CHAPTER X

SISTEMA DE LOCALIZACIÓN Y MAPEADO MEDIANTE VISIÓN ARTIFICIAL PARA ASISTIR A LA NAVEGACIÓN DE PERSONAS INVIDENTES

P. F. ALCANTARILLA[1], L. M. BERGASA[1], R. BAREA[1], E. LOPEZ[1], M. OCAÑA[1]

[1]Departamento de Electrónica - Universidad de Alcalá, {pablo.alcantarilla, bergasa, barea, elena, mocana}@depeca.uah.es

En este artículo, se plantea la aplicación de técnicas de SLAM (*Simultaneous Localization and Mapping*) utilizando un sistema sensorial basado en una cámara estéreo de bajo coste y un GPS para la ayuda a la asistencia a la navegación de personas invidentes. Para ello es necesario un sistema sin restricciones en el movimiento, al ser una persona quien transporta al sistema sensorial y fusionar la información visual con la información de un GPS, útil en entornos exteriores donde no existe ocultamiento de satélites. El primer requisito para poder proporcionar al usuario una localización fiable, consiste en el proceso de estimación de un mapa 3D del entorno basado en marcas visuales. Para tal fin, se utiliza un algoritmo SLAM de 6DOF basado en el *Filtro Extendido de Kalman* (EKF). El sistema combina información de orientación y profundidad mediante el uso de dos parametrizaciones distintas de las marcas: profundidad inversa y 3D respectivamente. Los puntos de interés son detectados en la imagen y clasificados como marcas 3D o profundidad inversa, a partir de un umbral de profundidad obtenido mediante un estudio de no-linealidad. En el momento de inicialización de cada marca, se calcula el vector normal del plano que contiene a la marca, utilizando la información del par estéreo, para posteriormente transformar la apariencia del parche ante los diversos puntos de vista de la cámara, en el proceso de seguimiento de las mismas. Finalmente, se presentan diversos resultados experimentales en interiores así como las principales conclusiones obtenidas.

1 Introducción

La investigación en técnicas SLAM (*Simultaneous Localization and Mapping*) en tiempo real es uno de los campos más importantes en la robótica actual. Recientemente el campo de Visual SLAM ha cobrado gran interés debido principalmente a tres razones: las cámaras son sensores más baratos que los tradicionales scan-lasers, proporcionan una rica información visual sobre el entorno y pueden ser fácilmente adaptados a sistemas *wearable*. Por lo tanto, el campo de aplicaciones de las técnicas SLAM, se ha extendido desde aplicaciones típicas robóticas hasta campos tales como la cirugía no-invasiva (Mountney, 2006), realidad aumentada (Klein, 2007) y localización de vehículos en grandes entornos (Schleicher, 2007).

La percepción visual de marcas o puntos de interés en un entorno, es crucial en muchos aspectos de la vida diaria. Esta percepción visual puede ser útil en aplicaciones tales como cruzar una carretera, bajar a una estación de metro o encontrar la oficina de un profesor en una universidad. Muchas de las actividades anteriores, a pesar de que pueden ser consideradas como sencillas, pueden suponer un gran desafío para las personas invidentes. Por lo tanto, para este tipo de usuarios es muy interesante el desarrollar un sistema *wearable* capaz de proporcionarles información visual sobre el entorno a partir de un medio o dispositivo no visual, para proporcionar a la persona invidente un mayor conocimiento de su entorno y por lo tanto una mejor navegación a través del mismo.

En este trabajo, se presenta un sistema de SLAM métrico 6DOF utilizando una cámara estéreo movida con la mano como único sensor, proporcionando al usuario tanto el mapa del entorno así como su localización en el mismo. Nuestro sistema sienta las bases de un sistema de SLAM de alto nivel para las personas invidentes, en el cuál la información visual se fusionará con la información procedente de un GPS para entornos exteriores. Debido a que los usuarios finales del sistema son personas, no existen restricciones especiales en el movimiento de la cámara, sin embargo, se espera que el movimiento de la cámara sea suave y el usuario camine a una velocidad normal entre 3 – 5 km/hora. Las principales ventajas de utilizar un sensor estéreo en vez de uno monocular, son descritas en (Schleicher, 2006).

La solución adaptada se basa en los trabajos previos en SLAM monocular en (Davison, 2007), pero adaptando la filosofía del trabajo anterior a un

sistema de visión estéreo. En este tipo de sistemas, únicamente un reducido grupo de marcas naturales de alta calidad, extraídas de la imagen utilizando algún detector de puntos de interés (Harris, 1988), son seguidas a lo largo del tiempo y utilizadas para calcular la pose de la cámara creando un mapa no denso de marcas naturales utilizando un Filtro Extendido de Kalman (EKF). Paz *et al* presentaron en (Paz, 2008) un sistema 6DOF EKF-SLAM utilizando un sensor estéreo movido con la mano para grandes entornos tanto en interiores como en exteriores. La parametrización de profundidad inversa propuesta por Civera *et al* en (Civera, 2008) para el SLAM monocular, se adapta en la versión estéreo proporcionando información de profundidad y orientación. Las marcas naturales son extraídas de la imagen y clasificadas como puntos 3D si la disparidad horizontal es superior a un umbral, o clasificadas como marcas de profundidad inversa en caso contrario. Su algoritmo Visual SLAM genera mapas locales condicionalmente independientes y finalmente, el mapa final es obtenido utilizando el algoritmo *Conditionally Independent Divide and Conquer*, que les permite obtener un tiempo de cómputo constante la mayoría del tiempo (Paz, 2007). Aunque los resultados son buenos considerando grandes entornos tanto en interiores como exteriores, la variedad de movimientos de la cámara es limitada, ya que no se realiza ninguna transformación de la apariencia del parche ante cambios de punto de vista de la cámara, realizándose únicamente correlaciones 2D de parches en la imagen en el proceso de matching. Por medio de un estudio empírico, sugieren utilizar un umbral de profundidad de 5 m, para conmutar entre marcas 3D y de profundidad inversa.

