límite de separación entre pared y suelo, no podemos saber cuando un punto pertenece a éste o no, por lo que es necesario diseñar un algoritmo que sea capaz de deducir las regiones más probables en las que deben encontrarse estos puntos.

Se hace evidente que la información aportada por las líneas almacenadas en los tres grupos definidos con anterioridad es redundante y tan solo unos pocos puntos pertenecientes a estas son necesarios para generar las hipótesis de contorno.

De esta forma se definen los siguientes tres nuevos grupos (figura 4):

1) De cada línea que formaba parte del grupo de segmentos verticales (azules), se selecciona el punto más cercano al centro de la imagen, con objetivo de que este punto recaiga en la zona limitante entre suelo y pared. El conjunto de estos puntos formará el nuevo grupo  $G_Z$ .

2) Tomando las líneas pertenecientes al grupo de seg-

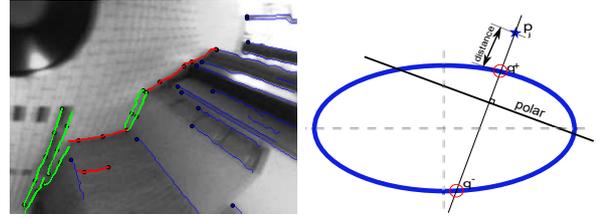


Figura 4: Izquierda: Selección de puntos clave, explicado en la sección 3.1. Derecha: Explicación gráfica de cómo medir la distancia entre punto y cónica de la sección 3.2.

mentos horizontales en dirección X (rojos), se seleccionan únicamente aquellas que están cercanas a alguno de los puntos del nuevo grupo  $G_Z$ , y estas líneas serán discretizadas tomando puntos cada cierto intervalo de longitud, los cuales pasarán a formar parte del grupo  $G_X$ .

3) Se repite el mismo proceso anterior para las líneas horizontales en dirección Y formando el grupo  $G_Y$ .

### 3.2 Generación de cónicas

Debido a que no conocemos que puntos se sitúan sobre la región del suelo, vamos a utilizar RANSAC para identificar las cónicas más votadas, candidatas a representar el borde deseado.

Como se mencionó anteriormente, solo son necesarios dos puntos, un VP y uno de los puntos de los grupos recién formados. El producto vectorial entre un VP y cualquiera de los puntos  $\mathbf{p}_i$  genera un vector normal  $\mathbf{n}_i$ , que define una cónica  $\bar{\Omega}_i$ , obteniendo finalmente  $\hat{\Omega}_i$  después de la transformación proyectiva  $H_C[1]$ :

$$\mathbf{n}_i = \begin{pmatrix} n_{ix} \\ n_{iy} \\ n_{iz} \end{pmatrix} = \begin{pmatrix} VP_x^S \\ VP_y^S \\ VP_z^S \end{pmatrix} \times \begin{pmatrix} P_{ix}^S \\ P_{iy}^S \\ P_{iz}^S \end{pmatrix} \quad (1)$$

$$\bar{\Omega}_i = \begin{bmatrix} n_{ix}^2 \rho - n_{iz}^2 \xi^2 & n_{ix} n_{iy} \rho & n_{ix} n_{iz} \\ n_{ix} n_{iy} \rho & n_{iy}^2 \rho - n_{iz}^2 \xi^2 & n_{iy} n_{iz} \\ n_{ix} n_{iz} & n_{iy} n_{iz} & n_{iz}^2 \end{bmatrix} \quad (2)$$

$$\rho = (1 - \xi^2)$$

$$\hat{\Omega}_i = H_C^{-t} \bar{\Omega}_i H_C^{-1} \quad (3)$$

Ahora, la distancia entre cada punto  $\mathbf{p}_j$  y la cónica generada por el punto  $\mathbf{p}_i$  con el punto de fuga VP, se mide usando una aproximación [4] de [14]. Se calcula la línea polar de un punto  $\mathbf{p}_j$  en la cónica  $\hat{\Omega}_i$ , se calcula la perpendicular a la línea polar, especificando que pase sobre el punto  $\mathbf{p}_j$ . Este segmento perpendicular corta a la cónica en dos puntos  $\mathbf{q}^+$  y  $\mathbf{q}^-$ , la mínima distancia Euclídea entre  $\mathbf{p}_j$  y  $\mathbf{q}^+$  o  $\mathbf{q}^-$  corresponde a la distancia entre punto y cónica, Fig. 4.

