

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 911 400**

51 Int. Cl.:

G06T 7/579

(2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.06.2017** **PCT/EP2017/065677**

87 Fecha y número de publicación internacional: **28.12.2017** **WO17220815**

96 Fecha de presentación y número de la solicitud europea: **26.06.2017** **E 17733432 (3)**

97 Fecha y número de publicación de la concesión europea: **26.01.2022** **EP 3475917**

54 Título: **Sistema SLAM basado en cámara RGB-D y método del mismo**

30 Prioridad:

24.06.2016 US 201662354251 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:

19.05.2022

73 Titular/es:

ROBERT BOSCH GMBH (100.0%)
Postfach 30 02 20
70442 Stuttgart, DE

72 Inventor/es:

KIM, SOOHWAN;
BABU, BENZUN PIOUS WISELY;
YAN, ZHIXIN y
REN, LIU

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 911 400 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema SLAM basado en cámara RGB-D y método del mismo

5 Campo

Esta divulgación se refiere en general a sistemas SLAM y, más en particular, a un sistema SLAM basado en cámara RGB-D y un método del mismo.

10 Antecedentes

“Dense Visual SLAM for RGB-D Cameras”, por Christian Kerl et al. divulga un método SLAM visual denso para cámaras RGB-D que minimiza tanto el error fotométrico como el error de profundidad en todos los píxeles. A diferencia de los escasos métodos basados en características, este método permite aprovechar mejor la información disponible en los datos de imagen, que conduce a una mayor precisión de postura. Además, se propone una medición de similitud basada en entropía para selección de cuadros clave y detección de cierre de bucle.

“Robust odometry estimation for RGB-D cameras”, Christian Kerl et al. divulga un método rápido y preciso para estimar el movimiento de la cámara a partir de imágenes RGB-D. El enfoque registra dos cuadros RGB-D consecutivos directamente uno sobre el otro al minimizar el error fotométrico. El movimiento de cámara se estima utilizando minimización no lineal en combinación con un esquema de grueso a fino. Para permitir ruido y valores atípicos en los datos de imagen, el documento propone utilizar una función de error robusta que reduce la influencia de los residuos grandes. Además, la formulación divulgada permite la inclusión de un modelo de movimiento que se puede basar en conocimiento previo, filtrado temporal o sensores adicionales como una IMU.

25 Breve descripción

La invención proporciona un método para calcular localización y mapeo simultáneos (SLAM) visuales, así como un sistema SLAM visual como se menciona en las reivindicaciones independientes. Se establecen realizaciones ventajosas en las reivindicaciones dependientes.

30 Breve descripción de los dibujos

Estas y otras características, aspectos, y ventajas de esta divulgación se comprenderán mejor cuando la siguiente descripción detallada se lee con referencia a las figuras anexas en las que caracteres similares representan técnicas similares a lo largo de las figuras, donde:

La figura 1 es un diagrama de bloques que ilustra un sistema SLAM visual;

40 La figura 2 es un diagrama de bloques que ilustra la estructura de un evaluador de restricción de bucle y gráfica de cuadros clave de ejemplo;

La figura 3 ilustra un modelo de sensor de cámara RGB-D;

45 La figura 4 es un diagrama de bloques de una propagación de incertidumbre; y

La figura 5 ilustra un ejemplo de un mapa generado por un sistema SLAM σ -DVO

50 Descripción detallada

La figura 1 es un diagrama de bloques que ilustra un sistema de localización y mapeo simultáneos (SLAM) visuales 100 dividido en extremo frontal 100a y extremo trasero 100b. En el extremo frontal 100a, el sistema 100 utiliza un enfoque de odometría visual al hacer un uso completo de toda la información de píxeles de una cámara RGB-D para generar una estimación de transformación local 112. Es decir, la odometría visual densa 108 o 110 proporciona una estimación de transformación por pares entre dos cuadros de imagen 102, 104, 106. Como se ilustra, la estimación de transformación por pares se realiza entre el cuadro clave 102 y el cuadro actual 104 usando odometría visual densa 108. La segunda estimación de transformación por pares se realiza entre el cuadro actual 104 y el cuadro anterior 106 usando odometría visual densa 110. Se utiliza un generador de cuadros clave 114 para generar un cuadro clave V_k con base en la calidad de la estimación de odometría. En el extremo trasero 100b del sistema 100, se crea una gráfica de cuadros clave $G \subset \{V_k\}$ 116 usando el generador de cuadros clave 114. En un evaluador de restricción de bucle 118, las restricciones basadas en el retorno, por ejemplo, el cierre de bucle a ubicaciones previamente visitadas, se añaden a la gráfica de cuadros clave para mejorar su conectividad. Luego, el optimizador de gráfica 120 optimiza la gráfica final con restricciones para producir un mapa optimizado con la trayectoria 122. Más adelante se describirán más detalles sobre la gráfica de cuadros clave 116 y el evaluador de restricción de bucle 118. Se utiliza un modelo de sensor probabilístico en el extremo frontal 100a y realiza la generación de cuadros clave 114, y detección de restricción de bucle 118 y optimización de gráfica 120 en el extremo trasero 100b.