Las dos principales contribuciones de nuestro trabajo, son la determinación de un umbral óptimo de profundidad para conmutar entre marcas de profundidad inversa y marcas 3D por medio de un estudio de no-linealidad, y la transformación de la apariencia del parche mediante una homografía 2D utilizando la información procedente de las dos cámaras. Este artículo esta estructurado de la siguiente manera: en primer lugar en la Sección 2 se realiza una descripción de la arquitectura global del sistema final, en la Sección 3 se describe el sistema de SLAM visual, mientras que las Secciones 4 y 5 presentan el análisis de no-linealidad y el método de homografía 2D. Finalmente, en las Secciones 6 y 7 se muestran los experimentos en interiores y las conclusiones y trabajos futuros respectivamente.

2 Descripción del Sistema

La localización por medio de medidas sensoriales, es un aspecto fundamental para una correcta navegación por un entorno. Sin embargo, en muchas aplicaciones las medidas de ciertos sensores son erróneas o no están disponibles. El error cometido utilizando únicamente localización basada en GPS puede ser muy elevado, especialmente en entornos urbanos en los que puede existir ocultamiento de satélites, y pérdida de señal debido a edificios altos y/o vegetación. Igualmente, para entornos en interiores la señal GPS se encuentra no disponible la mayoría de las ocasiones, lo cuál impide una localización y navegación fiables utilizando este único sensor, como por ejemplo para aplicaciones en las que una persona invidente pueda acceder a una determinada aula en un colegio o universidad.

En este trabajo se propone un sistema que ayude a la localización de personas invidentes combinando visión y medidas GPS. La localización en entornos interiores se basará en la información obtenida a partir de un sistema de visión estereoscópica, mientras que para entornos exteriores se realizará una fusión entre visión y GPS. En (Oh, 2004) se pueden ver ejemplos de localización en exteriores para personas invidentes utilizando únicamente señal GPS y mapas del entorno conocidos a priori. En nuestro caso, proponemos también el uso de auriculares ya que representa el medio más intuitivo de transmisión de la información visual para este tipo de usuarios. En la Fig. 1 se muestra un esquema del sistema final de ayuda a la navegación de personas invidentes:

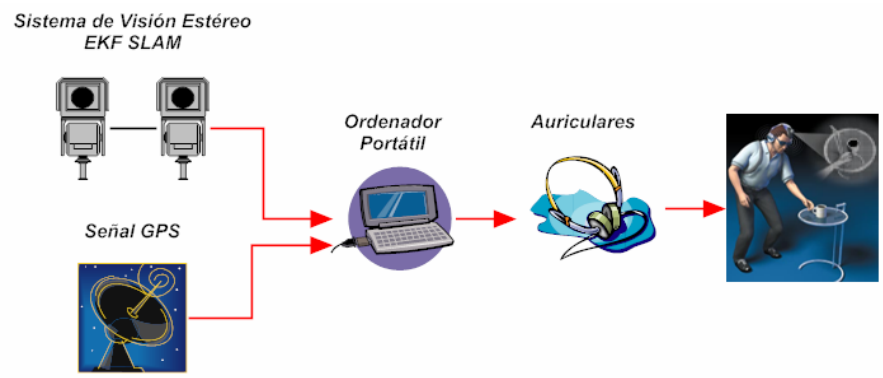

Fig. 1. Descripción del sistema de ayuda a la navegación de personas invidentes

3 Visual SLAM

El sistema desarrollado consiste en una cámara estéreo con lentes de gran angular transportada por la mano y un ordenador portátil para el procesado de las imágenes en tiempo real. La Fig. 2 muestra nuestro sistema estéreo y el tipo de entornos en interiores donde se han desarrollado los experimentos.

Fig. 2. Sistema de visión estéreo utilizado 6DOF SLAM y entorno de pruebas experimentales

El vector de estado del sistema X incorpora la información referente a la cámara izquierda y a las marcas que constituyen el mapa del entorno. El vector de estado correspondiente a la cámara consta de la posición 3D de la misma en coordenadas cartesianas, la orientación de la cámara expresada mediante un cuaternión, y la velocidad lineal y angular de la cámara, ya que ambas velocidades son necesarias en el modelo de movimiento impulsivo utilizado para modelar el movimiento de la cámara (Davison, 2007).

$$X_c = \left(X_{cam}, q_{cam}, v_{cam}, \omega_{cam}\right)^t \tag{1}$$

Se han utilizado dos tipos distintos de parametrización de las marcas, 3D y profundidad inversa, proporcionando información de profundidad u

orientación respectivamente. Dependiendo de la profundidad a la que se encuentran las marcas, estas son inicializadas como 3D o profundidad inversa, e incorporadas al vector de estado del filtro EKF:

$$X = \left(X_c, Y_{1,3D} \cdots Y_{n,3D}, Y_{1,INV} \cdots Y_{N,INV}\right)^t \quad (2)$$

