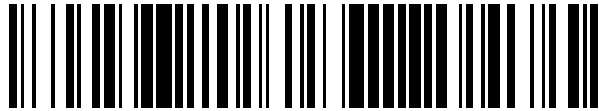


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 564 697**

51 Int. Cl.:

**G06T 7/00**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.10.2012 E 12775001 (6)**

97 Fecha y número de publicación de la concesión europea: **03.02.2016 EP 2769360**

54 Título: **Procedimiento de localización de objetos por resolución en el espacio tridimensional de la escena**

30 Prioridad:

**21.10.2011 FR 1159594**

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

**28.03.2016**

73 Titular/es:

**COMMISSARIAT À L'ÉNERGIE ATOMIQUE ET  
AUX ÉNERGIES ALTERNATIVES (100.0%)  
25, Rue Leblanc, Bâtiment "Le Ponant D"  
75015 Paris, FR**

72 Inventor/es:

**DHOME, YOANN;  
LUVISON, BERTRAND y  
PHAM, QUOC-CUONG**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 564 697 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

Procedimiento de localización de objetos por resolución en el espacio tridimensional de la escena

La invención se sitúa en el campo de la videovigilancia mediante cámaras calibradas. Esta se refiere a un procedimiento de localización de objetos de interés en una escena a partir de imágenes proporcionadas por una cámara. Los objetos de interés pueden ser en particular personas o vehículos.

Numerosas aplicaciones de tratamiento de imágenes pretenden detectar objetos de interés en una imagen o en un flujo de imágenes adquirido por una cámara. Estas aplicaciones se basan en unos métodos que se pueden clasificar en dos familias. La primera familia de métodos se basa en el reconocimiento de formas. El principio consiste en reconocer una o varias características muy específicas del objeto buscado, por ejemplo el contorno de una cabeza o la silueta de una persona. La búsqueda de estas características en el conjunto de la escena es una tarea difícil, por una parte, por las deformaciones geométricas causadas por las distorsiones ópticas de los sensores y por las diferencias de punto de vista de los objetos buscados y, por otra parte, por las ocultaciones entre objetos buscados. A título de ejemplo, la silueta de una persona vista de frente es muy diferente de la de una persona vista desde arriba. Las distorsiones ópticas están causadas por el tipo de cámara utilizada. Estas son especialmente pronunciadas en particular para las cámaras omnidireccionales y las cámaras denominadas "fisheye". Ahora bien, los métodos de reconocimiento de formas precisan un aprendizaje sobre unas bases acreditadas. Estas bases dan ejemplos de personas así como contraejemplos en función de un punto de vista particular y de un tipo de cámara dado. En consecuencia, la configuración de un sistema de localización de objetos de interés que utilizan un método de reconocimiento de formas es una tarea delicada, lo que precisa la realización de una base de aprendizaje específica para el punto de vista particular de la cámara. La segunda familia de métodos de detección de objetos en una imagen se basa en un criterio de optimización del espacio tridimensional (3D). La idea es maximizar, en la imagen, el solapamiento entre una máscara obtenida por sustracción de fondo, y la proyección de uno o varios modelos 3D del objeto buscado. Un ejemplo de dicho método se describe en el documento Alahi Alexandre, L. Jacques, Y. Boursier y P. Vanderghyest, "Sparsity-driven People Localization Algorithm: Evaluation in Crowded Scenes Environments", IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, Utah, 2009. En este documento, se considera una cuadrícula fija de posiciones 3D en el suelo así como un modelo geométrico de persona, en este caso un elipsoide que representa la parte superior de la persona y un cilindro para la parte inferior. De aquí en adelante, se llama máscara binaria a una imagen en la que cada píxel toma bien un primer valor, por ejemplo '0', o bien un segundo valor, por ejemplo '1', en función de un parámetro del píxel considerado. Según el método de Alahi, para cada posición de la cuadrícula, se calcula una máscara binaria de la proyección del modelo geométrico, llamado átomo. Cada máscara binaria toma el valor '1' en cada píxel correspondiente a la proyección del modelo geométrico en la imagen, y '0' en el resto. La localización de las personas en la imagen consiste entonces en minimizar la diferencia entre la máscara binaria obtenida por sustracción de fondo y una combinación lineal de átomos, estando cada átomo o presente o ausente. Dicho de otro modo, el método consiste en buscar el conjunto de las posiciones en la cuadrícula que da, mediante la proyección de un modelo geométrico en cada una de estas posiciones, la imagen más parecida a la imagen en la que se buscan unas personas. Uno de los principales inconvenientes de este método es su complejidad algorítmica. La búsqueda de las personas se realiza en el espacio de imagen, lo que implica la resolución de un sistema lineal cuya dimensión es igual al número de píxeles en la imagen multiplicado por el número de posiciones en la cuadrícula. En la práctica, el método precisa importantes recursos de cálculo. Incluso con un submuestreo de la imagen, el método no se puede aplicar para un tratamiento en tiempo real. Además, el método presenta el inconveniente de basarse en la utilización de una máscara binaria de sustracción de fondo. Sin dicha máscara puede fusionar grupos de personas separados, por ejemplo a causa de las sombras, y fragmentar grupos normalmente juntos, por ejemplo a causa de prendas de ropa cuyos colores son localmente cercanos a los colores del segundo plano. En consecuencia, la eficacia del método está limitada por la de la etapa de sustracción de fondo.

