

# UNA FORMULACIÓN CERRADA PARA ODOMETRÍA VISUAL PROBABILÍSTICA EN 6D CON UNA CÁMARA ESTÉREO

F. A. Moreno, J. L. Blanco y J. González  
Departamento de Ingeniería de Sistemas y Automática.  
Universidad de Málaga.  
famoreno@isa.uma.es, {jlblanco, jgonzalez}@ctima.uma.es

## Resumen

*Tradicionalmente, la estimación del movimiento de un robot se ha llevado a cabo mediante el uso de sistemas odométricos basados en codificadores. Sin embargo, estos sistemas sufren ciertos problemas como la existencia de errores acumulativos, posibles derrapes de las ruedas o la restricción a movimientos sobre un plano. En este trabajo presentamos un método alternativo para estimar el cambio incremental en la posición y orientación de un robot usando imágenes estéreo. A diferencia de otras propuestas para realizar odometría visual en 6D, basadas en métodos iterativos, proponemos utilizar una formulación cerrada y óptima, más precisa y eficiente, y que evita problemas de convergencia. Además, derivamos la expresión de la covarianza asociada a esta estimación, de manera que ésta pueda ser integrada en métodos de SLAM visual probabilístico. Nuestra propuesta combina la gran capacidad de distinción de los descriptores SIFT con la velocidad del método de seguimiento KLT, consiguiendo de esta manera una ejecución eficiente y robusta en tiempo real. Nuestro método ha sido validado mediante la serie de pruebas experimentales que se presentan.*

**Palabras Clave:** Odometría visual, sistema estéreo, formulación cerrada, SIFT, KLT.

## 1 INTRODUCCIÓN

La odometría es uno de los sistemas más extendidos para la estimación del desplazamiento de un robot móvil. Tradicionalmente, la odometría se ha basado en codificadores que miden las revoluciones de las ruedas del robot, proporcionando así información para estimar cambios incrementales en la posición y la orientación del robot (en este artículo denominaremos *pose* al conjunto de la posición y la orientación del robot). Por desgracia, el uso de la odometría basada en codificadores está limitado a robots con ruedas que se mueven por superficies planas, y los errores sistemáticos, el derrape de las ruedas y las diferencias físicas entre las ruedas del

robot producen errores incrementales en la estimación del desplazamiento que no pueden ser modelados, como es habitual, por una distribución Gaussiana de media cero. Sin embargo, este modelo erróneo es asumido en muchos filtros probabilísticos usados para la localización de robots y la realización de SLAM [15], pudiendo llevar en última instancia a la divergencia del filtro.

Para superar estas limitaciones de la odometría basada en codificadores, en los últimos años se han usado otros tipos de sensores no propioceptivos como escáneres láser [4], [14] y, más recientemente, sistemas basados en visión por computador [1], [16]. El uso de escáneres también se ve restringido a movimientos puramente planos, mientras que la odometría basada en visión explota las ventajas del mayor campo de visión de las cámaras. Hoy en día, las cámaras son sensores baratos capaces de adquirir una enorme cantidad de información del entorno. La existencia de métodos eficaces capaces de extraer y seguir características en las imágenes, junto con las ventajas de las cámaras anteriormente mencionadas, establecen un marco de trabajo adecuado para aplicar la visión a la estimación del movimiento.

Relacionados con este tema, podemos encontrar en la literatura técnica una cantidad considerable de propuestas que aplican diferentes métodos para estimar el desplazamiento de un robot móvil equipado con un sistema de visión a partir de las imágenes capturadas durante su navegación por el entorno. En [10] se propone un sistema de odometría visual tanto monocular como estéreo basado en métodos iterativos para estimar el cambio de la *pose* del robot en 3D, mientras que [1] realiza odometría visual monocular con cámaras de bajo coste sin calibrar bajo la asunción de movimiento puramente plano. En [13] se propone un método probabilístico para realizar SLAM que utiliza odometría visual como modelo de movimiento. Dicho método busca características en imágenes estéreo y les calcula sus descriptores SIFT para establecer correspondencias entre ellos. El movimiento de la cámara es estimado posteriormente usando un algoritmo de optimización iterativo que minimiza el error de reproyección de los puntos tridimensionales en las imágenes.

En este trabajo, proponemos un nuevo enfoque para la odometría visual que estima los cambios incrementales en la *pose* 6D ( $x, y, z$ , desviación, elevación y rotación) del robot entre pares de imágenes estéreo consecutivos. Nuestro método estima el conjunto completo de traslaciones y ángulos, sin asumir ningún modelo de movimiento concreto. Por lo tanto, no existen restricciones en el tipo de movimientos que puede realizar la cámara tal y como ocurre en otros métodos como el descrito en [2].

Nuestro sistema combina la velocidad de detección y seguimiento de características del método de Kanade-Lucas-Tomasi [12] con el poder de distinción de los descriptores SIFT [8] para emparejar características en las imágenes estéreo. Al llevar a cabo el proceso de emparejamiento estéreo basado en los SIFT tan sólo cuando el número de puntos característicos seguidos cae por debajo de un umbral, evitamos el alto coste computacional que implica el calcular y comparar la distancia Euclídea entre los descriptores SIFT de todas las características para cada par de imágenes estéreo. Otra ventaja de nuestro método sobre otros enfoques es la aplicación de una formulación cerrada para estimar los cambios en la orientación y la traslación, evitando así tanto la complejidad como los problemas de divergencia asociados a los métodos iterativos. Además, modelamos la incertidumbre de la estimación del cambio de *pose* mediante la propagación de la incertidumbre en las posiciones 3D

de los puntos observados.