Con todos aquellos puntos que tengan una distancia menor a un umbral de la cónica, se estima una cónica media y se repite el proceso hasta que converge (no se encuentran más puntos cercanos). Los puntos que votan a esta cónica media son eliminados de la lista, y se selecciona un nuevo punto  $p_i$  de los restantes para generar una nueva cónica repitiendo el proceso hasta que todos los puntos votan al menos a una cónica.

### 3.3 Hipótesis de contorno inicial

Una vez que se tienen los datos necesarios para generar hipótesis de las posibles zonas donde se encuentra el borde entre pared y suelo surge un nuevo problema: no se sabe cuántos bordes estamos buscando y un algoritmo no puede identificar el número de paredes que componen una habitación, sin ninguna información adicional.

Sin embargo, los entornos de interior suelen estar contruidos según una serie de patrones. Se puede distinguir entre pasillos o habitaciones, los primeros comúnmente son alargados en forma de I, o con ramificaciones en uno de sus extremos cuando nos acercamos al final, normalmente adquiriendo forma de L o de T. En cuanto a habitaciones se refiere, la forma cuadrangular es la forma por excelencia, pudiendo contar esta con irregularidades en alguna de sus caras. La figura 5 muestra algunos de los ejemplos más comunes.

Todas estas formas geométricas pueden ser descompuestas como un cuadrado central del que nacen ramificaciones en algunas de sus caras. Además, este cuadrado central cumple una condición, al tener las caras paralelas dos a dos, cada una de las caras paralelas entre sí, debe caer a cada uno de los lados formados por la línea imaginaria que une sus correspondientes puntos de fuga. Esto es debido a la definición de punto de fuga como lugar geométrico donde convergen las líneas paralelas.

Debido a esta propiedad, el algoritmo de votación de cónicas (Section 3.2) se ejecuta para cuatro casos particulares, con el objetivo de encontrar los 4 primeros bordes que definan el suelo de esta primera hipótesis (ver Fig. 6):

- Los bordes 1 y 3 serán las líneas que definen los límites del suelo con orientación  $X$ , y por tanto se usaran los puntos de los grupos  $G_Z$  y  $G_X$  para buscar cónicas a cada lado de la línea imaginaria formada al unir los  $VPs$  en dirección  $X$  (Fig. 6 (izquierda)).
- Los bordes 2 y 4 definirán los bordes con orientación  $Y$ , y las cónicas se buscaran a cada lado de la línea imaginaria formada al unir los  $VPs$  en dirección  $Y$  usando los puntos de los grupos  $G_Z$  y  $G_Y$  (Fig. 6 (centro)).

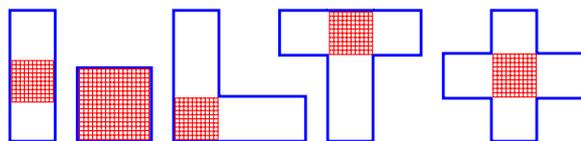


Figura 5: Formas más comunes de pasillos/habitaciones (vista superior). La rejilla roja representa el cuadrado básico central que se busca en la sección 3.3.

Otra propiedad es que los 4 puntos de fuga definen un círculo. Este círculo corresponde a los puntos de la escena situados a la misma altura que la cámara, por lo que los puntos que caen en el interior de este círculo estarán a una altura inferior y por tanto pueden corresponder a puntos del suelo, mientras que los que están en el exterior del círculo es seguro que no pueden pertenecer al suelo y son eliminados de manera que no formen parte de la votación en la elección de cónicas.