Un objetivo de la invención es, en particular, resolver todos o parte de los inconvenientes mencionados con anterioridad permitiendo un tratamiento de localización de objetos en tiempo real. Para ello, la invención tiene por objeto un procedimiento de localización de objetos de interés en una escena observada mediante un sistema de adquisición de imágenes, localizándose los objetos de interés a partir de un mapa de presencia (CP del francés

*carte de présence*) inicial  $p_{CP}^1$  que modeliza unas posiciones  $i$  en la escena y que comprende, para cada posición  $i$ , un valor  $p_{CP}^1(i)$  representativo de la probabilidad de que un objeto de interés se encuentre en la posición  $i$  considerada, obteniéndose cada valor  $p_{CP}^1(i)$  a partir de un criterio de localización definido en un espacio de imagen del sistema de adquisición de imágenes. El procedimiento se caracteriza porque comprende una iteración de las siguientes etapas sucesivas mientras uno al menos de los valores  $p_{CP}^k(i)$  del mapa de presencia  $p_{CP}^k$  considerado para la iteración actual  $k$  es superior a un umbral predeterminado:

- determinar la posición  $n^k$  en el mapa de presencia  $p_{CP}^k$  para la cual el valor  $p_{CP}^k(n^k)$  es máximo, considerándose un objeto de interés como presente en dicha posición  $n^k$ ,

5 a partir de átomos  $A_j$  predeterminados para cada posición  $j$  del mapa de presencia  $p_{CP}^k$ , el átomo  $A_j$  de una posición  $j$  que comprende, para cada posición  $i$ , un valor  $A_j(i)$  representativo del solapamiento entre la proyección  $m'(i)$  en el espacio de imagen de un modelo tridimensional  $M'(i)$  colocado en la posición  $i$  y la proyección  $m'(j)$  en el espacio de imagen de un modelo tridimensional  $M'(j)$  colocado en la posición  $j$ , siendo cada modelo tridimensional representativo de un objeto de interés, determinar un mapa de presencia  $p_{CP}^{k+1}$  cuyo cada valor  $p_{CP}^{k+1}(j)$  es igual al valor  $p_{CP}^k(j)$  menos una combinación lineal de los valores  $A_{nk}(j)$  de los átomos  $A_{nk}$  que corresponden a las posiciones  $n^k$  en las que unos objetos de interés se consideran como presentes.

10 Según una forma de realización relativamente simple, cada valor  $A_i(j)$  de un átomo  $A_i$  se determina como la relación entre el número de píxeles comunes entre las proyecciones  $m'(i)$  y  $m'(j)$  y el número de píxeles de la proyección  $m'(j)$ .

La etapa de determinación del mapa de presencia  $p_{CP}^{k+1}$  se realiza, por ejemplo, mediante una minimización de mínimos cuadrados entre las probabilidades  $p_{CP}^k(i)$  del mapa de presencia de la iteración actual  $k$  y una combinación lineal de los átomos  $A_{nk}$  que corresponde a las posiciones  $n^k$  en las que unos objetos de interés se consideran como presentes en la iteración  $k$  actual.

15 Según una forma particular de realización, el mapa de presencia inicial  $p_{CP}^1$  se determina mediante las siguientes etapas:

- 20 ▪ determinar una imagen de proyección  $B_{m(i)}$  para cada posición  $i$  del mapa de presencia inicial  $p_{CP}^1$ , presentando cada imagen de proyección  $B_{m(i)}$  unas dimensiones iguales a las de una imagen del sistema de adquisición de imágenes y que corresponden a la proyección  $m(i)$  en el espacio de imagen de un modelo tridimensional  $M(i)$  representativo de un objeto de interés colocado en la posición  $i$  considerada, tomando cada píxel de la imagen de proyección  $B_{m(i)}$  un primer valor cuando pertenece a la proyección  $m(i)$  y un segundo valor en caso contrario,
- 25 ▪ determinar, mediante un tratamiento de localización de objetos realizado a partir de la imagen del sistema de adquisición de imágenes, una imagen de presencia de objetos que tiene unas dimensiones iguales a las de la imagen del sistema de adquisición de imágenes, tomando cada píxel de la imagen de presencia de objetos un valor en función de su correspondencia con un objeto presente en la escena,
- determinar el mapa de presencia inicial  $p_{CP}^1$  en función de las imágenes de proyección  $B_{m(i)}$  y de la imagen de presencia de objetos.

30 Según una primera variante de realización, el tratamiento de localización de objetos es un tratamiento de sustracción de fondo en el que cada píxel de la imagen de presencia de objetos toma un primer valor cuando corresponde a un objeto presente en la escena y un segundo valor en caso contrario.

35 La etapa de determinación del mapa de presencia inicial  $p_{CP}^1$  puede entonces comprender una etapa que consiste en determinar cada valor  $p_{CP}^1(i)$  como siendo igual a la relación entre el número de píxeles que tienen a la vez el primer valor en la imagen de presencia de objetos y el primer valor en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada y el número de píxeles que tienen el primer valor en dicha imagen de proyección  $B_{m(i)}$ .

Según una segunda variante de realización, el tratamiento de localización de objetos es un tratamiento de sustracción de fondo en el que cada píxel de la imagen de presencia de objetos toma un valor en función de la probabilidad de que corresponda a un objeto presente en la escena.

40 La etapa de determinación del mapa de presencia inicial  $p_{CP}^1$  puede entonces comprender una etapa que consiste en determinar cada valor  $p_{CP}^1(i)$  como siendo igual a la relación entre la suma de los valores de los píxeles de la imagen de presencia de objetos que corresponde a los píxeles que tienen el primer valor en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada y el número de píxeles que tienen el primer valor en dicha imagen de proyección  $B_{m(i)}$ .

45 Según una forma particular de realización, la escena se observa mediante varios sistemas de adquisición de imágenes, determinándose el mapa de presencia inicial  $p_{CP}^1$  a partir de varios mapas de presencia iniciales  $p_{CP}^e$ , obteniéndose cada valor  $p_{CP}^e(i)$  de un mapa de presencia inicial  $p_{CP}^e$  a partir de un criterio de localización definido en un espacio de imagen del sistema de adquisición de imágenes considerado, siendo cada valor  $p_{CP}^1(i)$  igual a

una media del conjunto de los valores  $p_{CP}^E(i)$  de la posición  $i$  considerada. Esta forma particular de realización permite dar fiabilidad a los valores  $p_{CP}^1(i)$  del mapa de presencia inicial  $p_{CP}^1$ , y cubrir un mayor número de posiciones según unos puntos de vista diferentes.