El resto del artículo se organiza de la siguiente manera: la sección 2 presenta un resumen del método propuesto, siendo descrito con mayor detalle en la sección 3. En la sección 4 mostramos algunos resultados experimentales mientras que la sección 5 presenta algunas conclusiones y líneas de trabajo futuro.

## 2 DESCRIPCIÓN GENERAL

Nuestro método, esquematizado en la Figura 1, se puede resumir en las siguientes etapas:

1. Búsqueda de un conjunto de puntos característicos en el primer par de imágenes estéreo, y cálculo de sus descriptores SIFT.
2. Emparejamiento estéreo basado en la distancia Euclídea entre los descriptores y las restricciones de la geometría epipolar.
3. Proyección al espacio tridimensional de las características emparejadas de manera que se obtenga un conjunto de puntos 3D con coordenadas relativas al robot.
4. Seguimiento de las características en el siguiente par de imágenes estéreo. Nótese que este proceso de seguimiento evita un nuevo proceso de

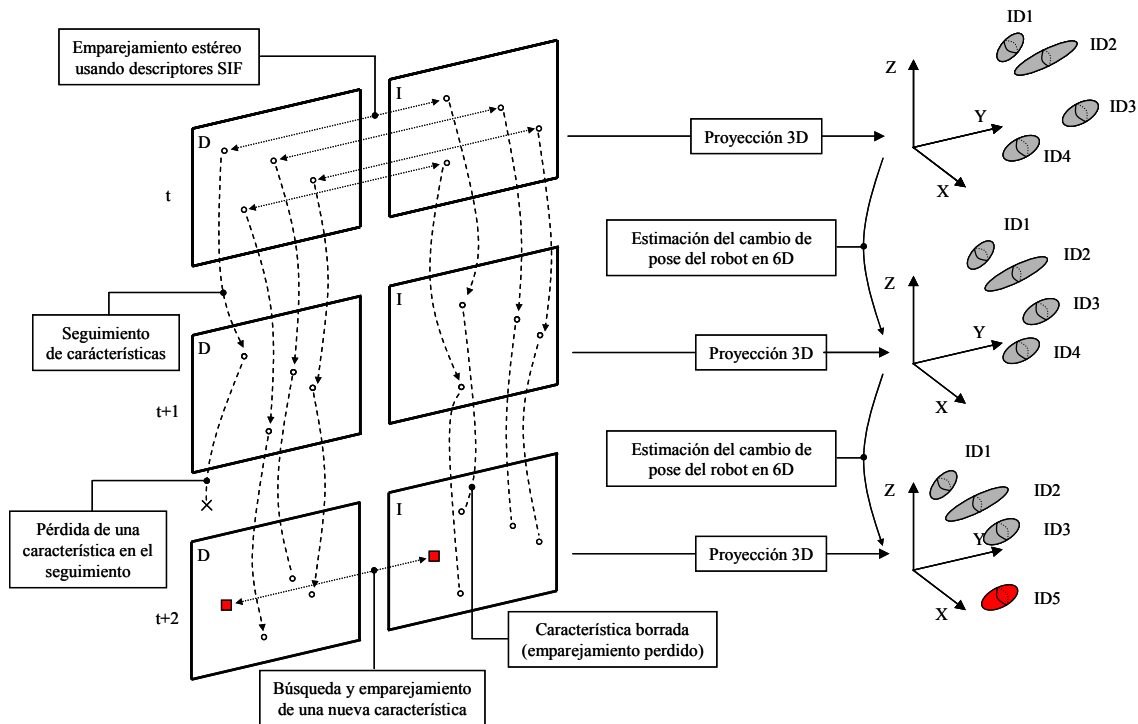


Figura 1. Esquema del procedimiento de odometría visual propuesto.

emparejamiento basado en SIFTs.

5. Proyección al espacio 3D de las características que han sido seguidas, obteniéndose un nuevo conjunto de puntos cuyas correspondencias con el anterior conjunto de puntos 3D son conocidas.
6. Estimación de la *pose* en 6D de la cámara (y, por lo tanto, del robot) mediante una formulación cerrada [6], dadas las correspondencias entre los dos conjuntos de puntos 3D.
7. Si el número de características seguidas cae por debajo de un umbral, se realiza una nueva búsqueda de puntos en las imágenes estéreo y se calculan sus descriptores SIFT. Posteriormente se emparejan usando los descriptores y se añaden al conjunto actual de puntos.
8. Repetir desde la etapa 4.

La siguiente sección muestra una descripción más detallada de todos los pasos del método propuesto.

### 3 DESCRIPCIÓN DETALLADA DEL MÉTODO

Esta sección presenta una descripción más detallada de las diferentes operaciones que son necesarias para realizar odometría visual con nuestro método.