De entre las cónicas más votadas para formar los límites del suelo de la hipótesis de habitación de 4 paredes se eligen aquellas más cercanas al centro de la imagen. Esto se debe a que al estar buscando el cuadrado central, es preferible seleccionar un cuadrado más pequeño que luego tiene posibilidades de ser ampliado en el proceso de expansión (Sección 3.4).

Una vez que las cuatro fronteras han sido halladas, se unen para formar las paredes y suelos de esta primera hipótesis (Fig. 6).

### 3.4 Proceso jerárquico de expansión

Denotemos como  $B_1$ ,  $B_2$ ,  $B_3$  y  $B_4$  las cuatro fronteras entre pared y suelo definidas en la sección 3.3. El área entre cada borde y el exterior de la imagen define cuatro sectores (Fig. 6). Estos sectores pueden corresponder a paredes de la escena si los bordes de separación coinciden con los de la imagen real, o cabe la posibilidad de que éstos puedan ser expandidos, entendiendo como expansión a reemplazar la frontera  $B_i$  por otras líneas cónicas que agranden el área del suelo de la primera hipótesis de forma que estas sucesivas hipótesis aproximen mejor a la forma real de la habitación. Para cada uno de estos sectores vamos a probar posibles combinaciones con las cónicas extraídas de la sección 3.3, de forma que se obtengan un máximo de 3 nuevos bordes que amplíen el área del suelo de ese sector. Vamos a denotar a estas posibles nuevas fronteras como  $B_i^L$ ,  $B_i^M$  y  $B_i^R$  siguiendo el sentido horario tal como se muestra en la Fig. 7.

Cuando tratamos de expandir pueden ocurrir 3 casos:

- Se dan condiciones bien definidas que permiten identificar los 3 nuevos bordes  $B_i^M$ ,  $B_i^L$  y  $B_i^R$ ; y por tanto habrá expansión en el sector actual.

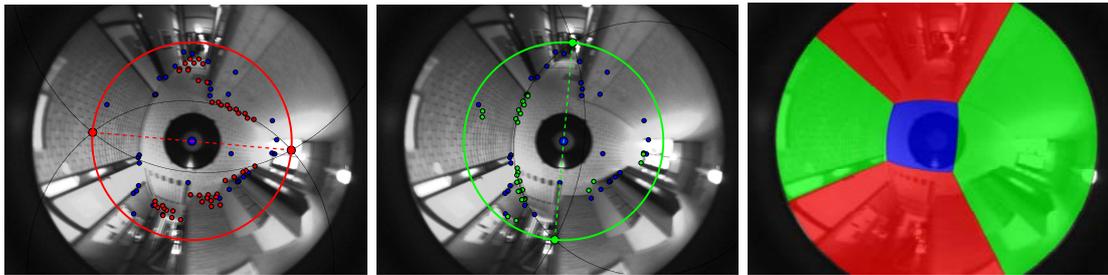


Figura 6: Las primeras dos imágenes muestran puntos de los grupos  $G_Z$  (azul),  $G_X$  (rojo) and  $G_Y$  (verde) bajo las condiciones expuestas en la sección 3.3. Los segmentos discontinuos rojo y verde son las líneas imaginarias que unen los respectivos VPs y dividen la imagen en dos partes. Finalmente, las cónicas negras representan la cónica ganadora para cada uno de los 4 casos. La imagen de la derecha muestra el resultado de combinar estas cónicas para generar suelos y paredes de la primera hipótesis.

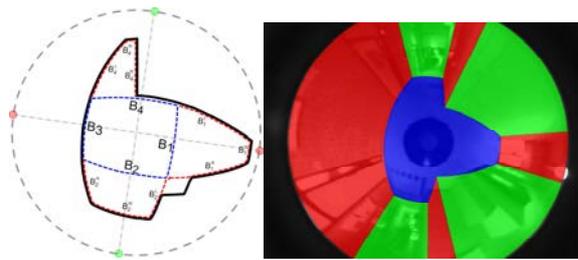


Figura 7: Izquierda: Ejemplo simulado que muestra los posibles casos ( $B_1$  y  $B_2$  son regiones que se pueden expandir,  $B_3$  no tiene expansión posible, y  $B_4$  corresponde a una esquina oculta). Las líneas negras representan el contorno real de la habitación, las rayas discontinuas azules muestran la primera hipótesis, y el resultado final viene marcado por las líneas discontinuas en rojo. Derecha: Resultado final de una imagen real.