Las marcas que componen el mapa 3D, se corresponden con puntos de interés en la imagen, y son detectados utilizando el detector de esquinas de Harris (Harris, 1988) más una posterior aproximación subpixélica. A medida que la cámara se mueve, se realiza un seguimiento de las marcas para actualizar el filtro a partir de las medidas de posición de las marcas en el plano imagen. Para poder realizar el seguimiento a una marca, se predice la posición en el plano imagen de la misma en ambas cámaras. Luego, la apariencia de la marca se transforma según el punto de vista actual de la cámara utilizando una homografía 2D (ver Sec. 2), y se realiza una correlación en un área de búsqueda delimitada por la incertidumbre actual de la marca y de la cámara, utilizando una correlación normalizada cruzada de media nula (ZMNCC). Si se obtiene un valor de correlación alto, por ejemplo superior a 0.8, la medida se da por buena. Se ha utilizado este tipo de correlación debido a su gran robustez e independencia a variaciones de iluminación (Faugeras, 93). Se ha utilizado un algoritmo de gestión inteligente del mapa, para poder eliminar del vector de estado la información de aquellas marcas que presentan una baja calidad a partir del ratio *número de medidas exitosas / número de intentos de medida*.

Debido al uso de lentes de gran angular, es necesario corregir la distorsión que presentan las imágenes originales. Al contrario que en algunos sistemas de SLAM como en (Schleicher, 2006, Davison, 2007), la distorsión radial y tangencial se corrigen utilizando tablas de búsqueda (LUT, *Look up tables*), de tal modo que se trabaja directamente con las imágenes sin distorsión. Las dos principales ventajas de utilizar LUTs son: primero, es un método de corrección de distorsión más rápido que trabajar con las imágenes distorsionadas y posteriormente corregir con modelos de distorsión las coordenadas en el plano imagen, y segundo, el proceso de *matching* es menos crítico si se utilizan las imágenes sin distorsión.

3.1 Marcas 3D

Para las marcas 3D, el vector de estado contiene la información referente a la posición 3D de la marca con respecto a un sistema de referencia global (por defecto se elige la posición de la cámara en el primer frame).

$$Y_{3D} = (x, y, z)^t \tag{3}$$

3.2 Marcas de profundidad inversa

Para las marcas con parametrización de profundidad inversa, el vector de estado contiene la información referente a la posición 3D de la cámara desde la cuál la marca fue inicializada (X_{ORI}), la orientación del rayo que va desde la cámara hasta la marca definido por los ángulos de azimuth (θ) y elevación (φ) y la inversa de la profundidad (ρ).

$$Y_{INV} = (X_{ORI}, \theta, \phi, \rho)^t \tag{4}$$

En la Fig. 3 se muestra la representación gráfica de este tipo de parametrización:

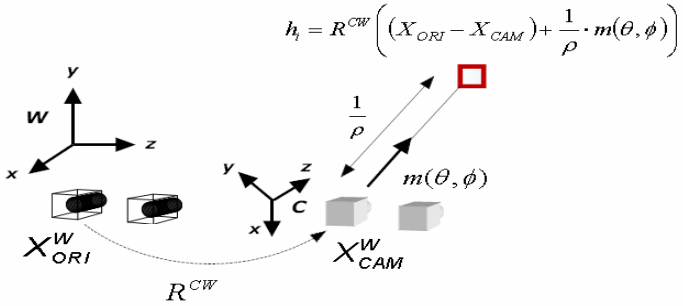

Fig. 3. Parametrización de profundidad inversa

En la figura anterior, m(θ, φ) es el vector unitario cuya dirección viene definida por la posición de la cámara y de la marca. Los ángulos de azimuth y elevación se definen a partir de las siguientes expresiones:

$$\theta = \tan^{-1}\left(\frac{z}{x}\right) \tag{5}$$

$$\phi = \tan^{-1}\left(\frac{\sqrt{x^2 + y^2}}{y}\right) \tag{6}$$

4 Análisis de No-Linealidad de la profundidad e información angular

Las investigaciones actuales en SLAM monocular han demostrado los beneficios de utilizar una parametrización de profundidad inversa, debido a que utilizando este tipo de parametrización se obtiene una inicialización sin retardo de las marcas (debido al problema existente de determinar la profundidad con una sola cámara) y además permite trabajar con marcas que se encuentran en el infinito así como cercanas a la cámara (Civera, 2008). Para el caso estéreo el utilizar o no una parametrización de profundidad inversa no es tan crítica como en el caso monocular, ya que la profundidad de las marcas puede ser obtenida utilizando la información del par estéreo. Sin embargo, existen ciertos beneficios al utilizar una parametrización de profundidad inversa en el caso de visión estereoscópica, ya que es una mejor representación de las marcas lejanas o en el infinito, debido a que la incertidumbre existente en la profundidad de una marca lejana medida con un par estéreo es grande, mientras que si esta marca lejana se modela en función de la orientación de marca con respecto a la cámara y la inversa de su profundidad, esta incertidumbre es mucho menor, obteniendo un sistema más lineal, lo que mejora los resultados de filtrado utilizando el EKF.

El principal problema de utilizar una parametrización de profundidad inversa, es el coste computacional que supone el representar una marca con un vector de 6 parámetros en vez de los 3 parámetros que constituyen la representación de una marca 3D. Este coste computacional extra, dependiendo del tipo de aplicación puede llegar a ser significativo. En el caso de visión estéreo, el problema radica en decidir la profundidad a la cuál una marca debe ser parametrizada como marca 3D o como marca con profundidad inversa, lo que es lo mismo, a partir de qué distancia es mejor utilizar información de profundidad o de orientación.