La figura 2 es un diagrama de bloques que ilustra una estructura de una gráfica de cuadros clave 200 de ejemplo que comprende una optimización de gráfica de extremo trasero 202 y un entorno local 204. En la optimización de gráfica de extremo trasero 202, se optimizan las restricciones de bucle L_{K_i, K_j} combinadas con las restricciones de odometría ponderadas por l_{K_i, K_j} . El cuadro clave K_1 reciente y los cuadros rastreados con respecto a K_1 , $(f_1, \dots, f_n)$ se incluyen en el entorno local 204. Los cuadros clave K_1 y los cuadros rastreados $(f_1, \dots, f_n)$ se determinan con base en la relación de entropías $H_{K_1, f_1}/H_{K_1, f_n}$. Cuando el cuadro actual no contiene suficiente información para rastrear un nuevo cuadro, se genera un nuevo cuadro clave al usar entropía de una estimación de pose de cámara. La estimación de postura de cámara genera un nuevo cuadro clave cuando la entrada estimada entre el cuadro clave y el cuadro actual cae por debajo de un umbral normalizado por la mayor entropía estimada en el entorno local 204. Se asume que la mayor entropía de estimación es la que se encuentra entre el cuadro clave y el primer cuadro. Se propone una estrategia adicional de generación de cuadros clave basada en la estimación de curva de la trayectoria de cámara. La estimación de curva $\rho_{i,k}$ entre los cuadros i y k se define como la relación de la suma de las traslaciones entre los cuadros $(\delta_{i,i-1})$ en el entorno local N con respecto a la traslación entre el cuadro clave y el último cuadro $(\delta_{i,k})$.

$$\rho = \frac{\sum_{i \in N}^k \delta_{i,i-1}}{\delta_{i,k}}$$

ecuación (1)

El retorno a una ubicación previamente visitada ayuda a identificar restricciones adicionales a la gráfica llamada cierre de bucle en el evaluador de restricciones de bucle 118 como se ilustra en la figura 2. Después de la optimización, la gráfica de postura se ajusta con base en los factores de ponderación de borde de diferentes restricciones en la gráfica. Una restricción de bucle errónea en algún momento puede conducir a una trayectoria final mal optimizada. Al extender los métodos anteriores de generación de restricciones de bucle, se pueden usar dos técnicas adicionales para reducir el impacto de restricciones de bucle incorrectas. En primer lugar, las restricciones de cierre de bucle se ponderan con base en el cuadrado inverso de la distancia métrica entre los cuadros clave que forman el cierre de bucle. Esto se basa en la intuición de que la restricción de bucle entre cuadros lejanos es propensa a un error mayor que los cuadros cercanos entre sí. En segundo lugar, se realiza filtrado de oclusión para eliminar falsas restricciones de cierre de bucle. La imagen de profundidad proporciona información de geometría que se puede usar para realizar el filtrado de oclusión entre dos cuadros clave. La desviación estándar de la incertidumbre de modelo de sensor de un punto de profundidad proporciona un límite en el desplazamiento de profundidad máximo posible de la siguiente ecuación:

$$\eta(Z_i) = \frac{q_{pix} b f}{2} \left[\frac{1}{\text{Rnd}(\frac{q_{pix} b f}{Z_i} - 0.5)} - \frac{1}{\text{Rnd}(\frac{q_{pix} b f}{Z_i} + 0.5)} \right]$$

ecuación (2)

Todos los puntos que violan esta suposición se consideran como oclusión.

En la generación de un nuevo cuadro clave, la gráfica de extremo trasero se actualiza con la información de cuadro clave anterior y se crea una estructura de gráfica de doble ventana 200. La gráfica de postura en el extremo trasero se optimiza utilizando, por ejemplo, una biblioteca de código abierto, g2o. Se realiza una optimización final en la terminación de la odometría visual para generar una estimación optimizada de trayectoria de cámara.

En general, las cámaras RGB-D proyectan patrones infrarrojos y recuperan la profundidad de las correspondencias entre dos vistas de imágenes con un pequeño paralaje. Durante este proceso, la disparidad se cuantifica en subpíxeles. Esto introduce un error de cuantificación en la medición de profundidad. El ruido debido al error de cuantificación en la medición de profundidad se define como

$$\eta(Z_i) = \frac{q_{pix} b f}{2} \left[\frac{1}{\text{Rnd}(\frac{q_{pix} b f}{Z_i} - 0.5)} - \frac{1}{\text{Rnd}(\frac{q_{pix} b f}{Z_i} + 0.5)} \right]$$

ecuación (3)