#### 3.1 EXTRACCIÓN Y EMPAREJAMIENTO DE CARACTERÍSTICAS ADECUADAS EN IMÁGENES ESTÉREO

En la literatura técnica se han propuesto muchos métodos para extraer puntos de interés de las imágenes, como los conocidos detectores de Kitchen & Rosenfeld [7] y Harris [5], basados en la primera y segunda derivada de la imagen, respectivamente. Más recientemente, el detector SIFT propuesto por Lowe [8] aborda este tema mediante la identificación de extremos locales en una pirámide de Diferencias de Gaussianas. Además, asocia a las características detectadas un descriptor que presenta invarianza a la rotación y a la escala, e invarianza parcial a cambios en la iluminación y distorsiones afines. En nuestro trabajo, la detección de puntos de interés en las imágenes se lleva a cabo mediante el método propuesto por Shi y Tomasi [12]. Además, se calculan sus correspondientes descriptores SIFT para conseguir que los puntos sean suficientemente distinguibles y se mejore la robustez del proceso de emparejamiento.

Una vez que los puntos han sido detectados en las imágenes izquierda y derecha, estos son emparejados de forma robusta según la similitud entre sus

descriptores y las restricciones impuestas por la geometría epipolar del sistema de visión. De forma más detallada, para cada punto de interés de la imagen izquierda se calcula la distancia Euclídea entre su descriptor y los de los puntos de interés de la imagen derecha. La primera restricción para que un par de puntos sea considerado un candidato a emparejamiento es que sus descriptores cumplan: a) ser suficientemente parecidos (su distancia debe estar por debajo de un determinado umbral) y b) ser suficientemente distintos a otros candidatos (con distancia mayor de otro umbral). Además, los puntos deben cumplir la restricción epipolar, es decir, deben caer en las líneas epipolares conjugadas (o estar suficientemente cerca de ellas). En un sistema de visión estéreo con ejes ópticos paralelos como el usado en nuestro trabajo, las líneas epipolares son paralelas y horizontales, por lo que la restricción epipolar se reduce a la comprobación de que ambos puntos están en la misma fila de píxeles (con una cierta tolerancia).

Finalmente, a cada par de puntos emparejados se le asigna un identificador (ID) único que será utilizado para distinguir el punto 3D que al proyectarse en las imágenes produjo los puntos emparejados.

#### 3.2 PROYECCIÓN AL ESPACIO 3D

Una vez que los puntos de interés han sido emparejados de forma robusta, el siguiente paso es determinar las coordenadas de los correspondientes puntos 3D a partir de las coordenadas de los puntos en las imágenes y los parámetros intrínsecos del sistema estéreo. Formalmente, sean  $(c, r)$  las coordenadas de un punto en la imagen izquierda (que será tomada como imagen de referencia en nuestro trabajo) y  $d$  la disparidad de su punto emparejado en la imagen derecha. Así, las coordenadas 3D  $(X, Y, Z)$  del punto proyectado se calculan mediante:

$$X = (c - c_0) \frac{b}{d} \quad Y = (r - r_0) \frac{b}{d} \quad Z = f \frac{b}{d} \quad (1)$$

donde  $(c_0, r_0)$  son las coordenadas en la imagen del punto principal de la imagen de referencia,  $b$  es la línea base (distancia entre los centros ópticos de las cámaras) y  $f$  representa la distancia focal de las cámaras (idéntica para ambas).

Los errores cometidos en el cálculo de las variables  $r$ ,  $c$  y  $d$  se modelan normalmente como variables Gaussianas de media cero e incorreladas entre sí [9]. Usando propagación de errores de primer orden para aproximar la distribución de las variables  $X$ ,  $Y$  y  $Z$  en (1) como Gaussianas multivariantes, obtenemos la siguiente expresión de la matriz de covarianza:

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{pmatrix} = \mathbf{J} \text{diag}(\sigma_c^2, \sigma_r^2, \sigma_d^2) \mathbf{J}^T \quad (2)$$

donde  $\mathbf{J}$  representa la matriz Jacobiana de las funciones mostradas en (1) y  $\sigma_X^2, \sigma_Y^2, \sigma_Z^2, \sigma_c^2, \sigma_r^2$ , y  $\sigma_d^2$  son las varianzas de las variables correspondientes. Expandiendo (2) llegamos a la siguiente expresión de  $\Sigma$ :

$$\Sigma = \begin{pmatrix} \frac{b^2\sigma_c^2}{d^2} + \frac{b^2(c-c_0)^2\sigma_d^2}{d^4} & \frac{(c-c_0)b^2\sigma_d^2(r-r_0)}{d^4} & \frac{(c-c_0)b^2\sigma_d^2f}{d^4} \\ \frac{(c-c_0)b^2\sigma_d^2(r-r_0)}{d^4} & \frac{b^2\sigma_r^2}{d^2} + \frac{b^2(r-r_0)^2\sigma_d^2}{d^4} & \frac{(r-r_0)b^2\sigma_d^2f}{d^4} \\ \frac{(c-c_0)b^2\sigma_d^2f}{d^4} & \frac{(r-r_0)b^2\sigma_d^2f}{d^4} & \frac{f^2b^2\sigma_d^2}{d^4} \end{pmatrix} \quad (3)$$

la cual modela aproximadamente la incertidumbre de las coordenadas 3D de los puntos calculados a partir de medidas ruidosas en un sistema de visión estéreo.

Finalmente, para distinguir los puntos proyectados, a cada punto 3D se le asigna el mismo ID asociado al emparejamiento del que fue generado.