- El borde medio  $B_i^M$  detectado está muy cercano a  $B_i$ , esto significa que la pared más votada sigue siendo la misma y no habrá expansión.
- De los tres bordes a encontrar solo 1 ó 2 están bien definidos. Esto se puede dar por diferentes causas y se debe estudiar antes de saber si habrá o no expansión del sector.

Este último caso se presenta cuando no tenemos suficientes datos, debido a líneas de la imagen que no han sido detectadas o a la presencia de esquinas ocultas (Fig. 7(izquierda)).

Si una de las fronteras no detectadas es uno de los bordes laterales ( $B_i^L$ ,  $B_i^R$ ) o uno de los laterales y el central ( $B_i^M$ ), es muy posible que se deba a la presencia de una esquina oculta. Esto se debe a que una de las paredes se encuentra en un ángulo de visión muerto para la cámara y es tapada por otra pared. Por esta razón, no se puede encontrar cónica que defina el borde no detectado y ha de ser definida como la línea radial que cruza desde el centro de la imagen hasta el punto, que votando a uno de los bordes bien detectados, se encuentre más cercano al VP característico del sector (ver ejemplo 2 de la Fig. 11).

Por otro lado, si no se dan las condiciones anteriores,

se asume que simplemente alguno de los límites no se detectó bien, por lo que si el borde perdido es uno de los laterales ( $B_i^L$  o  $B_i^R$ ), este será redefinido como la cónica resultante que cruza por el VP característico del sector, y el punto, que votando al borde  $B_i^M$  bien detectado, se encuentre más a la izquierda o a la derecha dependiendo de si el límite que estamos buscando es  $B_i^L$  o  $B_i^R$  respectivamente. Por el contrario, si la frontera no detectada es  $B_i^M$ , se considerará que es debido a que no hay suelo que expandir.

#### 4 Propagación mediante homografía

En las secciones previas se ha descrito el método propuesto para determinar la distribución espacial de un escenario a partir de una sola imagen omnidireccional extraída por un sistema hyper-catadióptrico.

Por otra parte, el objetivo final de este algoritmo es ser aplicado a una secuencia de imágenes, donde los cambios entre las sucesivas son pequeños, por lo que la distribución estructural obtenida para esta sucesión se pretende que varíe de forma coherente.

Con el objetivo de mejorar la precisión del método en imágenes donde los resultados no son los esperados, y poder hacer la información extraída de cada una de las imágenes de la secuencia lo más robusta posible, se hará uso de homografías[8] (Fig. 8).

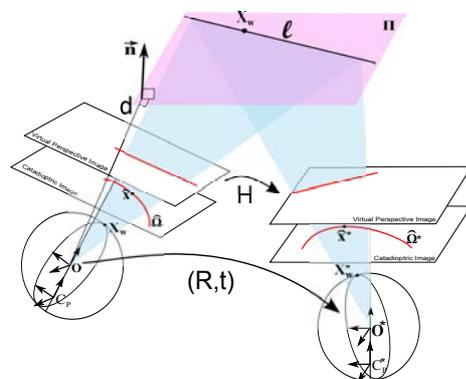


Figura 8: Homografía entre dos puntos de vista  $O$  y  $O^*$ .

## 4.1 Cálculo de la Homografía

Existen dos variaciones para calcular las homografías, por emparejamiento de puntos o de líneas [7]. Para este trabajo es más adecuada la segunda opción ya que las líneas que ejercen de borde pared-suelo pueden emplearse a la hora de calcular la matriz de homografía  $H$  sin necesidad de extraer características (por ejemplo mediante el descriptor SIFT [11]) únicamente para resolver la homografía.

En principio, se necesitan 4 parejas de puntos/líneas para resolver el sistema de ecuaciones lineales que determina la homografía, aunque es posible imponer condiciones adicionales que permiten eliminar términos de la matriz  $H$ .