En este trabajo se propone un estudio de no-linealidad para encontrar un umbral óptimo de profundidad para conmutar entre ambos tipos de parametrizaciones. Se dice que una función es lineal en un intervalo, si la primera derivada es constante en ese intervalo, y por lo tanto, la segunda derivada es nula en dicho intervalo. Considerando la expansión de *Taylor* para la primera derivada de una función continúa f que depende de la variable Z:

$$\frac{\partial f}{\partial Z}(z+\Delta z) \approx \left.\frac{\partial f}{\partial Z}\right|_z + \left.\frac{\partial^2 f}{\partial Z^2}\right|_Z \cdot \Delta z \tag{7}$$

A partir del cociente entre la segunda y la primera derivada, se puede obtener un índice de no-linealidad de la función *f* con respecto a la variable *Z*:

$$L_f = \left| \frac{\frac{\partial^2 f}{\partial Z^2} \cdot \Delta Z}{\frac{\partial f}{\partial Z}} \right| \tag{8}$$

La Ec. 7 puede ser expresada en función del índice de no-linealidad como:

$$\frac{\partial f}{\partial Z}(z+\Delta z) \approx \left.\frac{\partial f}{\partial Z}\right|_z (1+L_f) \tag{9}$$

Observando la ecuación anterior, se pueden obtener dos conclusiones importantes:

1. Si el índice de no-linealidad L_f es igual a cero en un punto Z_i, esto implica que la función *f* es lineal en el intervalo ΔZ.
2. Si el índice de no-linealidad L_f toma valores distintos de cero, implica que la función *f* es una función no lineal en el intervalo ΔZ.

4.1 Estudio de la no-linealidad de la profundidad

Considerando un sistema estéreo ideal, la profundidad de un punto, puede ser determinada a partir de la siguiente ecuación:

$$Z = \frac{f_L}{d_x} \cdot \frac{B}{u_R - u_L} = f_x \cdot \frac{B}{d_u} \tag{10}$$

En la ecuación anterior f_x es la distancia focal horizontal de las cámaras en píxeles, d_u es la disparidad horizontal en píxeles y *B* es la línea de base del par estéreo, en nuestra aplicación de 15 cm. El índice de no-linealidad para la profundidad es función de la disparidad horizontal y se puede calcular como sigue:

$$L_Z = \left| \frac{\frac{\partial^2 Z}{\partial d_{u_i}^2} \cdot \Delta d_u}{\frac{\partial Z}{\partial du_i}} \right| \tag{11}$$

Si aislamos la disparidad horizontal d_u en la Ec. 10, podemos expresar el índice de no-linealidad como función de la profundidad:

$$L_Z = \frac{2 \cdot \Delta d_u}{d_u} = \frac{2 \cdot Z \cdot \Delta d_u}{f_x \cdot B} \tag{12}$$

4.2 Estudio de la no-linealidad angular

El índice de no-linealidad angular L_a es calculado considerando los ángulos de azimuth y elevación.

$$L_a = L_\theta + L_\phi = \left| \frac{\frac{\partial^2 \theta_i}{\partial z^2} \cdot \Delta z}{\frac{\partial \theta_i}{\partial z}} \right| + \left| \frac{\frac{\partial^2 \phi_i}{\partial z^2} \cdot \Delta z}{\frac{\partial \phi_i}{\partial z}} \right| \tag{13}$$

Las expresiones de los índices de no-linealidad para los ángulos de azimuth y elevación son respectivamente:

$$L_\phi = \frac{x^4 - 2 \cdot z^4 + x^2 \cdot \left(y^2 - z^2\right)}{z \cdot \left(x^2 + z^2\right) \cdot \left(x^2 + y^2 + z^2\right)} \cdot \Delta z \tag{14}$$

$$L_\theta = \frac{2 \cdot z}{x^2 \cdot \left(1 + \frac{z^2}{x^2}\right)} \cdot \Delta z \tag{15}$$

4.3 Umbral de profundidad óptimo

Se ha realizado una simulación en la cuál se han calculado los valores de los índices de no-linealidad para diferentes valores de la profundidad *Z*. La

Fig. 4 muestra los índices de no-linealidad angular y profundidad, considerando los siguientes intervalos $\Delta d_u = \pm 1$ píxeles y $\Delta Z = \pm 1$m.

Como se puede observar en la Fig. 4 ambos índices de no-linealidad son iguales para un único valor de profundidad Z_t = 5.71 m. Para valores de profundidad mayores que dicho umbral, la información angular es más lineal que la profundidad, y por lo tanto para este intervalo de distancias, es mejor utilizar una parametrización de profundidad inversa. Por el contrario, para valores de profundidad menores a dicho umbral es mejor utilizar una parametrización 3D de las marcas.

Por lo tanto, y acorde con los resultados, sugerimos utilizar un umbral de profundidad Z_t = 5.71 m como el umbral óptimo para conmutar entre ambos tipos de parametrizaciones. Nuestro resultado se aproxima bastante al obtenido por Lina *et al*. en (Paz, 2008), en dónde a partir de un análisis empírico sugieren utilizar un umbral de 5 m considerando una línea de base de 12 cm.