### 3.3 SEGUIMIENTO DE PUNTOS

Los puntos detectados y emparejados son seguidos en imágenes estéreo consecutivas usando el conocido método de Kanade-Lucas-Tomasi [12] de manera que se obtengan sus coordenadas en el siguiente par de imágenes. Este algoritmo calcula el flujo óptico de un punto característico en dos imágenes consecutivas minimizando la diferencia entre las ventanas de píxeles centradas en el punto de interés usando un método de Newton-Raphson.

El correcto seguimiento de un par de puntos emparejados en las imágenes izquierda y derecha en el instante  $k$  produce un nuevo par de puntos emparejados en el instante  $k+1$ . Para evitar la presencia de emparejamientos no fiables, se hace uso de la restricción epipolar, detectando así puntos que no han sido seguidos adecuadamente. Usando este método de seguimiento eludimos tanto la búsqueda de características como el emparejamiento estéreo basado en los SIFT para cada par de imágenes. De esta forma, nuestro método acelera el proceso de extraer y emparejar puntos, reduciendo considerablemente el coste computacional del proceso global de odometría visual.

El conjunto resultante de puntos seguidos es proyectado al espacio tridimensional mediante el método descrito en la sección 3.2, generando un nuevo conjunto de puntos 3D que conservan los IDs de sus puntos asociados en las imágenes, de manera

que se mantengan unas relaciones de emparejamiento implícitas con los puntos 3D del conjunto del instante de tiempo anterior.

Si el número de puntos seguidos cae por debajo de un determinado umbral, el algoritmo busca nuevas características en las imágenes para mantener un número adecuado de elementos en los conjuntos de puntos 3D.

### 3.4 ESTIMACIÓN PROBABILÍSTICA DEL CAMBIO DE POSE

En esta sección presentamos un método para estimar la distribución de probabilidad del cambio de *pose* del robot entre dos instantes de tiempo consecutivos a partir de dos conjuntos de puntos 3D conseguidos tal y como se ha descrito en las secciones anteriores.

Formalmente, sea  $\mathbf{X}_k$  un conjunto de puntos 3D obtenidos en el instante  $k$ :

$$\mathbf{X}_k = \{\mathbf{X}_k^i\}_{i=1\dots N_k} \quad (4)$$

donde la posición de cada punto 3D  $\mathbf{X}_k^i$  se asume sigue una distribución Gaussiana con media  $\mu_{\mathbf{X}_k^i} = \langle X_k^i, Y_k^i, Z_k^i \rangle$  y covarianza  $\Sigma_{\mathbf{X}_k^i}$  determinada por las ecuaciones (1) y (3) respectivamente:

$$\mathbf{X}_k^i \sim N(\mu_{\mathbf{X}_k^i}, \Sigma_{\mathbf{X}_k^i}) \quad (5)$$

Por otro lado, definimos la variable aleatoria  $q_{k,k+1}$  que modela el cambio de *pose* entre los instantes de tiempo  $k$  y  $k+1$  como una función de los conjuntos de puntos proyectados ( $\mathbf{X}_k$  y  $\mathbf{X}_{k+1}$ ):

$$q_{k,k+1} = f(\mathbf{X}_k, \mathbf{X}_{k+1}); \quad q_{k,k+1} \sim N(\mu_q, \Sigma_q) \quad (6)$$

Siguiendo una aproximación lineal de propagación del error,  $q_{k,k+1}$  viene modelada por una distribución Gaussiana con matriz de covarianza  $\Sigma_q$  y media  $\mu_q = \langle \Delta x, \Delta y, \Delta z, \Delta \alpha, \Delta \beta, \Delta \gamma \rangle$  donde  $\Delta x$ ,  $\Delta y$  y  $\Delta z$  son los incrementos en las coordenadas  $X$ ,  $Y$  y  $Z$  respectivamente, y  $\Delta \alpha$ ,  $\Delta \beta$  y  $\Delta \gamma$  representa los incrementos en los ángulos de desviación, elevación y rotación, respectivamente.

#### 3.4.1 Estimación de la media $\mu_q$

En este trabajo, proponemos calcular  $\mu_q$  mediante el método propuesto por Horn en [6], donde se deriva una solución de formulación cerrada al problema de mínimos cuadrados de encontrar la relación entre dos sistemas de coordenadas usando las medidas de las coordenadas de un conjunto de puntos en ambos sistemas. Usaremos los valores medios  $\mu_{X_k^i}$  de las

posiciones de los puntos 3D como entradas al algoritmo. Esta solución de formulación cerrada contrasta con otras propuestas para realizar odometría visual basada en métodos iterativos [11], [13] que necesitan una estimación inicial del cambio de *pose*.