donde q_{pix} es la resolución de subpíxeles del dispositivo, b es la línea de referencia y f es la longitud focal. Este error incrementa cuadráticamente con el intervalo Z_i , evitando así el uso de observaciones de profundidad de objetos lejanos. El ruido de sensor 3D de las cámaras RGB-D se puede modelar con una distribución gaussiana multivariable de media cero cuya matriz de covarianza tiene los siguientes componentes diagonales:

$$\sigma_{11}^2 = \tan\left(\frac{\beta_x}{2}\right) Z_i, \quad \sigma_{22}^2 = \tan\left(\frac{\beta_y}{2}\right) Z_i, \quad \sigma_{33}^2 = \eta(Z_i)^2$$

ecuación (4)

donde la dirección σ_{33}^2 está a lo largo del rayo, y β_x y β_y denotan las resoluciones angulares en direcciones x e y.

La figura 3 ilustra un modelo de sensor de cámara RGB-D. La cámara se encuentra en el origen y está mirando hacia arriba en la dirección z. Para cada intervalo de 1, 2 y 3 metros, se toman muestras de 80 puntos y sus incertidumbres se expresan con elipsoides. El error en la dirección de rayo incrementa cuadráticamente.

La figura 4 es un diagrama de bloques de una propagación de incertidumbre. Cada punto 3D $\mathbf{p}_i$ en la figura 4 se asocia con una distribución gaussiana cuyas matrices de covarianza son Σ_i y Σ'_i , respectivamente,

$$p(\mathbf{p}_i) = \mathcal{N}(\bar{\mathbf{p}}_i, \Sigma_i) \quad \text{ecuación (5)}$$

donde

$$\Sigma_i = R_{\text{ray}} \begin{bmatrix} \sigma_{11}^2 & 0 & 0 \\ 0 & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{bmatrix} R_{\text{ray}}^T \quad \text{ecuación (6)}$$

R_{ray} denota la matriz de rotación entre el rayo y las coordenadas de cámara.

Se utiliza un método de linealización para propagar la incertidumbre a los residuos y la función de probabilidad se puede expresar como una distribución gaussiana,

$$p(\mathbf{r} | \xi) = \mathcal{N}(\mu_i, \Sigma_i) \quad \text{ecuación (7)}$$

donde

$$\mu_i = \begin{bmatrix} \mu_i^I \\ \mu_i^Z \end{bmatrix} = \begin{bmatrix} I_2(\pi(g(\bar{\mathbf{p}}_i, \xi))) - I_1(\mathbf{x}_i) \\ Z_2(\pi(g(\bar{\mathbf{p}}_i, \xi))) - [g(\bar{\mathbf{p}}_i, \xi)]_Z \end{bmatrix} \quad \text{ecuación (8)}$$

$$\Sigma_i = \mathbf{J}_i \Sigma_i \mathbf{J}_i^T + \text{diag}(0, [\Sigma'_i]_{3,3}) \quad \text{ecuación (9)}$$

$$\mathbf{J}_i^T = [\nabla r_i^I \quad \nabla r_i^Z] = \begin{bmatrix} \frac{\partial r_i^I}{\partial \mathbf{p}_i} & \frac{\partial r_i^Z}{\partial \mathbf{p}_i} \end{bmatrix} \quad \text{ecuación (10)}$$

En la presente, $[\Sigma_i]_{3,3}$ denota la varianza del punto $\mathbf{q}_i$ retroproyectado en el eje z de las coordenadas de cámara actuales como se muestra en la figura 4. La estimación de probabilidad máxima es,

$$\hat{\xi}_{\text{Sensor}} = \underset{\xi}{\text{argmin}} \sum_{i=1}^n \mathbf{r}_i^T \Sigma_i^{-1} \mathbf{r}_i \quad \text{ecuación (11)}$$

La matriz de precisión individual se divide en dos raíces cuadradas $\Sigma_i^{-1} = \Sigma_i^{-1/2} \Sigma_i^{-1/2}$ y se normaliza al aplicar la matriz de precisión individual de los residuos ponderados Σ^{-1} como

$$\hat{\xi}_{\text{Sensor}} = \underset{\xi}{\text{argmin}} \sum_{i=1}^n \mathbf{r}_i^T \Sigma_i^{-1/2} \Sigma^{-1} \Sigma_i^{-1/2} \mathbf{r}_i \quad \text{ecuación (12)}$$

Los errores fotométricos y geométricos se pueden definir como,

$$\mathbf{r}_i = \begin{bmatrix} r_i^I \\ r_i^Z \end{bmatrix} = \begin{bmatrix} I_2(\pi(g(\pi^{-1}(\mathbf{x}_i, Z_i), \xi))) - I_1(\mathbf{x}_i) \\ Z_2(\pi(g(\pi^{-1}(\mathbf{x}_i, Z_i), \xi))) - [g(\pi^{-1}(\mathbf{x}_i, Z_i), \xi)]_Z \end{bmatrix} \quad \text{ecuación (13)}$$

donde $Z_i = Z_i(\mathbf{x}_i)$ y $[\cdot]_z$ denota el componente z del vector.