5 La invención tiene, en particular, como ventaja que esta permite utilizar cualquier tipo de tratamiento de localización para construir el mapa de presencia, y no únicamente un tratamiento de sustracción de fondo. Permite, además, combinar varios tratamientos y varios puntos de vista.

Se entenderá mejor la invención y se mostrarán otras ventajas en la lectura de la descripción que viene a continuación, hecha en relación con los dibujos adjuntos, en los que:

- 10
- la figura 1 representa de manera esquemática un mapa de presencia que modeliza unas posiciones de ocupación posible en una escena para unos objetos que se pretende localizar;
  - la figura 2 representa, en forma de diagrama de flujo, un ejemplo de etapas implementadas en el procedimiento de localización de objetos de interés según la invención;
  - la figura 3 ilustra el enmascaramiento de objetos entre sí y el problema asociado de localización de los objetos en el espacio de imagen de una cámara.
- 15

La figura 1 representa de manera esquemática un ejemplo de escena en la que se pretende localizar unos objetos de interés. En la figura 1, la escena 10 representa una habitación de un edificio. La escena puede, no obstante, ser cualquier tipo de lugar en el que se desee buscar unos objetos. Se trata, por ejemplo, de una autopista en la que se pretende contabilizar un número de vehículos, o un hall de una estación en el que se pretende localizar unas personas. Un sistema de adquisición de imágenes, como una cámara de vídeo 11, está situado y orientado de modo que se observe al menos una parte de la escena. La escena se observa entonces desde un punto de vista determinado. La cámara 11 proporciona unas imágenes digitales a partir de las cuales se buscan los objetos de interés. Cada imagen está, por ejemplo, formada por una matriz de píxeles que constan cada uno de un valor representativo de la intensidad luminosa en un punto de la escena. La cámara 11 está calibrada. Dicho de otro modo, se conoce la relación entre las coordenadas espaciales de cada punto de la escena y el píxel asociado en la imagen.

20

25

El procedimiento de localización de objetos de interés según la invención se basa en la utilización de un mapa de probabilidad de la presencia de objetos de interés en la escena. Este mapa, llamado mapa de presencia y escrito como  $P_{CP}$ , comprende un conjunto de posiciones  $i$  definidas en el espacio tridimensional de la escena y, para cada una de estas posiciones, un valor  $P_{CP}(i)$  representativo de la probabilidad de que un objeto de interés se encuentre en la posición  $i$  considerada. De este modo las posiciones del mapa de presencia forman unas posiciones de ocupación posibles para los objetos de interés. Las posiciones están, por ejemplo, dispuestas de modo que formen una cuadrícula. El mapa de presencia 12 está representado en la figura 1 en forma de un conjunto de cruces cada una representando una posición de ocupación posible en la escena para un objeto de interés. El procedimiento de localización de objetos según la invención utiliza, por otra parte, unos modelos tridimensionales (3D) representativos de la ocupación de los objetos de interés en la escena. A título de ejemplo, se considera que los objetos de interés son unas personas y que las posiciones de ocupación posibles corresponden a las ubicaciones en las que las personas tendrían sus pies. El mapa de presencia 12 puede por tanto llamarse mapa de presencia en el suelo. Las posiciones del mapa de presencia pueden también, según el tipo de objetos de interés buscados y la geometría de la escena, estar situados en otro lugar distinto de un suelo plano. El modelo 3D representativo de una persona es, por ejemplo, un elipsoide. El mapa de presencia 12 se puede modelizar matemáticamente en forma de un vector del que cada componente está asociada a una posición  $i$  en la escena y cuyo valor  $P_{CP}(i)$  define la probabilidad de presencia de un objeto de interés. El valor de las componentes es, por ejemplo, igual a '0' cuando es seguro que ninguna persona se encuentra en la posición considerada, e igual al valor '1' cuando es seguro que ahí se encuentra una persona.

30

35

40

45

La figura 2 representa, en forma de diagrama de flujo, un ejemplo de etapas implementadas en el procedimiento de localización de objetos de interés según la invención. A continuación en la descripción, se escribe como  $M(i)$  el modelo 3D seleccionado para representar un objeto de interés en la posición  $i$  en la escena, y  $m(i)$  su proyección en un espacio de imagen de la cámara. En una primera fase 21, se determina el mapa de presencia  $P_{CP}$ . De este modo la fase 21 consiste en determinar, para cada posición  $i$ , la probabilidad  $P_{CP}(i)$  de que un objeto de interés se encuentre en esta posición  $i$  considerada. La fase 21 comprende, por ejemplo, las siguientes etapas. En una primera etapa 211, se determina la proyección  $m(i)$  del modelo 3D  $M(i)$  individualmente para cada posición  $i$ . Estas proyecciones se determinan mediante la relación entre el espacio tridimensional de la escena y el espacio de imagen de la cámara, es decir entre las coordenadas espaciales de la escena y los píxeles de la imagen. Se pueden definir una imágenes, llamadas imágenes de proyección  $B_{m(i)}$ , con unas dimensiones iguales a las de la imagen proporcionada por la cámara a partir de las proyecciones  $m(i)$ . Cada imagen de proyección  $B_{m(i)}$  toma el valor '1' en el interior del modelo proyectado  $m(i)$  y '0' en el exterior. Escribiendo como  $M \times N$  las dimensiones de la imagen, siendo por ejemplo  $M$  el número de líneas y  $N$  el número de columnas de píxeles, las imágenes de proyección  $B_{m(i)}$  se pueden modelizar mediante la siguiente relación:

50

55

$$B_{m(i)} \in \{0,1\}^{M \times N}$$

5 En una segunda etapa 212, se aplica un tratamiento de localización de objetos en la imagen suministrada por la cámara y en la que se buscan unos objetos de interés. Este tratamiento de localización de objetos produce una imagen de presencia de objetos con las mismas dimensiones que las de la imagen proporcionada por la cámara. Se trata, por ejemplo, de un tratamiento de sustracción de fondo. Este tratamiento consiste en diferenciar los píxeles que corresponden al fondo, es decir a la escena en ausencia de objetos, de los píxeles que corresponden a unos objetos introducidos en la escena. Estos objetos pueden o no ser objetos de interés. La imagen de presencia de objetos producida por el tratamiento de sustracción de fondo se llama imagen de sustracción de fondo  $B_{SdF}$ . Cada píxel de la imagen  $B_{SdF}$  toma, por ejemplo, el valor '0' si corresponde al fondo, y el valor '1' si corresponde a un objeto. De este modo la imagen  $B_{SdF}$  se puede calificar de máscara binaria. La imagen  $B_{SdF}$  se puede modelizar mediante la siguiente relación:

$$B_{SdF} \in \{0,1\}^{M \times N}$$

15 Las etapas 211 y 212 se pueden realizar en cualquier orden. Sin embargo, la etapa 211 de determinación de las imágenes de proyección se realiza de manera ventajosa de forma previa a las otras etapas del procedimiento. En efecto, para una cámara situada en una escena dada, las imágenes de proyección  $B_{m(i)}$  solo dependen del modelo 3D utilizado. De este modo, se pueden determinar los mapas de forma previa para los diferentes modelos 3D susceptibles de utilizarse en el procedimiento de localización. En una tercera etapa 213, el mapa de presencia  $P_{CP}$  se determina a partir de la imagen de presencia de objetos, por ejemplo la imagen de sustracción de fondo  $B_{SdF}$ , y de las imágenes de proyección  $B_{m(i)}$ . Para cada posición  $i$  del mapa de presencia  $P_{CP}$ , la probabilidad  $P_{CP}(i)$  se determina, por ejemplo, como siendo la relación entre el número de píxeles que tienen a la vez el valor '1' en la imagen de sustracción de fondo  $B_{SdF}$  y el valor '1' en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada, y el número de píxeles que tienen el primer valor en dicha imagen de proyección  $B_{m(i)}$ :

$$p_{CP}(i) = \frac{|B_{m(i)} \cap B_{SdF}|}{|B_{m(i)}|}$$

en la que el operador  $||$  define el número de píxeles en la imagen  $I$ .

25 La imagen de sustracción de fondo  $B_{SdF}$  determinada en la etapa 212 presenta el inconveniente de generar una pérdida de información aprovechable en unas etapas posteriores del procedimiento según la invención. De manera ventajosa, la imagen  $B_{SdF}$  se puede sustituir por una imagen denominada de distancia  $D_{SdF}$  en la que los píxeles no toman necesariamente el valor '0' o '1', sino que pueden tomar cualquier valor comprendido entre '0' y '1', en función de la probabilidad de que los píxeles considerados correspondan al fondo o a un objeto. Los píxeles toman, por ejemplo, el valor '0' si es seguro que estos corresponden al fondo, el valor '1' si es seguro que estos corresponden a un objeto, o un valor intermedio en función de la probabilidad de que correspondan a un objeto. La imagen  $D_{SdF}$  se puede modelizar mediante la siguiente relación:

$$D_{SdF} \in [0,1]^{M \times N}$$

35 Las probabilidades  $P_{CP}(i)$  del mapa de presencia pueden entonces determinarse como siendo iguales a la relación entre la suma de los valores de los píxeles de la imagen de distancia  $D_{SdF}$  que corresponde a los píxeles que tienen el valor '1' en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada, y el número de píxeles que tienen el valor '1' en esta imagen de proyección  $B_{m(i)}$ :

$$p_{CP}(i) = \frac{\sum_{(x,y)|B_{m(i)}(x,y)=1} D_{SdF}(x,y)}{|B_{m(i)}|}$$

en la que los pares  $(x, y)$  designan los píxeles en la línea  $x$  y en la columna  $y$  de la imagen considerada.

40 Por otra parte, el mapa de presencia se puede determinar a partir de varias cámaras que observan la escena desde unos puntos de vista diferentes. Se escribe como  $C$  un conjunto de cámaras. Se determina un mapa de presencia  $P_{CP}^c$  para cada una de las cámaras  $c$  del conjunto  $C$ , como se ha indicado con anterioridad. Los mapas de presencia se pueden determinar a partir de imágenes de sustracción de fondo  $B_{SdF}$  o de imágenes de distancia  $D_{SdF}$ . El mapa de presencia  $P_{CP}^c$  debe tener en cuenta el hecho de que cada posición  $i$  del mapa  $P_{CP}^c$  no es necesariamente visible por todas las cámaras. De este modo, se puede determinar un operador  $v^c(i)$  de modo que indique, para cada posición  $i$ , si un modelo 3D  $M(i)$  se proyecta o no en la imagen de una cámara  $c$ . Para cada posición  $i$ , se define un

subconjunto  $C(i)$  del conjunto  $C$  que consta de las cámaras que proporcionan imágenes en las que las proyecciones  $m(i)$  son visibles correctamente. El mapa de presencia  $p_{CP}$  se puede determinar entonces a partir del conjunto de los mapas de presencia  $p_{CP}^c$ . La probabilidad  $p_{CP}(i)$  del mapa de presencia global en cada posición  $i$  es, por ejemplo, la media geométrica de las probabilidades  $p_{CP}^c(i)$ :

$$p_{CP}(i) = \sqrt{|C(i)| \prod_{c \in C(i)} p_{CP}^c(i)}$$

5 La imagen de presencia de objetos se puede determinar de una forma diferente de un tratamiento de sustracción de fondo. En particular, se puede determinar mediante un tratamiento de clasificación. Los datos de clasificación asociados a cada posición  $i$  se deben convertir en probabilidades con el fin de poder aprovecharse en el resto del procedimiento de localización. El documento "Galdi Giovanni, A. Prati, R. Cucchiara, Multi-stage Sampling with Boosting Cascades for Pedestrian Detection in Images and Videos. In Proceedings of ECCV (6)'2010, págs. 196-209" proporciona un ejemplo de una conversión de este tipo.