El algoritmo de formulación cerrada se compone de los siguientes pasos:

1. Cálculo de los centroides ( $\mathbf{c}_k$  y  $\mathbf{c}_{k+1}$ ) de los dos conjuntos de puntos y resta de sus coordenadas para manejar únicamente puntos relativos a sus centroides:

$$\begin{aligned}\bar{\mathbf{X}}_k^i &= \langle \bar{X}_k^i, \bar{Y}_k^i, \bar{Z}_k^i \rangle \\ \bar{\mathbf{X}}_{k+1}^i &= \langle \bar{X}_{k+1}^i, \bar{Y}_{k+1}^i, \bar{Z}_{k+1}^i \rangle\end{aligned}\quad (7)$$

2. Para el punto 3D  $i$ -ésimo, cálculo de los siguientes nueve productos de sus coordenadas en los instantes  $k$  y  $k+1$ :

$$\begin{aligned}P_{XX}^i &= \bar{X}_k^i \bar{X}_{k+1}^i & P_{XY}^i &= \bar{X}_k^i \bar{Y}_{k+1}^i & P_{XZ}^i &= \bar{X}_k^i \bar{Z}_{k+1}^i \\ P_{YX}^i &= \bar{Y}_k^i \bar{X}_{k+1}^i & P_{YY}^i &= \bar{Y}_k^i \bar{Y}_{k+1}^i & P_{YZ}^i &= \bar{Y}_k^i \bar{Z}_{k+1}^i \\ P_{ZX}^i &= \bar{Z}_k^i \bar{X}_{k+1}^i & P_{ZY}^i &= \bar{Z}_k^i \bar{Y}_{k+1}^i & P_{ZZ}^i &= \bar{Z}_k^i \bar{Z}_{k+1}^i\end{aligned}\quad (8)$$

3. Suma de los productos en (8) para obtener los nueve valores siguientes:

$$\begin{aligned}S_{XX} &= \sum_i P_{XX}^i & S_{XY} &= \sum_i P_{XY}^i & S_{XZ} &= \sum_i P_{XZ}^i \\ S_{YX} &= \sum_i P_{YX}^i & S_{YY} &= \sum_i P_{YY}^i & S_{YZ} &= \sum_i P_{YZ}^i \\ S_{ZX} &= \sum_i P_{ZX}^i & S_{ZY} &= \sum_i P_{ZY}^i & S_{ZZ} &= \sum_i P_{ZZ}^i\end{aligned}\quad (9)$$

4. Formación de una matriz de dimensión 4x4 con los elementos en (9):

$$\mathbf{N} = \begin{bmatrix} S_{xx} + S_{yy} + S_{zz} & S_{yz} - S_{zy} & S_{zx} - S_{xz} & S_{xy} - S_{yx} \\ S_{yz} - S_{zy} & S_{xx} - S_{yy} - S_{zz} & S_{xy} + S_{yx} & S_{zx} + S_{xz} \\ S_{zx} - S_{xz} & S_{xy} + S_{yx} & -S_{xx} + S_{yy} - S_{zz} & S_{yz} + S_{zy} \\ S_{xy} - S_{yx} & S_{zx} + S_{xz} & S_{yz} + S_{zy} & -S_{xx} - S_{yy} + S_{zz} \end{bmatrix}\quad (10)$$

5. Búsqueda del autovector correspondiente al mayor autovalor de  $\mathbf{N}$ . Este autovector, interpretado como un cuaternión, determina el incremento en rotación entre los instantes  $k$  y  $k+1$ .

6. Cálculo de la matriz de rotación  $\mathbf{R}$  asociada a dicho cuaternión y cálculo de la traslación  $\mathbf{t} = (\Delta x, \Delta y, \Delta z)^T$  como la diferencia entre el centroide en el instante  $k$  y el centroide (escalado y rotado) en el instante  $k+1$ :

$$\mathbf{t} = \mathbf{c}_k - \mathbf{R}\mathbf{c}_{k+1}\quad (11)$$

Finalmente, los valores de los incrementos en los ángulos de desviación, elevación y rotación  $\langle \Delta\alpha, \Delta\beta, \Delta\gamma \rangle$  entre las *poses* se obtienen de la matriz de rotación  $\mathbf{R}$ , completando de esta manera todos los componentes de la media  $\mu_q$ .

### 3.4.2 Estimación de la matriz de covarianza $\Sigma_q$

Típicamente, las matrices de covarianza se obtienen mediante una aproximación lineal de las funciones implicadas en una transformación dada entre variables (ver, por ejemplo, la sección 3.2). Sin embargo en el caso de la solución de formulación cerrada descrita anteriormente, la función no puede ser linealizada debido al cálculo del autovector.

Por lo tanto, en este trabajo proponemos el uso de la versión linealizada del problema, la cual puede ser vista como la minimización del error cuadrático del sistema

$$\begin{pmatrix} X_{k+1}^i \\ Y_{k+1}^i \\ Z_{k+1}^i \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \hline 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_k^i \\ Y_k^i \\ Z_k^i \\ 1 \end{pmatrix}\quad (12)$$

para las variables que determinan el cambio de *pose*, es decir,  $\mu_q = \langle \Delta x, \Delta y, \Delta z, \Delta\alpha, \Delta\beta, \Delta\gamma \rangle$ . Expandiendo (12) obtenemos la posición del punto  $i$ -ésimo en el instante  $k+1$  como una función de su posición en el instante  $k$  (representada por  $\mathbf{X}_k^i$ ) y los incrementos en  $X, Y, Z$ , desviación, elevación y rotación entre los dos instantes de tiempo:

$$\begin{aligned}X_{k+1}^i &= f(\mu_q, \mathbf{X}_k^i) \\ Y_{k+1}^i &= f(\mu_q, \mathbf{X}_k^i) \\ Z_{k+1}^i &= f(\mu_q, \mathbf{X}_k^i)\end{aligned}\quad (13)$$