Para encontrar la postura relativa de cámara que minimiza los errores fotométricos y geométricos, la función de energía es la suma de los errores cuadrados ponderados como

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \sum_{i=1}^n \mathbf{r}_i^T \mathbf{W} \mathbf{r}_i \quad \text{ecuación (14)}$$

donde n es el número total de píxeles válidos, y $\mathbf{W} \in \mathbb{R}^{2 \times 2}$ denota los factores de ponderación para diferentes errores.

Dado que la función de energía no es lineal con respecto a la postura relativa de cámara ξ , el algoritmo de Gauss-Newton se aplica en general para encontrar numéricamente la solución óptima y la ecuación (14) se actualiza ahora a:

$$\xi_{k+1} = \xi_k + \Delta \xi, \quad (\mathbf{J}^T (\mathbf{I}_n \otimes \mathbf{W}) \mathbf{J}) \Delta \xi = -\mathbf{J}^T (\mathbf{I}_n \otimes \mathbf{W}) \mathbf{r} \quad \text{ecuación (15)}$$

donde $\otimes$ denota el producto de Kronecker, $\mathbf{r} = (n, \dots, r_n)^T \in \mathbb{R}^{2n \times 1}$, y la matriz jacobiana se define como

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_n \end{bmatrix} \in \mathbb{R}^{2n \times 6}, \quad \mathbf{J}_i = \begin{bmatrix} \frac{\partial r_i^x}{\partial \xi_1} & \dots & \frac{\partial r_i^x}{\partial \xi_6} \\ \frac{\partial r_i^y}{\partial \xi_1} & \dots & \frac{\partial r_i^y}{\partial \xi_6} \end{bmatrix} \in \mathbb{R}^{2 \times 6} \quad \text{ecuación (16)}$$

La ecuación (14) es equivalente con estimación de probabilidad máxima donde cada residuo es independiente y sigue una distribución gaussiana idéntica,

$$\hat{\xi}_{\text{ML}} = \underset{\xi}{\operatorname{argmax}} \sum_{i=1}^n \log p(\mathbf{r}_i | \xi) \quad \text{ecuación (17)}$$

donde $p(\mathbf{r}_i | \xi) = \mathcal{N}(\mathbf{0}, \Sigma)$. Se señala que esto corresponde al caso de $\mathbf{W} = \Sigma^{-1}$ en la ecuación (14). La ecuación (17) se puede reescribir como:

$$\hat{\xi}_{\text{DVO}} = \underset{\xi}{\operatorname{argmin}} \sum_{i=1}^n w_i \mathbf{r}_i^T \Sigma^{-1} \mathbf{r}_i \quad \text{ecuación (18)}$$

donde $w_i = (v + 2)/(v + \mathbf{r}_i^T \Sigma^{-1} \mathbf{r}_i)$. Se señala que esto corresponde al caso de $\mathbf{W} = w_i \Sigma^{-1}$ en la ecuación (14).

Una distribución T para errores fotométricos y propagar un modelo de sensor de una distribución gaussiana para errores geométricos al combinar la ecuación (11) y ecuación (18) para ahora definirse como odometría visual σ -densa (σ -DVO):

$$\hat{\xi}_{\text{Hybrid}} = \underset{\xi}{\operatorname{argmin}} \sum_{i=1}^n \mathbf{r}_i^T \mathbf{W}_i^{1/2} \Sigma^{-1} \mathbf{W}_i^{1/2} \mathbf{r}_i \quad \text{ecuación (19)}$$

donde la matriz de factores de ponderación $\mathbf{W}_i = \operatorname{diag}(w_i^I, w_i^Z)$ y

$$w_i^I = \frac{v + 1}{v + \left(\frac{r_i^I}{\sigma}\right)^2} \quad \text{ecuación (20)}$$

$$w_i^Z = \frac{1}{\nabla r_i^Z \Sigma_i^{-1} \nabla r_i^{Z^T} + [\Sigma_i']_{3,3}} \quad \text{ecuación (21)}$$

El algoritmo σ -DVO se puede implementar en cualquier dispositivo cliente adecuado, tal como teléfono inteligente, tableta, teléfono móvil, asistente digital personal (PDA) y cualquier dispositivo. Volviendo a la figura 1, el sistema SLAM 100 con algoritmo σ -DVO integrado utiliza un número menor de cuadros clave y se debe a una desviación reducida en el sistema. Un número reducido de cuadros clave indica menos requisitos de cálculo en el extremo trasero del sistema.

La figura 5 ilustra un ejemplo de un mapa generado por un sistema SLAM σ -DVO 100. Como se puede ver, se genera una trayectoria consistente usando el sistema SLAM σ -DVO 100.