**a) Condición de movimiento plano:** Se considera que la cámara se mueve sobre un plano horizontal, por lo que la matriz de giro se restringe a rotar sobre el eje vertical, el vector normal al plano solo tendrá componente vertical, y el desplazamiento solo constará de movimiento en coordenadas  $x$  e  $y$ .

$$R = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$\mathbf{n}_F = [n_{F_x}, n_{F_y}, n_{F_z}]^T = [0, 0, 1]^T \quad (5)$$

$$\mathbf{t} = [t_x, t_y, t_z]^T = [t_x, t_y, 0]^T \quad (6)$$

Sustituyendo en la expresión  $H = \mathbf{R}(\mathbf{I} + \mathbf{t}\mathbf{n}_F^T/d)$ :

$$H = \begin{bmatrix} \cos \theta & \sin \theta & t_x/d \\ -\sin \theta & \cos \theta & t_y/d \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

**b) Imposición de rotación:** Hacer “tracking” de los puntos de fuga proporciona información de giro entre imágenes (ángulo de rotación  $\theta$ ), el cual podemos introducir en  $H$  de forma que las únicas incógnitas restantes son:  $t_x$ ,  $t_y$  y  $d$ , pero están agrupadas en dos términos y cómo para esta aplicación no interesa conocer su valor particular se consigue un sistema de ecuaciones resoluble a partir de un solo emparejamiento.

Sea  $\mathbf{n}$  la normal que defina una línea en la imagen  $I$ , y  $\mathbf{n}^*$  la normal que representa el emparejamiento de  $\mathbf{n}$  visto en la imagen  $I^*$ , estas líneas están relacionadas por:

$$\mathbf{n}^* \propto H^{-T} \mathbf{n} \quad (8)$$

Desarrollando la matriz inversa y transpuesta de  $H$ ;  $H^{-T} = [h_{ij}]$  con  $i, j = 1, 2, 3$ , se observa que ésta sigue dependiendo únicamente de dos parámetros desconocidos ( $h_{31}$  y  $h_{32}$ ) compuestos por una combinación lineal de los parámetros de  $H$  pero cuyo valor individual es irrelevante.

El sistema de ecuaciones resultantes es:

$$\begin{bmatrix} n_x n_y^* & n_y n_y^* & \lambda_1 \\ -n_x n_x^* & -n_y n_x^* & \lambda_2 \end{bmatrix} \begin{bmatrix} h_{31} \\ h_{32} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (9)$$



Figura 9: A la izquierda se muestra la hipótesis de suelo de una imagen  $I$  (rojo). En el centro se muestra la hipótesis de suelo de una imagen  $I^*$  en blanco y dos casos de posibles homografías al proyectar la hipótesis de la imagen  $I$  sobre la imagen  $I^*$ . La foto del centro corresponde con una mala homografía y la de la derecha con la homografía ganadora.

$$\lambda_1 = n_z n_y^* - n_x n_z^* \sin \theta - n_y n_z^* \cos \theta$$

$$\lambda_2 = n_x n_z^* \cos \theta - n_y n_z^* \sin \theta - n_z n_x^*$$

que se puede resolver mediante Descomposición en Valores Singulares (SVD), de donde se obtiene  $h_{31}$  y  $h_{32}$ , componiendo así la matriz  $H^{-T}$ , y al deshacer la inversión y transposición se recupera  $H$ .

A priori no se conocen los emparejamientos entre las líneas de ambas hipótesis, por lo que una opción sería calcular las homografías obtenidas para todas las posibles combinaciones entre éstas. Sin embargo, muchas combinaciones no tienen sentido físico (líneas de puntos opuestos de la imagen), por lo que solo es necesario calcular homografías para aquellas líneas cuyas normales indiquen cercanía. Cada combinación dará como resultado una Homografía que será más o menos buena (Fig. 9) de las cuales se elegirá cómo ganadora aquella que proporciona mayor similitud entre la proyección del suelo hipótesis de la imagen  $I$  sobre  $I^*$  y la propia hipótesis de la imagen  $I^*$ .