Linealizando estas ecuaciones llegamos a la siguiente expresión para  $\Sigma_q$ :

$$\Sigma_q^{-1} = \mathbf{H}^T \Sigma^{-1} \mathbf{H} = \sum_i \mathbf{H}^{iT} (\Sigma^i)^{-1} \mathbf{H}^i\quad (14)$$

donde  $\mathbf{H}^i$  representa la matriz Jacobiana de las ecuaciones en (13) para el  $i$ -ésimo punto 3D relativo a  $\mu_q$  y  $\Sigma^i = \Sigma_k^i + \Sigma_{k+1}^i$  es la suma de las matrices de covarianza de las posiciones del punto  $i$ -ésimo en los instantes de tiempo  $k$  y  $k+1$  tal y como se definieron en la ecuación (3). Nótese que, ya que los puntos 3D son incorrelados, la primera expresión en (14) puede ser dividida en la suma de sus elementos bloque diagonales.

## 4 RESULTADOS EXPERIMENTALES

Hemos realizado diferentes experimentos para comparar la odometría clásica basada en codificadores con nuestro método de odometría visual en entornos de interiores. En este trabajo, presentamos uno de estos experimentos donde nuestro robot Sancho, equipado con una cámara estéreo PointGrey Bumblebee, es guiado a través de una habitación mientras va adquiriendo imágenes estéreo y medidas odométricas. También utilizamos

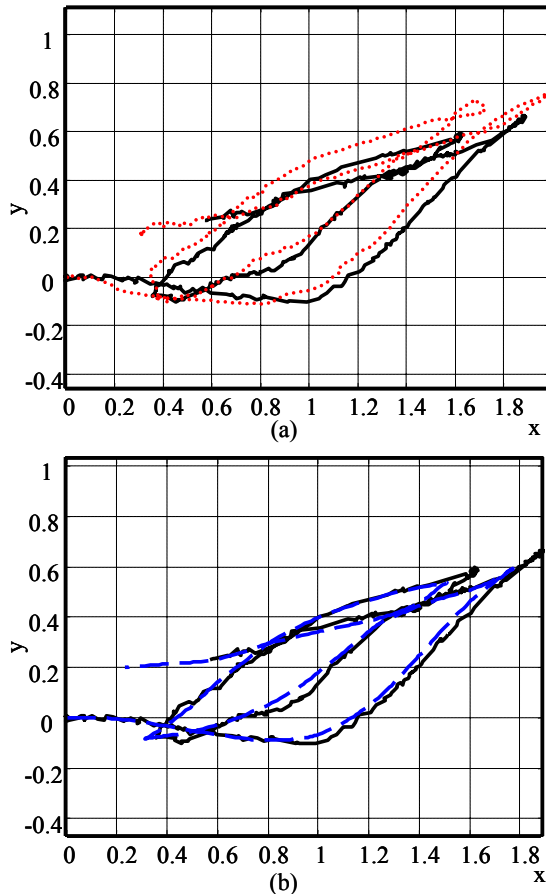


Figura 2. (a) Camino del robot estimado por *scans* láser (línea gruesa) y nuestro propuesta de odometría visual (línea de puntos). (b) Camino del robot estimado por *scans* láser (línea gruesa) y la odometría basada en codificadores (línea discontinua). (c) Ejemplo de imágenes utilizadas en el experimento.

*scans* láser para construir un mapa del entorno y estimar el camino real del robot, el cual se tomará como referencia (líneas gruesas en la Figura 2(a)-(b)). En la Figura 2(c) se muestra un ejemplo de las imágenes usadas en el experimento.

Para comparar las prestaciones de los métodos de odometría, calculamos los errores cometidos por ambos métodos en cada instante de tiempo como la diferencia entre las estimaciones y los valores reales de la *pose*.

En la Figura 3 se muestran los histogramas de los errores en la posición 3D para ambos enfoques. Podemos observar que los dos métodos se comportan de manera muy similar, con la mayoría de los errores cometidos en  $\Delta x$  y  $\Delta y$  por debajo de 5 cm. Nótese que, al seguir el robot un movimiento plano,  $\Delta z$  debería ser igual a cero para todo el experimento. Consecuentemente, nuestro algoritmo proporciona una estimación coherente siempre cercana a  $\Delta z = 0$  con un pequeño error (típicamente 1 cm.), tal y como puede verse en la Figura 3. La distribución del error en la posición 3D se muestra en la última gráfica de la Figura 4.

En cuanto a la estimación de la orientación, la odometría visual alcanza un error en la desviación (el único grado de libertad rotacional en un robot planar) similar a la odometría convencional. Sin embargo, debemos resaltar la precisión de nuestro algoritmo en las otras componentes de la orientación, donde el error máximo está por debajo de  $1^\circ$  (ver los histogramas para los ángulos de elevación y rotación en la Figura 4).

Volviendo a los caminos estimados por el robot mostrados en la Figura 2 para los dos métodos de odometría, podemos destacar su precisión similar a pesar de la mayor dimensionalidad de la odometría visual, lo que, a priori, podría producir una mayor acumulación de errores. Podemos concluir, por tanto, que la razón que explica este buen comportamiento es el pequeño error de la odometría visual en las dimensiones no involucradas en un movimiento plano, es decir,  $\Delta z$ ,  $\Delta \beta$  y  $\Delta \gamma$ .