Las realizaciones también se pueden poner en práctica en entornos informáticos distribuidos donde se realizan tareas mediante dispositivos de procesamiento locales y remotos que se enlazan (ya sea mediante enlaces cableados, enlaces inalámbricos o mediante una combinación de los mismos) a través de una red de comunicaciones.

Las instrucciones ejecutables por computadora incluyen, por ejemplo, instrucciones y datos que hacen que una computadora de propósito general, computadora de propósito especial o dispositivo de procesamiento de propósito especial realicen una cierta función o grupo de funciones. Instrucciones ejecutables por computadora también incluyen módulos de programa que se ejecutan por computadoras en entornos independientes o de red. En general, los módulos de programa incluyen rutinas, programas, objetos, componentes y estructuras de datos, etc. que realizan tareas particulares o implementan tipos de datos abstractos particulares. Las instrucciones ejecutables por computadora, estructuras de datos asociadas y módulos de programa representan ejemplos de los medios de código de programa para ejecutar los pasos de los métodos divulgados en la presente. La secuencia particular de estas instrucciones ejecutables o estructuras de datos asociadas representa ejemplos de actos correspondientes para implementar las funciones descritas en estos pasos.

REIVINDICACIONES

1. Un método para calcular localización y mapeo simultáneos (SLAM) visuales que comprende:
- 5 generar, mediante un módulo de odometría visual (100A), una estimación de odometría (112);
- crear, mediante un generador de cuadros clave (114), cuadros clave y una gráfica de cuadros clave (116);
- 10 añadir restricciones a la gráfica de cuadros clave (116), usando un evaluador de restricción de bucle (118), con base en un cierre de bucle, donde el cierre de bucle es el retorno a ubicaciones previamente visitadas; y
- optimizar, utilizando un optimizador de gráfica (120), la gráfica de cuadros clave (116) obtenida en el paso anterior, con restricciones para producir un mapa optimizado con una trayectoria (112);
- 15 caracterizado porque
- las restricciones de cierre de bucle se ponderan con base en el cuadrado inverso de la distancia métrica entre los cuadros clave que forman el cierre de bucle.
- 20 2. El método de la reivindicación 1, que comprende además:
- generar un nuevo cuadro clave entre un cuadro clave y un cuadro actual antes de generar la estimación de odometría (112).
- 25 3. Un sistema de localización y mapeo simultáneos (SLAM) visuales, que comprende:
- un módulo de odometría visual (100A) configurado para generar una estimación de odometría (112);
- un generador de cuadros clave (114) configurado para crear cuadros clave y una gráfica de cuadros clave (116);
- 30 un evaluador de restricción de bucle (118) configurado para añadir restricciones a la gráfica de cuadros clave (116) con base en un cierre de bucle, donde el cierre de bucle es el retorno a ubicaciones previamente visitadas; y
- un optimizador de gráfica (120) configurado para optimizar la gráfica de cuadros clave (116) con restricciones para producir un mapa optimizado con una trayectoria (112);
- 35 caracterizado porque
- el evaluador de restricción de bucle (118) se adapta para ponderar las restricciones de cierre de bucle con base en el cuadrado inverso de la distancia métrica entre los cuadros clave que forman el cierre de bucle.
- 40