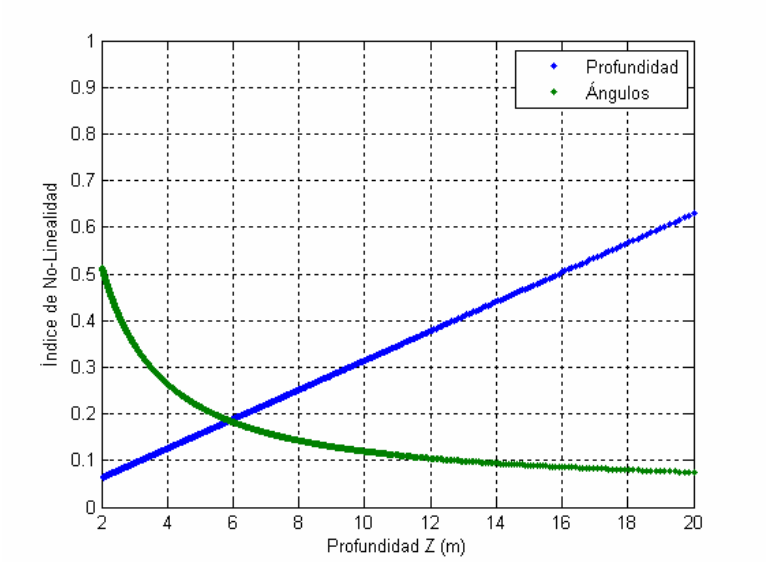

Fig. 4. No-Linealidad angular y en profundidad

5 Transformación de la apariencia del parche mediante homografía 2D

Cuando se va a realizar la medida de una marca, la posición y orientación de la cámara izquierda (accesibles en el vector de estado del EKF), y el vector normal del plano al cuál pertenece la marca son utilizados para transformar el valor inicial de apariencia de la plantilla 2D de la marca (debido a los cambios en el punto de vista de la cámara) mediante una homografía 2D. Nuestra propuesta se basa en los trabajos previos desarrollados en (Liang, 2002, Molton 2004).

Considerando dos sistemas de referencia genéricos, la transformación entre ambos sistemas X_1 y X_2 viene definida por la siguiente ecuación:

$$X_2 = R \cdot X_1 + T \tag{16}$$

en donde R y T son la matriz de rotación y el vector de traslación que definen la transformación relativa entre ambos sistemas de coordenadas. Sea X_1 un punto perteneciente al plano definido por la Ec. 17:

$$\pi : a \cdot x_1 + b \cdot y_1 + c \cdot z_1 + 1 = 0 \tag{17}$$

El plano anterior es un plano que no pasa por el origen de coordenadas, y el vector $n = (a, b, c)^t$ es el vector normal al plano. Según esto, se puede encontrar la siguiente relación:

$$n^t \cdot X_1 = -1 \tag{18}$$

Utilizando la ecuación anterior, la Ec. 16 se puede expresar de la siguiente manera:

$$X_2 = R \cdot X_1 - T \cdot n^t \cdot X_1 = \left(R - T \cdot n^t\right) \cdot X_1 \tag{19}$$

Y por lo tanto, las posiciones en el plano imagen para dos poses de cámara diferentes, quedan relacionadas por la siguiente homografía 2D:

$$U_2 = C_2 \cdot \left(R - T \cdot n^t\right) \cdot C_1^{-1} \cdot U_1 \tag{20}$$

La Fig. 5 muestra la información geométrica del par estéreo, así como la problemática de obtener el vector normal al plano, así como la homografía 2D para transformar la apariencia inicial del parche, utilizando la información de las dos cámaras.

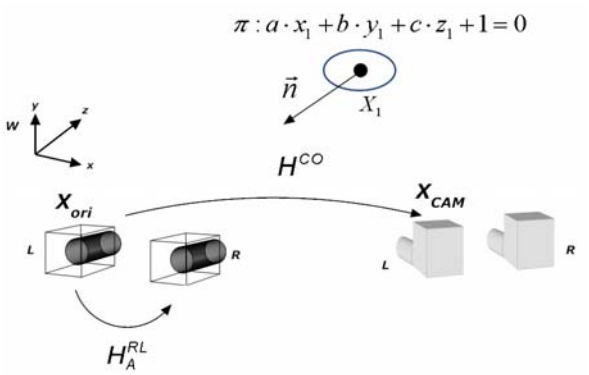

Fig. 5. Proceso de transformación de la apariencia del parche utilizando la información del par estéreo

La Eq. 21 muestra la relación existente entre los sistemas de coordenadas asociados a las cámaras izquierda y derecha.

$$U_R = C_R \cdot \left(R^{RL} - T^{RL} \cdot n^t\right) \cdot C_L^{-1} \cdot U_L \tag{21}$$

La ecuación anterior depende de la matriz de rotación R^{RL} y el vector de traslación T^{RL} existentes entre ambas cámaras. Los valores de la matriz de rotación y el vector de traslación son conocidos con gran exactitud, ya que han sido previamente calculados en el proceso de calibración del par estéreo. Suponiendo que existe una transformación afín entre los parches 2D correspondientes a la imagen izquierda y derecha, la transformación afín H_A^{RL} puede ser expresada como:

$$H_A^{RL} = C_R \cdot \left(R^{RL} - T^{RL} \cdot n^t\right) \cdot C_L^{-1} \tag{22}$$

Esta transformación afín puede ser calculada fácilmente a partir de 3 correspondencias de puntos no-colineales y bajo la aproximación de parches localmente planos. Como se puede observar, la Ec. 22 depende del vector normal al plano *n*. A partir de la Ec. 22 se puede despejar el producto $T^{RL} \cdot n^t$. Denominando a este producto X, su expresión es la siguiente:

$$X = T^{RL} \cdot n^t = R^{RL} - C_R^{-1} \cdot H_A^{RL} \cdot C_L \tag{23}$$

Todos los parámetros de la ecuación anterior son conocidos, ya que la transformación afín, H_A^{RL}, previamente calculada, y el resto de matrices implicadas son conocidas gracias al proceso previo de calibración estéreo.