## 5 CONCLUSIONES

Este artículo ha presentado un nuevo método para realizar odometría visual mediante el cálculo del cambio en 6D entre las *poses* de una cámara en instantes de tiempo distintos. Nuestro método combina la velocidad de la detección y el seguimiento del método de Kanade-Lucas-Tomasi con la capacidad de distinción de características de los descriptores SIFT. Otra contribución de este trabajo en comparación con otros enfoques existentes

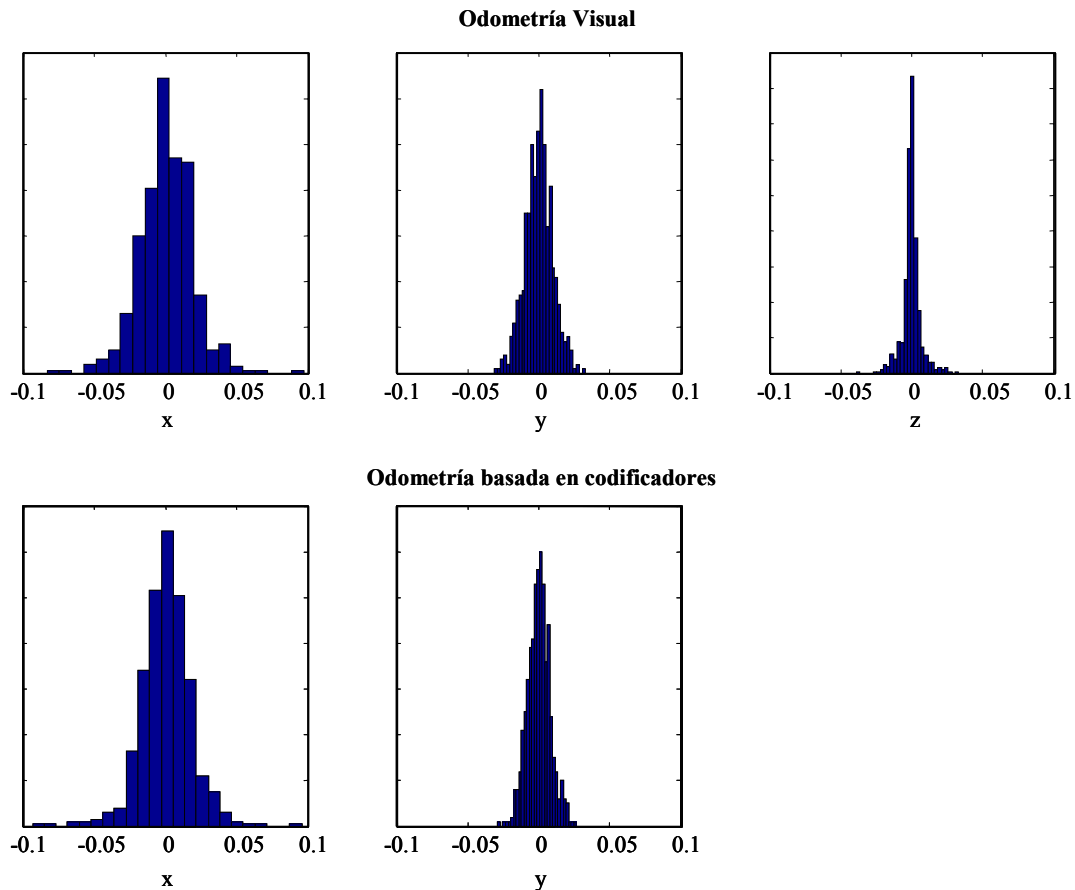


Figura 3. Histogramas de los errores cometidos en la estimación de los cambios de posición del robot para los métodos de odometría visual (gráficas superiores) y odometría basada en codificadores (gráficas inferiores).

es el empleo de una solución óptima y de formulación cerrada al problema de encontrar la transformación 6D entre dos conjuntos de puntos correspondientes. Los resultados demuestran que las prestaciones de nuestro enfoque de odometría visual son bastante similares a las de la odometría convencional para entornos planos, permitiendo además movimientos en 6D. Las líneas de trabajo futuras se centrarán en integrar el método presentado dentro de marcos de trabajo relacionados con el SLAM.

## Referencias

- [1] Campbell, J., Sukthankar, R., Nourbakhsh, I., Pahwa, A., (2005) "A Robust Visual Odometry and Precipice Detection System Using Consumer-grade Monocular Vision", *Int. Conference on Robotics and Automation*, pp. 3421-3427.
- [2] Davison, A.J., Reid, I., Molton, N., Stasse, O., (2007) "MonoSLAM: Real-Time Single Camera SLAM", *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [3] Fernandez, D., Price, A., (2004) "Visual Odometry for an Outdoor Mobile Robot", *Conference on Robotics, Automation and Mechatronics*, pp. 816-821.
- [4] Hahnel, D., Burgard, W., Fox, D., Thrun, S., (2003) "An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements", *Proc. of Int. Conference on Intelligent Robots and Systems (IROS)*.
- [5] Harris, C. J., Stephens, M., (1988) "A combined edge and corner detector", *Proceedings of 4<sup>th</sup> Alvey Vision Conference*, Manchester, pp. 147-151.
- [6] Horn, B.K.P., (1987) "Closed-form solution of absolute orientation using unit quaternions", *Journal of the Optical Society of America A 4*, pp. 629-642.
- [7] Kitchen, L., Rosenfeld, A., (1982) "Gray-level corner detection", *Pattern Recognition Letters*, 1, pp. 95-102.
- [8] Lowe, D. G., (2004) "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, n° 2, pp. 91-110.