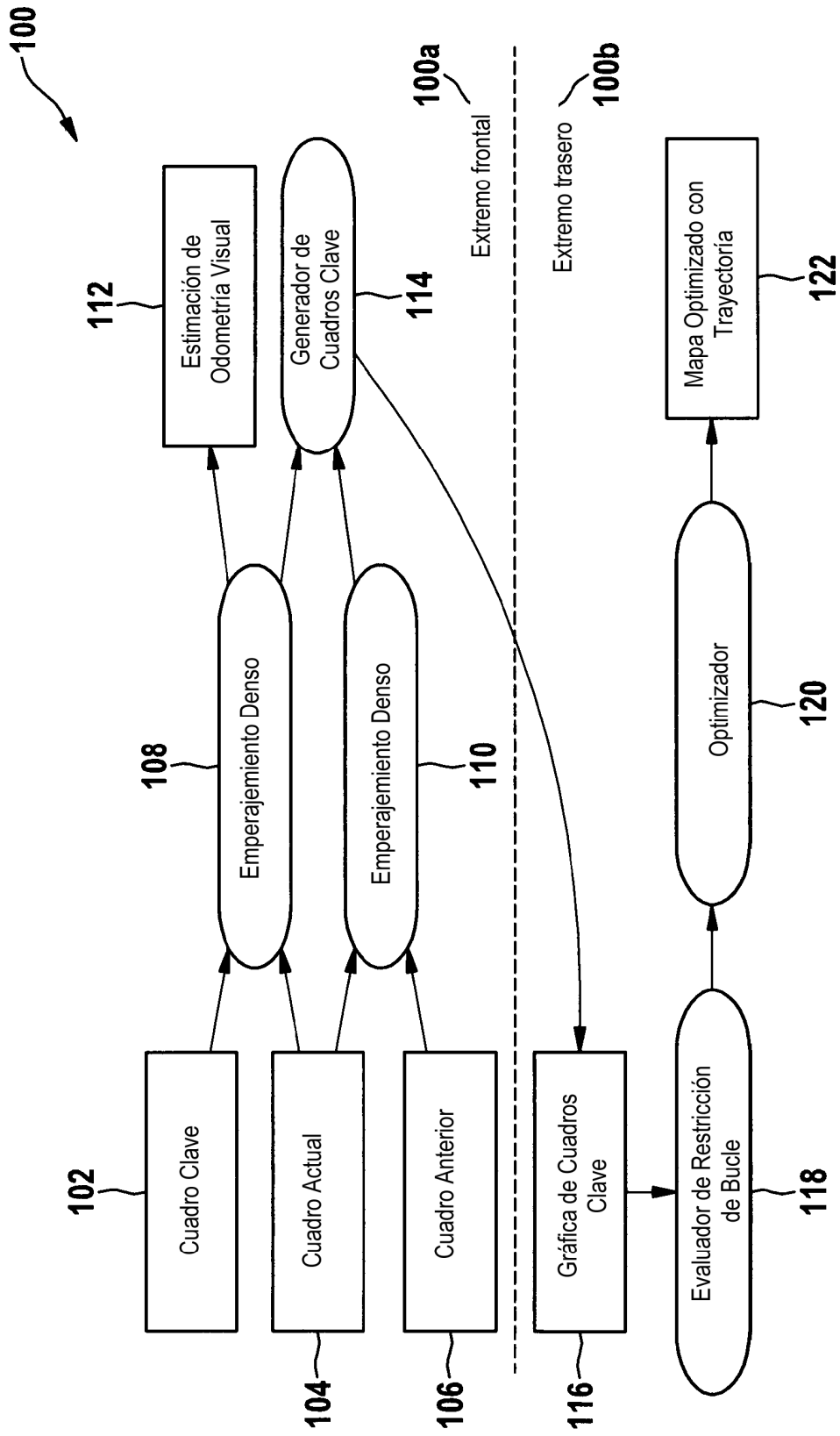


Fig. 1

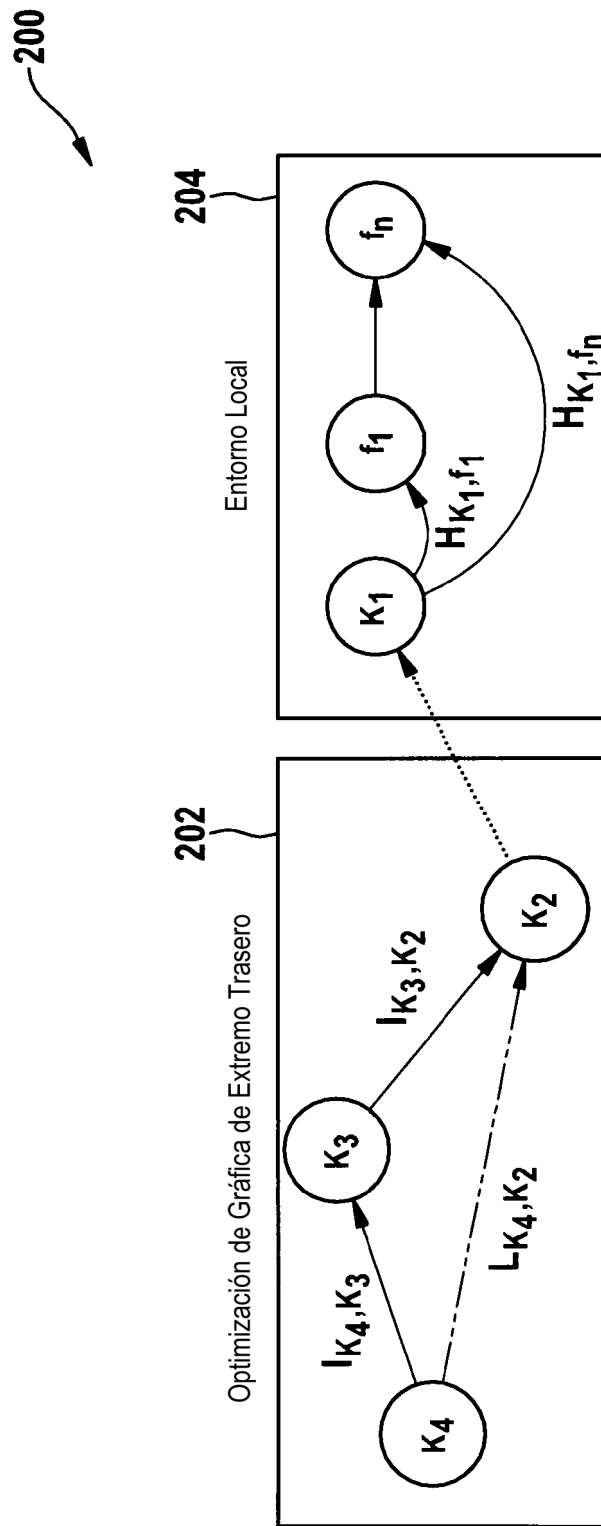