## 4.2 Aplicación secuencial

Este último paso consiste en comparar las hipótesis de varias imágenes, de forma que se pueda obtener una “hipótesis media” que sea la que mejor se adapte al conjunto de contornos generados.

Para poder hacer la comparación es necesario que todos los fotogramas se encuentren proyectados en la misma imagen, por lo que hacemos uso de las homografías recién definidas de forma que proyectemos sobre la imagen  $I$  un número  $N$  de hipótesis de imágenes anteriores ( $h_1, \dots, h_N$ ). En este momento podemos comparar las hipótesis entre sí, de esta forma, si existen hipótesis con distinto número de bordes pared-suelo, se elegirán cómo contorno base aquel que tenga más rasgos comunes con el resto (Fig. 10(centro)).

Finalmente este contorno base se promedia con aquellas hipótesis que tengan la forma más parecida a él, de forma que las hipótesis malas apenas tienen relevancia en el resultado final (Fig. 10(dcha)).

Este resultado corresponde a la hipótesis final ( $h^F$ ) que muestra el contorno del suelo de la escena  $I$  bajo

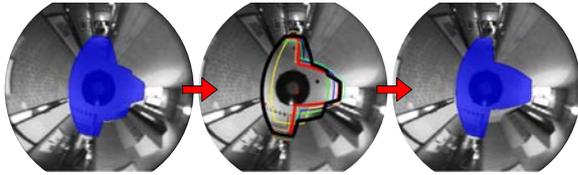


Figura 10: Izqda. Ejemplo de una hipótesis con defectos (el suelo abarca zonas que son pared). Centro: Sobre la hipótesis de suelo actual (negro) se proyectan hipótesis de imágenes anteriores, y se elige la que más se parece al global (rojo). Dcha: Se promedia la hipótesis ganadora (rojo) con las cercanas para dar el resultado final.

estudio. A su vez, este resultado obtenido se guardará, de forma que al iterar el proceso, se propague para ser considerado en las próximas votaciones entre hipótesis. De esta forma podemos guardar un número  $M < N$  de hipótesis finales ( $h_1^F, \dots, h_M^F$ ) de forma que cuanto mayor sea  $M$  los resultados serán más robustos ante errores, pero a su vez el algoritmo será menos flexible a detectar cambios rápidos en la forma de las habitaciones.

En nuestros experimentos hemos tomado cómo valores  $N = 7$  imágenes anteriores a la frame actual que estamos estudiando, y  $M = 2$  hipótesis finales consideradas para aumentar la robustez de el resultado, sin afectar en exceso a cambios de la escena.

## 5 Resultados

Nuestros experimentos se han ejecutado en Matlab, en un tiempo de 3s por imagen, y con la base de datos COGNIRON compuestas por imágenes de interior con una resolución de  $768 \times 1024$  donde hay presentes una gran cantidad de habitaciones diferentes. Estas imágenes han sido tomadas por un sistema de cámara perspectiva y espejo hiper-catadióptrico montadas sobre un robot móvil. La base de datos y calibración del sistema están disponibles en [16].

Algunos de los resultados en diferentes tipos de escenas de interior son mostrados en la Fig. 11. Los dos primeros ejemplos corresponden a pasillos con forma T y L (cómo los que se describen en la Fig. 5), las paredes no están muy saturadas de objetos, por lo que los resultados son precisos. En el segundo ejemplo cabe destacar que se observa una esquina oculta en la parte superior de la imagen.

En el último caso se muestra un recibidor con un escritorio y una estantería. Nuestro algoritmo es capaz de reconocer estos obstáculos, sin embargo, no llega a detectar la puerta abierta que se encuentra en la parte superior de la imagen, probablemente al exceso de iluminación que entra a través de ésta. Una tabla comparando la precisión de los resultados de nuestro algoritmo puede encontrarse en [12].

También se muestra una secuencia de 7 imágenes

(Fig. 12) usando y sin usar el algoritmo basado en homografía, dónde se aprecia cómo los resultados aplicando homografía son más homogéneos evitando así malas clasificaciones.