Por lo tanto, se obtiene un sistema de 9 ecuaciones y 3 incógnitas, que son las componentes del vector normal al plano:

$$\left.\begin{array}{lll} n_x = \dfrac{X_{11}}{T_x} & n_x = \dfrac{X_{21}}{T_y} & n_x = \dfrac{X_{31}}{T_z} \\ n_y = \dfrac{X_{12}}{T_x} & n_y = \dfrac{X_{22}}{T_y} & n_y = \dfrac{X_{32}}{T_z} \\ n_z = \dfrac{X_{13}}{T_x} & n_z = \dfrac{X_{23}}{T_y} & n_z = \dfrac{X_{33}}{T_z} \end{array}\right\} \quad (24)$$

En el momento de la inicialización de una nueva marca, el vector normal al plano se calcula de la manera que se ha explicado. Una vez que se ha estimado este vector normal, la homografía 2D entre los distintos puntos de vista de cámara, puede ser determinada utilizando la estimación actual de la posición y orientación de la cámara izquierda con respecto a la posición y orientación de la cámara izquierda en el momento de inicialización de la marca:

$$U_{CAM} = C_L \cdot \left(R^{CO} - T^{CO} \cdot n^t\right) \cdot C_L^{-1} \cdot U_{ORI} \quad (25)$$

En la ecuación anterior R^{CO} y T^{CO} son las matrices de rotación y traslación entre la pose actual de la cámara izquierda y la pose de referencia en el que la marca fue inicializada.

6 Experimentos en Interiores

Para poder evaluar el funcionamiento del sistema, se han realizado numerosas pruebas en entornos de interiores. En este trabajo, se presentan solamente los resultados de 3 secuencias. Las cámaras utilizadas fueron del tipo Unibrain Fire-i IEEE1394 con lentes adicionales de gran angular de 1.9 mm que proporcionan un campo de vista de alrededor de 100º en horizontal y en vertical. Las calibraciones para cada una de las cámaras son calculadas en un proceso *offline* siguiendo el proceso descrito en (Heikkila, 1997). La línea de base del par estéreo es de 15 cm, la resolución de las imágenes procesadas de 320x240 píxeles, utilizando únicamente imágenes en blanco y negro. El *frame rate* de adquisición utilizado es de 30 *frames/seg*. Las pruebas experimentales han sido llevadas a cabo en un ordenador portátil con procesador Intel Core 2 Duo a 2.4 GHz. El algoritmo de SLAM Visual ha sido implementado en C/C++ y

es capaz de funcionar en tiempo real bajo entornos pequeños de aproximadamente 150 marcas.

La primera secuencia se trata de un pasillo de una longitud de 10 m, en la cuál la cámara se mueve siguiendo una trayectoria recta desplazada hacia la izquierda. Este es un escenario propicio para comprobar el funcionamiento de la parametrización de profundidad inversa, ya que se pueden encontrar muchas marcas lejanas. Se ha realizado una comparación considerando los dos tipos de parametrizaciones de marcas. Se han estudiado tres casos: sin parametrización de profundidad inversa, es decir únicamente con parametrización 3D, y combinando ambas parametrizaciones utilizando dos umbrales diferentes de profundidad Z_t = 10 m y Z_t = 5.7 m. La segunda de las secuencias es una trayectoria en forma de L de dimensiones aproximadas 3 m en el eje X y 6 m de largo en el eje Z. Finalmente, la última de las secuencias es un bucle de dimensiones 4.8 m a lo largo del eje X y 6 m a lo largo del eje Z. La Fig. 6 muestra el tamaño del vector de estado a lo largo del tiempo correspondiente a la secuencia con trayectoria en forma de L. Como se puede observar, el tamaño del vector de estado en el caso de considerar una parametrización de profundidad inversa y un umbral Z_t = 5.7 m es mayor que en el resto de experimentos, debido al coste computacional extra introducido al utilizar este tipo de parametrización.

El mapa y trayectorias finales de las secuencias de L y bucle se muestran en la Fig. 7. La Tabla 1 muestra los resultados de la comparativa entre los diferentes casos analizados. El significado de los parámetros en la tabla son los siguientes:

- % Marcas Inversas: Este porcentaje es un indicador del número de marcas que han sido inicializadas con una parametrización de profundidad inversa con respecto al número total de marcas que constituyen el mapa final.
- ε_i: Es el error absoluto medio en m, para las coordenadas cartesianas (X, Z).
- Traza Media P_{YY}: Es el valor de la traza media de la matriz de covarianza P_{YY} de todas las marcas que componen el mapa final. Este parámetro es indicativo de la incertidumbre existente en la posición de las marcas, es decir, de la calidad final del mapa en términos de incertidumbre en la posición 3D de las marcas.