10 Según una forma de realización especialmente ventajosa, el mapa de presencia  $p_{CP}$  se determina en cada posición  $i$  a partir de varios mapas de presencia, por ejemplo el mapa de presencia  $p_{CP}^{SAF}$  obtenido mediante un tratamiento de sustracción de fondo y el mapa de presencia  $p_{CP}^{class}$  obtenido mediante un tratamiento de clasificación. Los diferentes mapas de presencia se combinan, por ejemplo, formando, posición por posición, el producto de las probabilidades asociadas a cada tratamiento. El mapa de presencia  $p_{CP}$  así obtenido es, por lo tanto, menos dependiente del tratamiento aplicado y, por lo tanto, más seguro.

15 El mapa de presencia  $p_{CP}$  proporciona una primera estimación de la presencia de objetos de interés en la escena observada. Esta primera estimación se obtiene a partir de un criterio de localización definido en el espacio de imagen de la cámara. El criterio de localización es, por ejemplo, un criterio asociado a un tratamiento de sustracción de fondo o un criterio asociado a un tratamiento de clasificación. El principal inconveniente es que la presencia de un objeto de interés en la escena en una posición dada corre el riesgo de provocar la detección errónea de objetos de interés en cada una de las posiciones situadas detrás de la posición efectiva. En efecto, la presencia de un objeto de interés en una posición dada puede provocar la ocultación de todo o parte de otros objetos de interés. Es, por lo tanto, necesaria una segunda fase en el procedimiento de localización de objetos de interés con el fin de tener en cuenta los fenómenos de ocultación y determinar la o las posiciones en la(s) que los objetos de interés están probablemente situados. Esta fase se caracteriza, en particular, en que la determinación de las posiciones efectivas de los objetos de interés se realiza en el espacio tridimensional del mapa de presencia.

20 La figura 3 ilustra el problema del enmascaramiento de los objetos entre sí. Un mapa 31 de presencia similar al de la figura 1 y dos modelos 32 y 33 3D están representados según un punto de vista de una cámara, es decir en el espacio de imagen de la cámara. El primer modelo 32 3D está localizado en una posición 34 del mapa 31 de presencia, y el segundo modelo 33 3D está situado en una posición 35. Los modelos 32 y 33 3D son unos elipsoides. De este modo, en el sistema de referencias de imagen de la cámara, los modelos 32 y 33 3D se proyectan en forma de elipses. Están situados de tal modo enfrentados a la cámara que sus proyecciones en el sistema de referencias de imagen convergen ampliamente. La zona de convergencia entre las dos elipses está representada por una zona sombreada. El objetivo de la segunda fase es, por lo tanto, suprimir del mapa de presencia la contribución de un modelo 3D considerado como presente en una posición dada en cada una de las demás posiciones.

25 La segunda fase 22 consta de las siguientes etapas. En una primera etapa 221, un modelo 3D representativo de un objeto de interés se coloca sucesivamente en cada posición  $i$ , y se determina su influencia sobre cada una de las demás posiciones en el espacio de imagen de la cámara. La influencia de un objeto de interés en una posición  $i$  dada sobre cada una de las demás posiciones, escritas como  $j$ , del mapa de presencia  $P_{CP}$  se modeliza mediante una entidad llamada "átomo"  $A_i$ . Se escribe como  $N$  el número de posiciones  $i$  o  $j$  en el mapa de presencia,  $M'(i)$  un modelo 3D representativo de un objeto de interés colocado en la posición  $i$ , y  $m'(i)$  su proyección en el espacio de imagen de la cámara. El modelo 3D  $M'(i)$  puede ser idéntico al modelo 3D  $M(i)$  utilizado en la primera fase 21. La proyección  $m'(i)$  en el espacio de imagen es en este caso idéntica a la proyección  $m(i)$ . Los átomos  $A_i$  se modelizan de manera ventajosa matemáticamente en forma de un vector para cada posición  $i$ , siendo el valor  $A_i(j)$  de la componente de un vector  $A_i$  representativo de la influencia de la presencia de un modelo 3D  $M'(i)$  en la posición  $i$  considerada sobre la posición  $j$ . El valor  $A_i(j)$  de un átomo  $A_i$  se determina, por ejemplo, mediante la siguiente relación:

$$A_i(j) = \frac{|m'(i) \cap m'(j)|}{m'(j)}$$

Esta relación define, para cada par de posiciones  $(i, j)$ , la relación entre el número de píxeles en común entre dos proyecciones  $m'(i)$  y  $m'(j)$  de modelos 3D  $M'(i)$  y  $M'(j)$ , y el número de píxeles de la proyección  $m'(i)$  del modelo 3D  $M'(i)$ . La relación se ilustra en la figura 3 mediante las dos elipses y su zona de convergencia. La primera etapa 221 se puede realizar de forma previa a cualquier otra etapa del procedimiento. En el marco de un tratamiento en tiempo real sobre un flujo de imágenes, la etapa 221 puede realizarse una única vez para el conjunto de las imágenes, siendo los átomos  $A_i(j)$  independientes de las imágenes. En una segunda etapa 222, se determina, en el mapa de presencia, la posición  $n$  en la que un objeto de interés tiene las mayores posibilidades de encontrarse. Se considera entonces que un objeto de interés se encuentra en esta posición  $n$ . La posición  $n$  se puede determinar buscando la más alta probabilidad  $P_{CP}(i)$  en el mapa de presencia  $P_{CP}$ . Se trata entonces de buscar la probabilidad máxima en el mapa de presencia. La determinación de la posición  $n$  en la que un objeto de interés tiene el máximo de posibilidades de encontrarse puede constar de una etapa previa que consiste en promediar las probabilidades  $P_{CP}(i)$  en un entorno local. Esta etapa previa permite alisar las probabilidades  $P_{CP}(i)$  anormalmente elevadas en algunas posiciones, por ejemplo a causa del tratamiento de sustracción de fondo. La búsqueda de una posición  $n$  también puede ser objeto de otras etapas previas que pretenden afinar las probabilidades  $P_{CP}(i)$  del mapa de presencia  $P_{CP}$ . En una tercera etapa 223, se determina un nuevo mapa de presencia suprimiendo del mapa de presencia inicial la contribución aportada por un objeto 3D en posición  $n$  en cada posición  $i$ .

Las etapas 222 y 223 se pueden reiterar de modo que se localicen todos los objetos de interés en la escena. En la figura 2, la repetición de las etapas 222 y 223 se realiza por medio de un contador de valor  $k$ , estando el mapa de presencia indizado por el valor  $k$  ( $p_{CP}^k$ ). En una etapa de inicialización 224, realizada a continuación de la etapa 221 o, llegado el caso, de la etapa 213, el valor  $k$  se pone en el valor '1'. El mapa de presencia determinado en la etapa 213 se escribe, por lo tanto,  $p_{CP}^1$ . La primera iteración de la etapa 222 se realiza a continuación de la etapa 224 con el mapa de probabilidades  $p_{CP}^1$ . La etapa 223 se realiza a continuación de cada iteración de la etapa 222. Para una iteración  $k$  dada, las probabilidades del nuevo mapa de presencia, escrito  $p_{CP}^{k+1}$ , se determinan en función del mapa de presencia  $p_{CP}^k$  y de la última posición  $n^k$  en la que un objeto de interés se ha considerado como presente. A continuación de cada etapa 223, el valor  $k$  del contador se incrementa en una unidad en una etapa 225. Para una iteración  $k$  dada, la etapa 223 se puede modelizar mediante la siguiente relación:

$$\forall i \in N, p_{CP}^{k+1}(i) = p_{CP}^k(i) - A_{n^k}(i)$$

En la práctica, una simple sustracción no da buenos resultados. En efecto, un átomo representa la influencia de un modelo 3D, por definición teórica, sobre otros modelos 3D idénticos. Esta influencia es asimilable a una probabilidad de presencia. Por lo tanto, la influencia es perfecta, mientras que el mapa de presencia determinado a partir de criterios de imagen no es perfecto. En consecuencia, se debe ponderar el átomo  $A_{n^k}(i)$  con el fin de tener en cuenta este fenómeno. El peso se determina de modo que la probabilidad  $p_{CP}^{k+1}(n^k)$  tome el valor nulo o casi nulo, siendo total la interacción entre dos modelos colocados en la misma posición. Las probabilidades  $p_{CP}^{k+1}(i)$  de las posiciones vecinas a la posición  $n^k$  disminuyen, mientras que las probabilidades de las posiciones más alejadas se mantienen sin cambios. De manera más general, la etapa 223 consiste en determinar las probabilidades  $p_{CP}^{k+1}(i)$  del nuevo mapa de presencia, inferiores al anterior, en función de la interacción entre un modelo 3D colocado en la posición  $n^k$  y un modelo 3D colocado sucesivamente en cada posición  $i$  del mapa de presencia. Se realiza una etapa 226 después de la etapa 225. Esta consiste en determinar si unos objetos de interés aun pueden encontrarse en la escena. Para ello, la probabilidad ( $\max_N p_{CP}^k(i)$ ) máxima en el mapa de presencia considerado se compara con un umbral  $S$  (del francés *seuil*) predeterminado. Este umbral  $S$  es, por ejemplo, igual a 0,5. Si la probabilidad máxima es superior al umbral, las etapas 222, 223, 225 y 226 se reiteran con el fin de determinar la nueva posición  $n^k$  en el nuevo mapa de presencia en la que un objeto de interés tiene más probabilidades de encontrarse. Si la probabilidad máxima es inferior al umbral, esto significa que todos los objetos de interés en la escena han tenido que localizarse. Por lo tanto, se termina el procedimiento de localización en una etapa 227. La etapa 226 podría, por ejemplo, como alternativa, realizarse determinando si una al menos de las probabilidades  $p_{CP}^{k+1}(i)$  en el mapa de presencia considerado es superior al umbral  $S$ .

Según una forma particular de realización, la etapa 223 de determinación de las probabilidades  $p_{CP}^{k+1}(i)$  del mapa de presencia  $p_{CP}^{k+1}$  se realiza mediante una minimización de mínimos cuadrados entre las probabilidades  $p_{CP}^1(i)$  del mapa de presencia de la iteración inicial y una combinación lineal de los átomos  $A_{n^k}$  que corresponden a las posiciones  $n^k$  en las que unos objetos de interés se consideran como presentes en la iteración  $k$  actual.