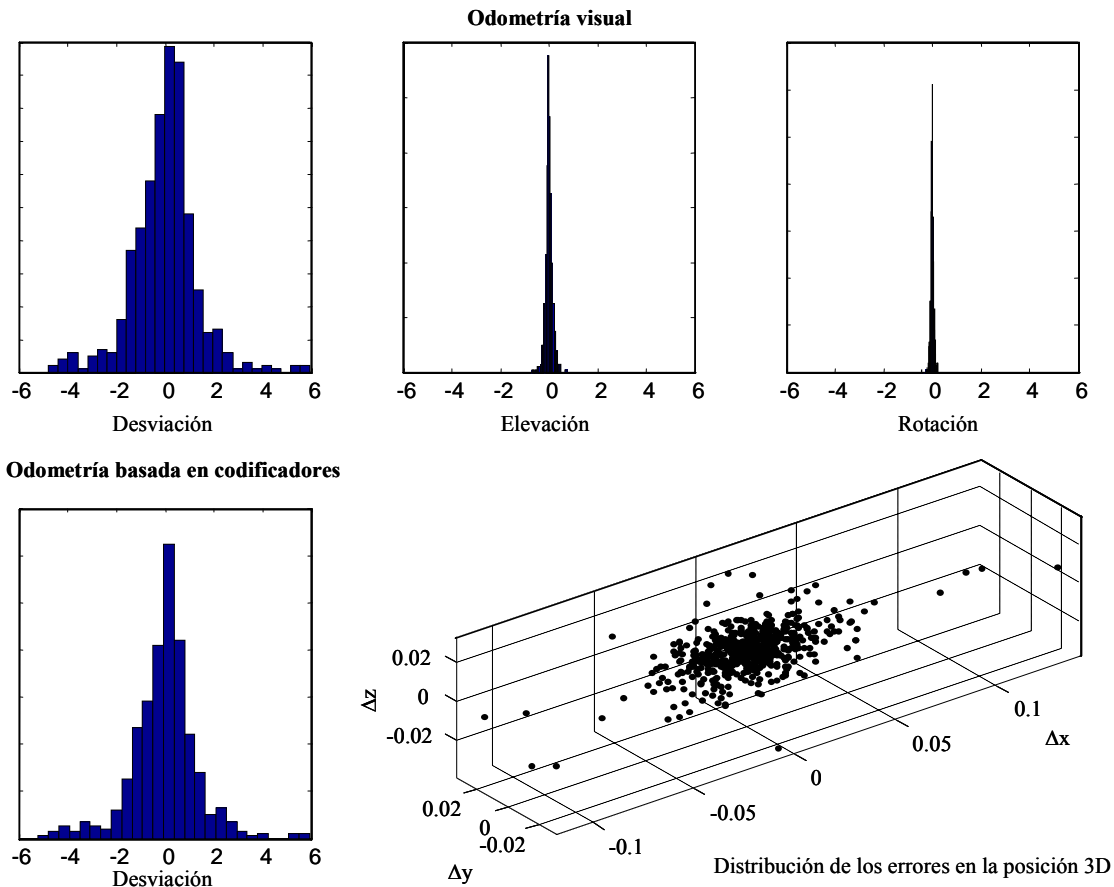


Figura 4. Histogramas de los errores cometidos en la estimación de los cambios de orientación del robot para los métodos de odometría visual (gráficas superiores) y odometría basada en codificadores (gráficas inferiores). (Gráfica inferior de la derecha) Distribución de los errores en la estimación del cambio de posición 3D del robot para el método de la odometría visual.

- [9] Matthies, L., Shafer, S. A., (1987) "Error modeling in Stereo Navigation", *IEEE Journal of Robotics and Automation*, vol. RA-3, nº 3.
- [10] Nistér, D., Naroditsky, O., Bergen, J., (2004) "Visual Odometry", *Computer Society Conference on Computer Vision and Pattern Recognition*, 1, pp. 652-659.
- [11] Olson, C. F., Matthies, L. H., Schoppers, M., Maimone, M. W., (2003) "Rover navigation using stereo ego-motion", *Robotics and Autonomous Systems*, vol. 43, nº 4, pp. 215-229.
- [12] Shi, J., Tomasi, C., (1994) "Good features to track", *Proc. Computer Vision and Pattern Recognition*, pp. 593-600.
- [13] Sim, R., Elinas, P., Griffin, M., Little, J.J., (2005) "Vision-based SLAM using the Rao-Blackwellised Particle Filter", *IJCAI Workshop Reasoning with Uncertainty in Robotics*, Edinburgh, Scotland.
- [14] Stachniss, C., Grisetti, G., Burgard, W., (2005) "Recovering Particle Diversity in a Rao-Blackwellized Particle Filter for SLAM After Actively Closing Loops", *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*.
- [15] Thrun, S., Burgard, W., Fox, D., (2006) *Probabilistic Robotics*. The MIT Press.
- [16] Wang, H., Yuan, K., Zou, W., Zhou, Q., (2005) "Visual Odometry Based on Locally Planar Ground Assumption", *Int. Conference on Information Acquisition*, pp. 59-64.