## 6 Conclusiones

Hemos propuesto un nuevo método para detectar puntos de fuga y clasificar líneas sobre imágenes catadióptricas, que es más rápido que los métodos existentes. A su vez, se propone un método para extraer la distribución estructural (suelo y paredes) de una escena a partir de una sola imagen omnidireccional. Dicho método se complementa con un proceso secuencial basado en homografías que permite eliminar errores de segmentación mejorando los resultados obtenidos.

Nuestro algoritmo es útil en muchas aplicaciones de navegación, detección de obstáculos y reconstrucción 3D ya que conociendo dónde se encuentran los límites entre pared-suelo y a que altura se sitúa la cámara, es posible calcular la situación exacta de cada punto de la imagen.

### Agradecimientos

El trabajo ha sido financiado por el Ministerio de Ciencia e Innovación/Unión Europea a través del proyecto DPI2009-14664-C02-01 y por DGA-FSE (grupo T04).

## Referencias

- [1] J. Barreto. *General Central Projection Systems: Modeling, Calibration and Visual Servoing*. PhD thesis, 2003.
- [2] J. C. Bazin, I. Kweon, C. Démonceaux, and P. Vasseur. A robust top-down approach for rotation estimation and vanishing points extraction by catadioptric vision in urban environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 346–353, 2008.
- [3] J. Bermudez, L. Puig, and J. J. Guerrero. Line extraction in central hyper-catadioptric systems. In *OMNIVIS*, 2010.
- [4] J. Bermudez-Cameo, L. Puig, and J. J. Guerrero. Hypercatadioptric line images for 3d orientation and image rectification. *Robotics and Autonomous Systems*, 60(6):755–768, 2012.
- [5] J. M. Coughlan and A. L. Yuille. Manhattan world: Compass direction from a single image by bayesian inference. In *Int. Conf. on Computer Vision*, pages 941–947, 1999.
- [6] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical applications. In *ECCV (2)*, pages 445–461, 2000.
- [7] J. J. Guerrero and C. Sagüés. From lines to homographies between uncalibrated images. In *IX Spanish*