Con respecto a los métodos de la técnica, el procedimiento de localización de objetos según la invención permite determinar las posiciones de los objetos de interés mediante unas operaciones realizadas directamente en el

espacio del mapa de presencia, sin pasar por el espacio de imagen. Los átomos que modelizan las influencias de la presencia de objetos de interés entre las diferentes posiciones se pueden determinar de forma previa. De este modo, a partir de un mapa de presencia inicial, la localización de objetos de interés se realiza en un espacio con unas dimensiones iguales al número de posiciones en el mapa de presencia. Por otra parte, el procedimiento de localización según la invención es un procedimiento denominado constructivo, es decir que parte de una configuración inicial en la que ningún objeto de interés está presente, determinándose de forma iterativa las posiciones en las que un objeto está efectivamente presente a partir de los máximos de probabilidad en los diferentes mapas de presencia calculados durante el algoritmo. Por lo tanto, se determinan unas posiciones correctas a medida que se realizan las iteraciones sin un cuestionamiento posterior. A la inversa, los métodos de la técnica se denominan degresivos, es decir que estos parten de una configuración en la que unos objetos se consideran presentes en todas las posiciones posibles del mapa de presencia, eliminándose de forma iterativa las posiciones poco probables.



REIVINDICACIONES

1. Procedimiento de localización de objetos de interés en una escena (10) observada por un sistema (11) de adquisición de imágenes, estando los objetos de interés localizados a partir de un mapa (12) de presencia inicial  $p_{CP}^1$  que modeliza unas posiciones  $i$  en la escena y que comprende, para cada posición  $i$ , un valor  $p_{CP}^1(i)$  representativo de la probabilidad de que un objeto de interés se encuentre en la posición  $i$  considerada, obteniéndose cada valor  $p_{CP}^1(i)$  a partir de un criterio de localización definido en un espacio de imagen del sistema de adquisición de imágenes,
- estando **caracterizado** el procedimiento **porque** comprende una iteración de las etapas sucesivas siguientes mientras uno al menos de los valores  $p_{CP}^k(i)$  del mapa de presencia  $p_{CP}^k$  considerado para la iteración actual  $k$  es superior a un umbral predeterminado:
- determinar (222) la posición  $n^k$  en el mapa de presencia  $p_{CP}^k$  para la cual el valor  $p_{CP}^k(n_k)$  es máximo, considerándose un objeto de interés como presente en dicha posición  $n^k$ ,
  - a partir de átomos  $A_j$  predeterminados para cada posición  $j$  del mapa de presencia  $p_{CP}^k$ , el átomo  $A_j$  de una posición  $j$  que comprende, para cada posición  $i$ , un valor  $A_j(i)$  representativo del recubrimiento entre la proyección  $m(i)$  en el espacio de imagen de un modelo (32) tridimensional  $M(i)$  colocado en la posición  $i$  y la proyección  $m'(j)$  en el espacio de imagen de un modelo (33) tridimensional  $M'(j)$  colocado en la posición  $j$ , siendo cada modelo tridimensional representativo de un objeto de interés, determinar (223) un mapa de presencia  $p_{CP}^{k+1}$  donde cada valor  $p_{CP}^{k+1}(j)$  es igual al valor correspondiente del mapa de presencia inicial  $p_{CP}^1$  o de uno de los mapas de presencia determinado en una iteración anterior, disminuido de una combinación lineal de los valores  $A_{nk}(j)$  de los átomos  $A_{nk}$  que corresponden a las posiciones  $n^k$  en las que unos objetos de interés son considerados como presentes.
2. Procedimiento según la reivindicación 1, en el que cada valor  $A_j(j)$  de un átomo  $A_j$  se determina como siendo la relación entre el número de píxeles comunes entre las proyecciones  $m(i)$  y  $m'(j)$ , y el número de píxeles de la proyección  $m'(j)$ .
3. Procedimiento según una de las reivindicaciones 1 y 2, en el que la etapa (223) de determinación del mapa de presencia  $p_{CP}^{k+1}$  se realiza mediante una minimización de mínimos cuadrados entre, por una parte, las probabilidades  $p_{CP}^k(i)$  del mapa de presencia de la iteración actual  $k$  o las probabilidades  $p_{CP}^1(i)$  del mapa de presencia inicial  $p_{CP}^1$  y, por otra parte, una combinación lineal de los átomos  $A_{nk}$  que corresponde a las posiciones  $n^k$  en las que unos objetos de interés se consideran como presentes en la iteración  $k$  actual.
4. Procedimiento según una de las reivindicaciones anteriores, en el que el mapa de presencia inicial  $p_{CP}^1$  se determina mediante las siguientes etapas:
- determinar (211) una imagen de proyección  $B_{m(i)}$  para cada posición  $i$  del mapa de presencia inicial  $p_{CP}^1$ , presentando cada imagen de proyección  $B_{m(i)}$  unas dimensiones iguales a las de una imagen del sistema (11) de adquisición de imágenes y que corresponden a la proyección  $m(i)$  en el espacio de imagen de un modelo tridimensional  $M(i)$  representativo de un objeto de interés colocado en la posición  $i$  considerada, tomando cada píxel de la imagen de proyección  $B_{m(i)}$  un primer valor cuando pertenece a la proyección  $m(i)$  y un segundo valor en caso contrario,
  - determinar (212), mediante un tratamiento de localización de objetos realizado a partir de la imagen del sistema (11) de adquisición de imágenes, una imagen de presencia de objetos que tiene unas dimensiones iguales a las de la imagen del sistema (11) de adquisición de imágenes, tomando cada píxel de la imagen de presencia de objetos un valor en función de su correspondencia con un objeto presente en la escena,
  - determinar (213) el mapa de presencia inicial  $p_{CP}^1$  en función de las imágenes de proyección  $B_{m(i)}$  y de la imagen de presencia de objetos.
5. Procedimiento según la reivindicación 4, en el que el tratamiento de localización de objetos es un tratamiento de sustracción de fondo, tomando cada píxel de la imagen de presencia de objetos un primer valor cuando corresponde a un objeto presente en la escena y un segundo valor en caso contrario.
6. Procedimiento según la reivindicación 5, en el que la etapa (213) de determinación del mapa de presencia inicial  $p_{CP}^1$  comprende una etapa que consiste en determinar cada valor  $p_{CP}^1(i)$  como siendo igual a la relación entre el

número de píxeles que tienen a la vez el primer valor en la imagen de presencia de objetos y el primer valor en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada, y el número de píxeles que tienen el primer valor en dicha imagen de proyección  $B_{m(i)}$ .