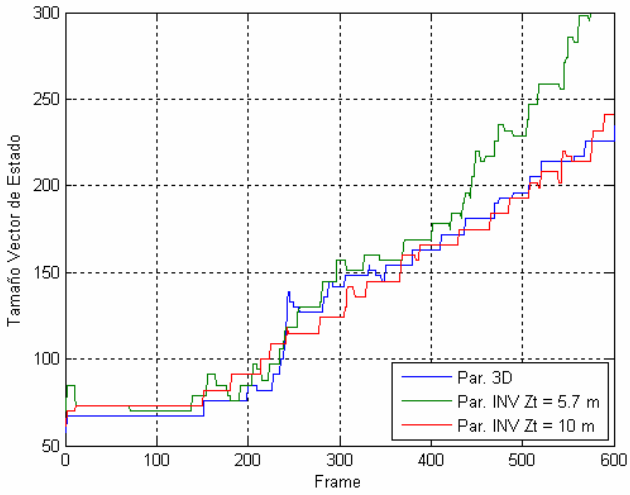

Fig. 6. Comparativa del tamaño del vector de estado

Secuencia	**Caso**	**%Marcas Inversas**	**ε_X (m)**	**ε_Y (m)**	**Traza Media P_{YY}**
Pasillo	Sin Par. Inversa	0.00	0.9394	0.4217	0.1351
Pasillo	Par. Inversa Zt = 10 m	5.23	0.9259	0.4647	0.0275
Pasillo	Par. Inversa Zt = 5.7 m	24.32	0.7574	0.3777	0.0072
L	Sin Par. Inversa	0.00	0.5047	0.3985	0.1852
L	Par. Inversa Zt = 10 m	7.85	0.5523	0.1017	0.0245
L	Par. Inversa Zt = 5.7 m	19.21	0.5534	0.2135	0.0078
Bucle	Sin Par. Inversa	0.00	0.4066	0.9801	0.2593
Bucle	Par. Inversa Zt = 10 m	5.27	0.3829	0.6303	0.0472
Bucle	Par. Inversa Zt = 5.7 m	12.36	0.2191	0.3778	0.0310

Tabla 1. Comparativa entre parametrización 3D y de profundidad inversa: errores de trayectoria absolutos e incertidumbre en el mapa

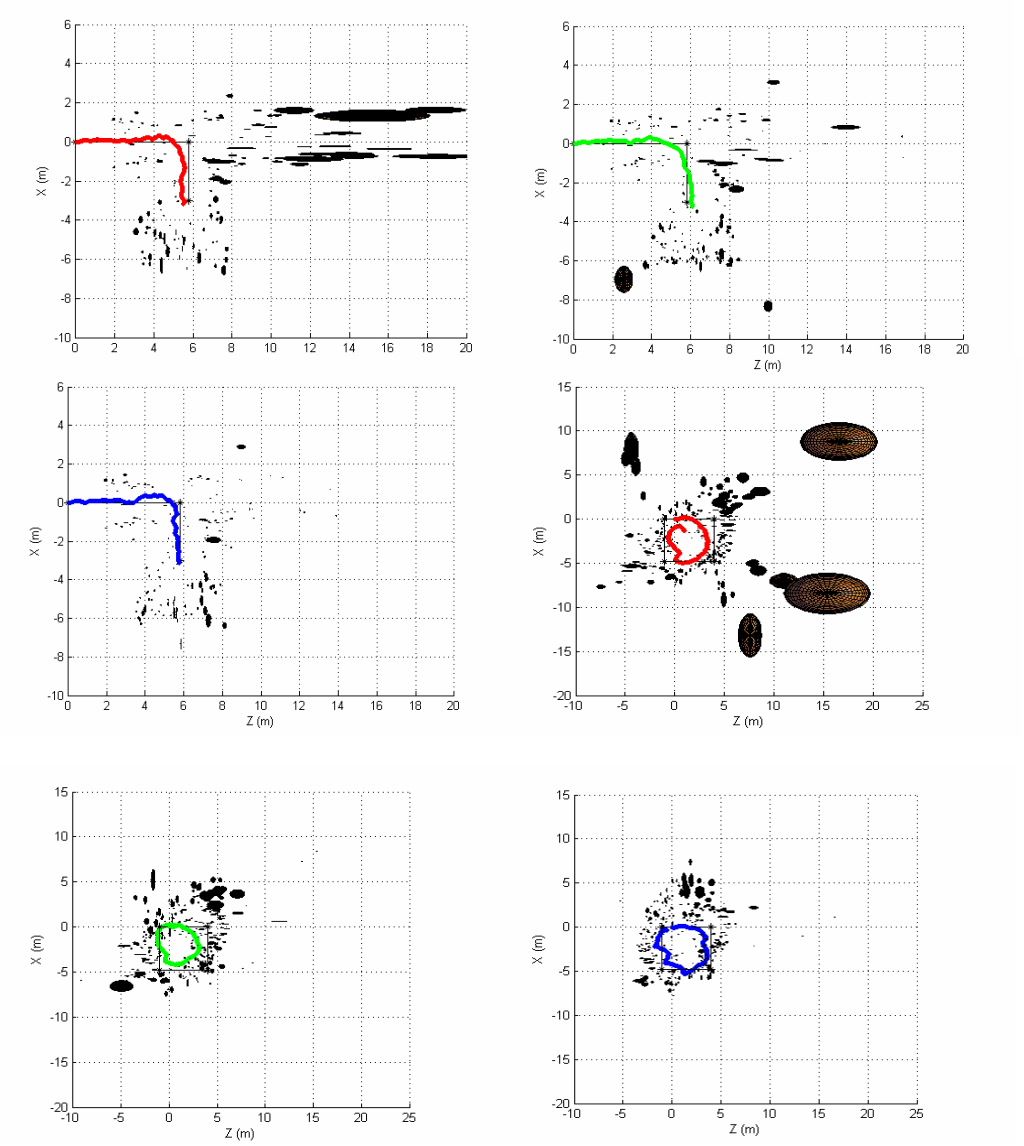

Fig. 7. Comparativa trayectoria y mapa final de la cámara. De arriba abajo y de izquierda a derecha. a) L sin par. de profundidad inversa. b) L con par. de profundidad inversa Zt = 10 m. c) L con par. de profundidad inversa Z_t = 5.7 m. d) Bucle sin par. de profundidad inversa. e) Bucle con par. de profanidad inversa Z_t = 10 m. f) Bucle con par. de profundidad inversa Z_t = 5.7 m.