Fig. 2

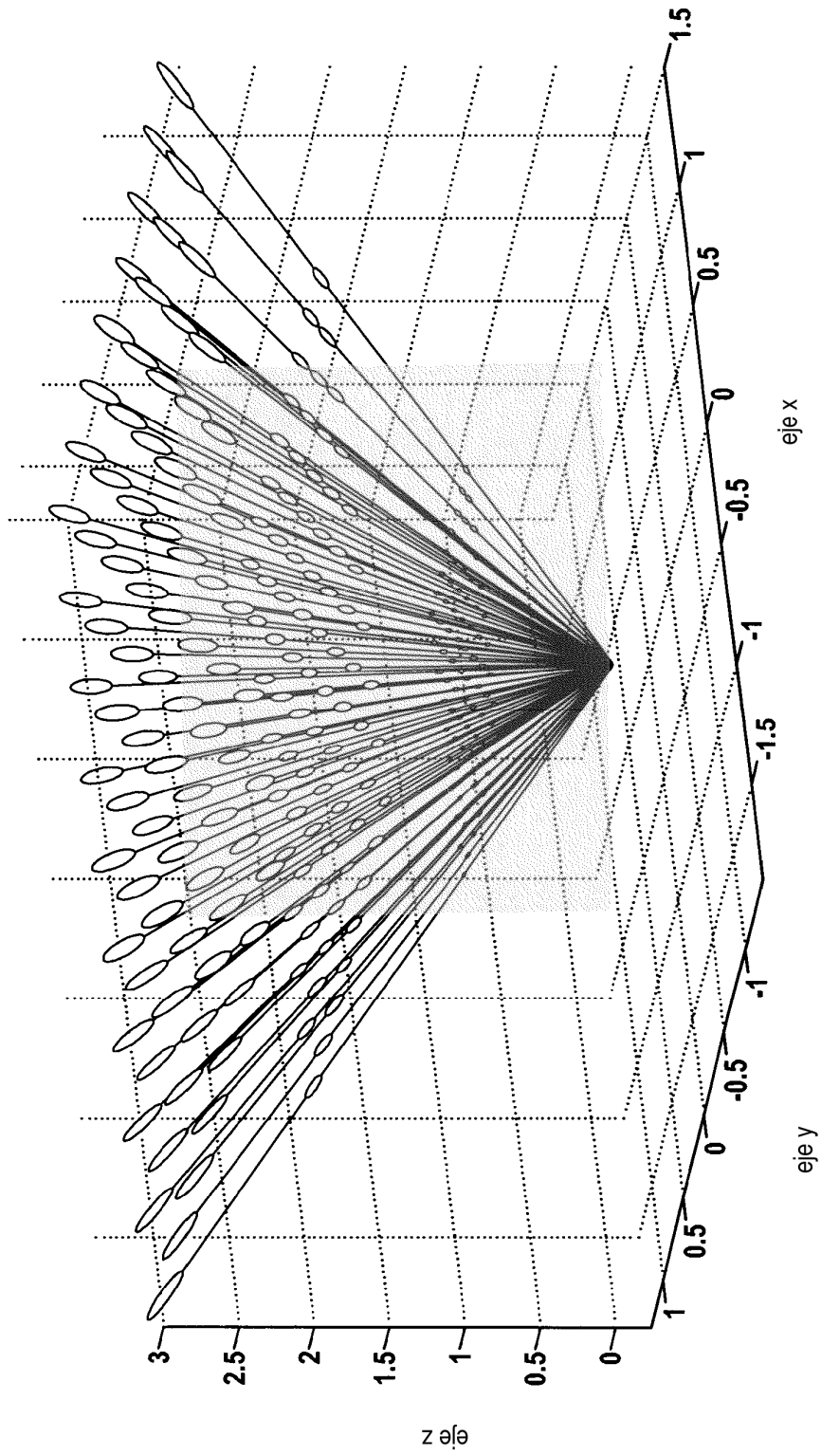


Fig. 3