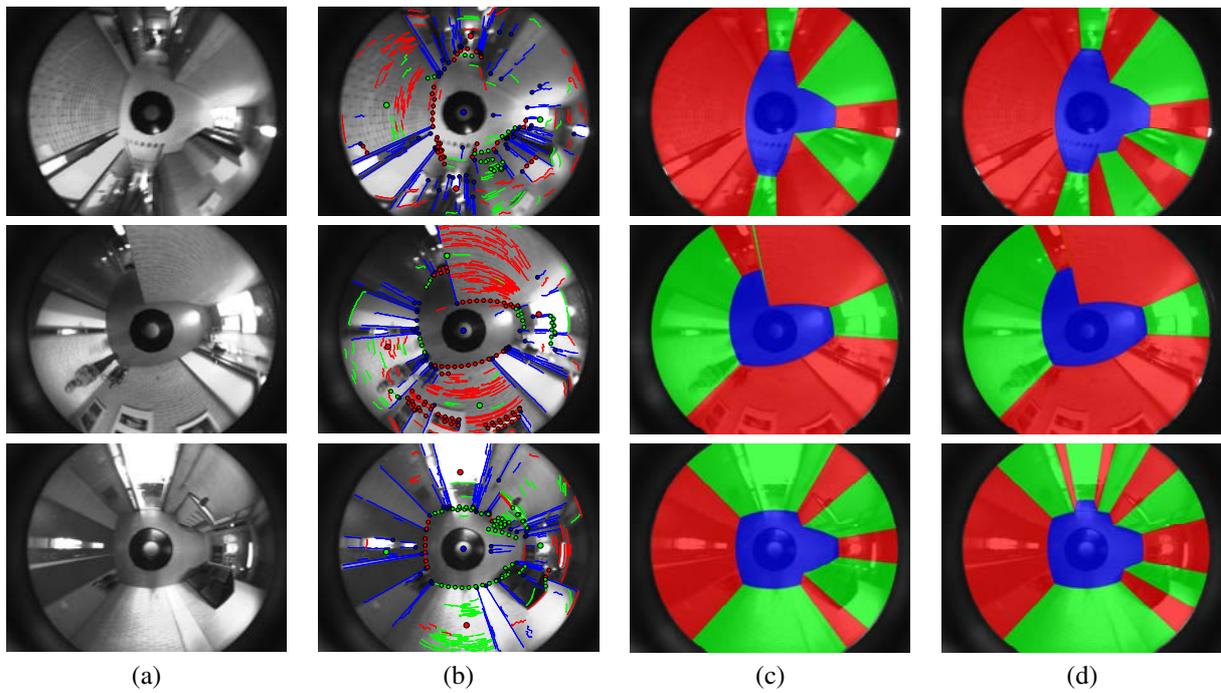


Figura 11: Ejemplos de resultados experimentales obtenidos para tres imágenes diferentes. (a) Imagen original. (b) Clasificación de líneas y extracción de puntos que votarán en la elección de bordes. (c) Salida final de nuestro algoritmo. (d) Resultado deseado, etiquetado manualmente.

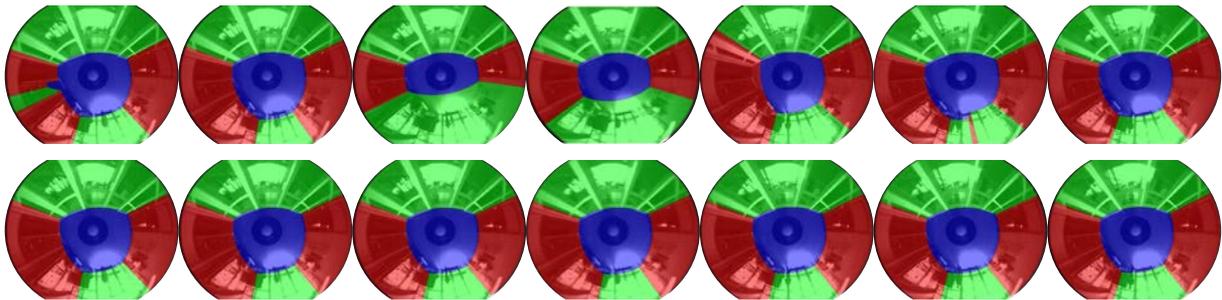


Figura 12: Secuencia de 7 imágenes seguidas. La primera fila muestra resultados obtenidos sin aplicar la homografía. En la segunda fila se puede ver cómo incluyendo el uso de las homografías los resultados son más homogéneos y se corrigen los posibles errores de las hipótesis originales.

- Symposium on Pattern Recognition and Image Analysis*, pages 233–240, 2001.
- [8] H. Hadj-Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet. Decoupled homography-based visual servoing with omnidirectional cameras. In *IROS*, pages 2332–2337, 2006.
- [9] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. In *IEEE International Conference on Computer Vision*, pages 1849–1856, 2009.
- [10] D. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2136–2143, June 2009.
- [11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [12] J. Omedes, G. López-Nicolás, and J. J. Guerrero. Omnidirectional vision for indoor spatial layout recovery. In *12th IAS Intelligent Autonomous Systems Conference*, pages 1–5, Jeju Island, June, 2012.
- [13] N. D. Ozisik, G. López-Nicolás, and J. J. Guerrero. Scene structure recovery from a single omnidirectional image. In *ICCV Workshops*, pages 359–366, 2011.
- [14] P. Sturm and P. Gargallo. Conic fitting using the geometric distance. In *Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan*, pages 784–795, 2007.
- [15] G. Tsai, C. Xu, J. Liu, and B. Kuipers. Real-time indoor scene understanding using bayesian filtering with motion cues. In *ICCV*, pages 121–128, 2011.
- [16] Z. Zivkovic, O. Booij, and B. Krose. From images to rooms. *Robotics and Autonomous Systems*, 55(5):411–418, 2007.