7 Conclusiones y Trabajos Futuros

En este trabajo se ha presentado un sistema que permite obtener la localización de una cámara estéreo movida con la mano, a la vez que se obtiene un mapa de marcas 3D del entorno. Una de las contribuciones del presente trabajo, es la obtención de un umbral de profundidad para conmutar entre los distintos tipos de parametrizaciones de marcas mediante un estudio de no-linealidad. Además, se han demostrado los beneficios de utilizar una parametrización de profundidad inversa en el caso de visión estereoscópica para poder trabajar de una manera más eficiente con las marcas lejanas. Sin embargo, dependiendo del tipo de aplicación (tipo de entorno y posibles restricciones de tiempo real), la sobrecarga introducida por el uso de una parametrización de profundidad inversa utilizando el umbral de profundidad óptimo, puede reducirse utilizando umbrales de profundidad mayores, siempre que la calidad del mapa final no se vea afectada. De acuerdo con los resultados de la Tabla 1, la simulación considerando la parametrización de profundidad inversa con el umbral óptimo ha resultado ser la que presenta una menor incertidumbre en el mapa final y menores errores absolutos. Además, en la secuencia del bucle, no se logra cerrar el lazo utilizando únicamente una parametrización 3D.

Como se ha comentado anteriormente, el utilizar el umbral óptimo de profundidad puede exceder en determinadas ocasiones las restricciones de tiempo real debido al coste computacional extra asociado a dicha parametrización. Estamos muy interesados en la posibilidad de utilizar un umbral de profundidad dinámico en función del tipo de entorno, en vez del umbral estático que se usa actualmente, con el fin de obtener la misma calidad final del mapa manteniendo las restricciones de tiempo real.

Considerando parches 2D y el vector normal al plano que contiene a la marca mejora el seguimiento de las marcas, permitiendo transformar la apariencia del parche inicial de la marca ante distintos puntos de vista de la cámara. Sin embargo y debido a que actualmente el vector normal se estima una única vez por marca, una actualización de este vector podría ser de gran interés para trabajos futuros.

Como trabajo futuro, se desarrollará un SLAM de alto nivel que permita trabajar con grandes entornos tanto en interiores como en exteriores. Además, se planea realizar la fusión del sistema estéreo con diversos sistemas inerciales tales como acelerómetros y/o GPS para los

experimentos en exteriores. También se plantearán alternativas para sustituir al modelo de movimiento actual (modelo impulsivo), debido a la gran variabilidad de movimientos que puede presentar una persona invidente que transporte el sistema.

Finalmente, debido a que estamos interesados en la aplicación de técnicas de SLAM visual como ayuda a la navegación de personas invidentes, se desea obtener cierta realimentación de organizaciones de personas invidentes así como la realización de pruebas experimentales con los usuarios finales del sistema.

Agradecimientos

Los autores desean expresar su gratitud a la Comunidad de Madrid por su financiación a través del proyecto RoboCity2030 (CAM-S-0505/DPI/000176) y al Ministerio de Ciencia e Innovación por su financiación a través del proyecto DRIVER-ALERT (TRA2008-03600/AUT).

Referencias

Mountney P., Stoyanov D., Davison A. J. y Yang G. Z. 2006. *Simultaneous stereoscope localization and soft-tissue mapping for minimally invasive surgery. In Proceedings of the Medical Image Computing and Computer Assisted Intervention*.

Klein G. y Murray D. 2007. *Parrallel tracking and mapping for small AR workspaces In Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality*.

Schleicher D., Bergasa L. M., Barea R., López E., Ocaña M. y Nuevo J. 2007. *Real-time wide-angle stereo visual SLAM on large environments using SIFT features correction. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems*.

Schleicher D., Bergasa L.M., Barea R., López E. y Ocaña M. 2006. *Real-time Simultaneous Localization and Mapping with a wide-angle stereo camera and adaptive patches. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems*.

Min Oh, S., Tariq S., Walker B. y Dellaert F. 2004. *Map-based Priors for Localization*. *In Proceedings of the IEEE International Conference on Intelligent Robots and Systems*.

Davison A. J., Reid I. D., Molton N. D. y Stasse O. 2007. *MonoSLAM: Real-time single camera SLAM. IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 29, no. 6*.

Harris C. y Stephens M. 1988. *A combined corner and edge detector. In Proceedings of Fourth Alvey Vision Conference*.

Paz L. M., Piniés P., Tardós J. D. y Neira J. 2008. *Large scale 6DOF SLAM with stereo in hand. IEEE Transactions on Robotics, vol 24, no. 5*.

Civera J., Davison A. J. y Montiel J. M. 2008. *Inverse depth parametrization for monocular SLAM. IEEE Transactions on Robotics*, vol 24, no 5.

Paz L. M., Guivant J., Tardós J. D. y Neira J. 2007. *Data association in O(n) for divide and conquer SLAM. In Proceedings of Robotics Science and Systems*.

Faugeras O., Hotz B., Mathieu H., Vieville T., Zhang Z., Pascal F., Theron E., Moll L., Berry G., Vuillemin J., Bertin P. y Proy C. 1993. *Real time correlation-based stereo: Algorithm, implementations and applications. INRIA Technical Report*.

Liang B. y Pears N. 2002. *Visual navigation using planar homographies. In Proceedings of International Conference on Robotics and Automation*.

Molton N., Davison A. J. y Reid I. 2004. *Locally planar patch features for real-time structure from motion. In Proceeding of British Machine Vision Conference*.

Heikkila y Silven A. 1997. *A Four-Step Camera Calibration Procedure with Implicit Image Correction. In Proceedings of Computer Vision and Pattern Recognition*.