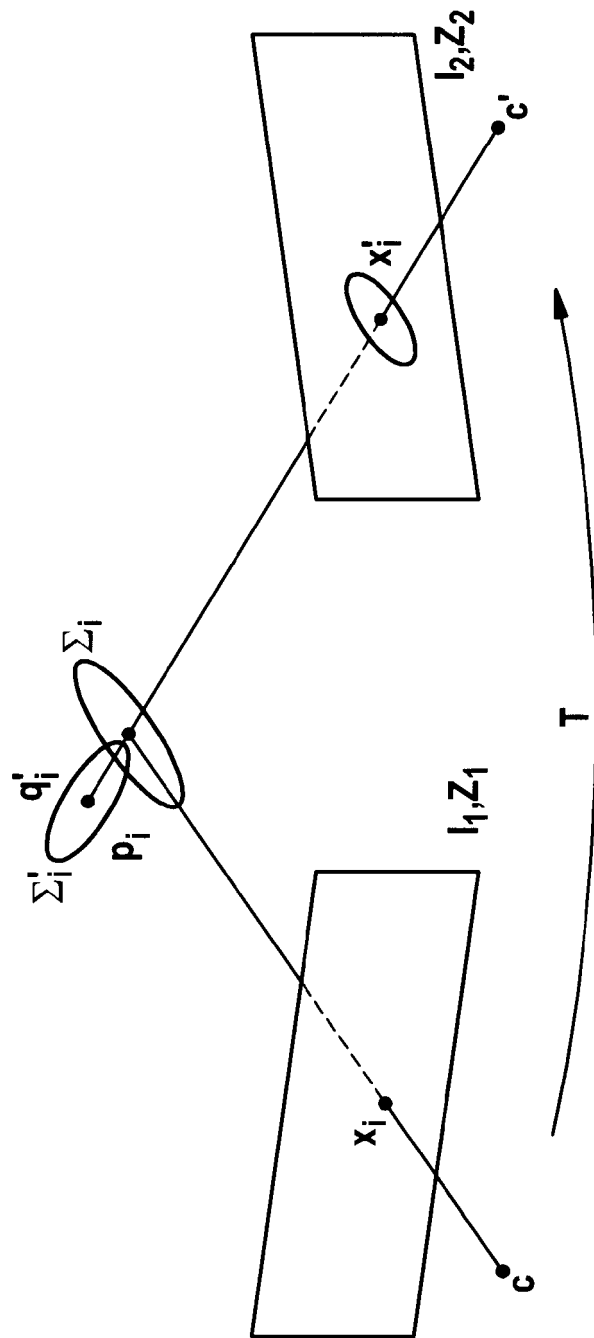


Fig. 4

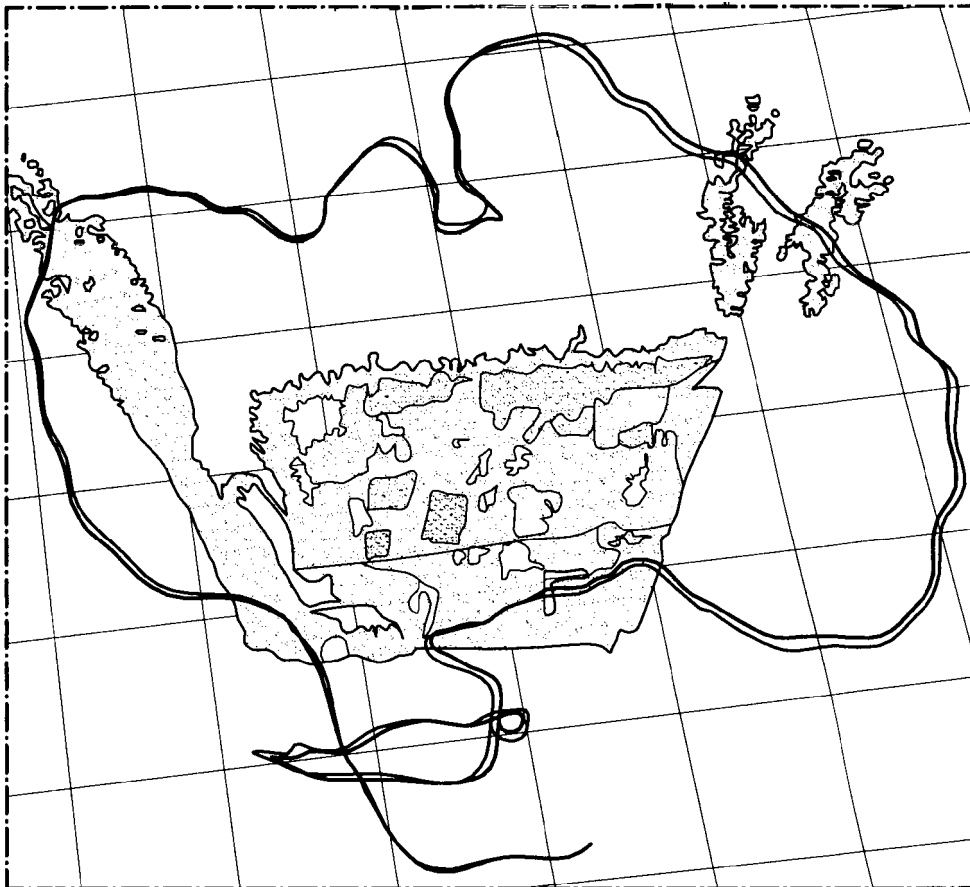


Fig. 5