5 7. Procedimiento según la reivindicación 4, en el que el tratamiento de localización de objetos es un tratamiento de sustracción de fondo, tomando cada píxel de la imagen de presencia de objetos un valor en función de la probabilidad de que corresponda a un objeto presente en la escena.

10 8. Procedimiento según la reivindicación 7, en el que la etapa (213) de determinación del mapa de presencia inicial  $p_{CP}^1$  comprende una etapa que consiste en determinar cada valor  $p_{CP}^1(i)$  como siendo igual a la relación entre la suma de los valores de los píxeles de la imagen de presencia de objetos que corresponde a los píxeles que tienen el primer valor en la imagen de proyección  $B_{m(i)}$  de la posición  $i$  considerada, y el número de píxeles que tienen el primer valor en dicha imagen de proyección  $B_{m(i)}$ .

15 9. Procedimiento según una de las reivindicaciones anteriores, en el que la escena (10) se observa mediante varios sistemas (11) de adquisición de imágenes, determinándose el mapa de presencia inicial  $p_{CP}^1$  a partir de varios mapas de presencia iniciales  $p_{CP}^E$ , obteniéndose cada valor  $p_{CP}^E(i)$  de un mapa de presencia inicial  $p_{CP}^E$  a partir de un criterio de localización definido en un espacio de imagen del sistema de adquisición de imágenes considerado, siendo cada valor  $p_{CP}^1(i)$  igual a una media del conjunto de los valores  $p_{CP}^E(i)$  de la posición  $i$  considerada.

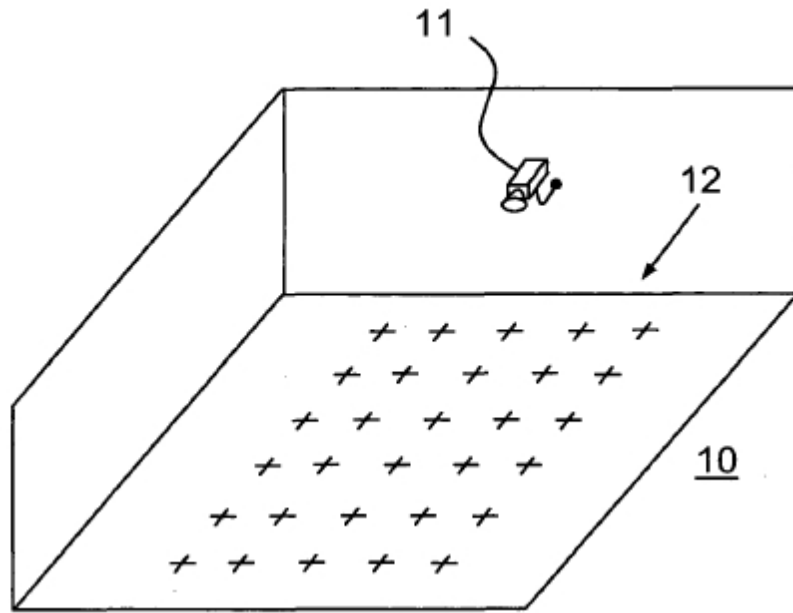


FIG. 1

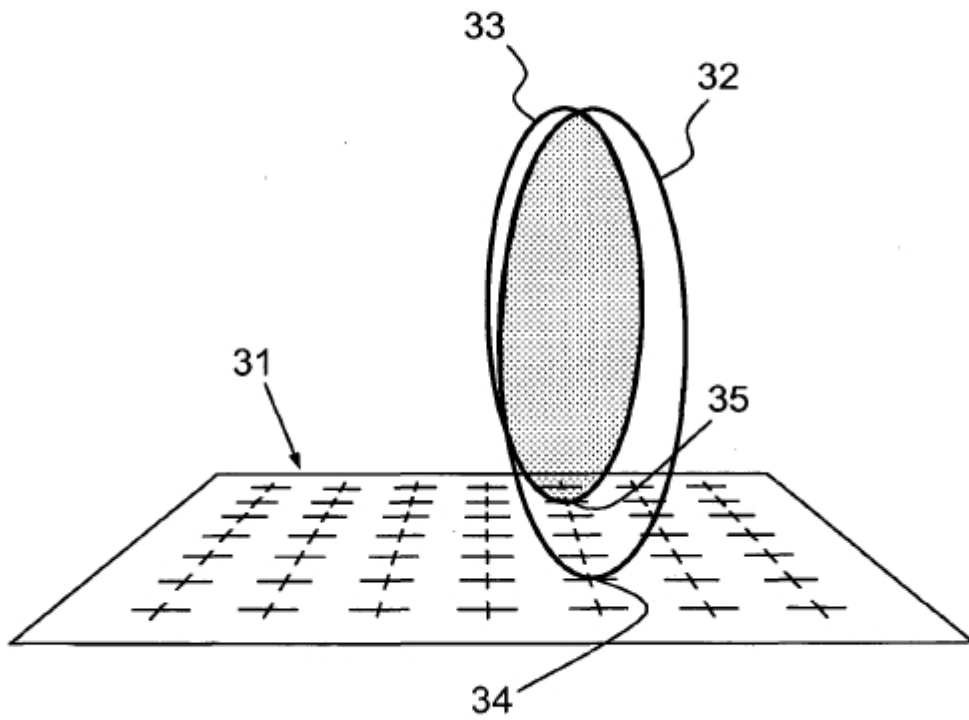


FIG. 3

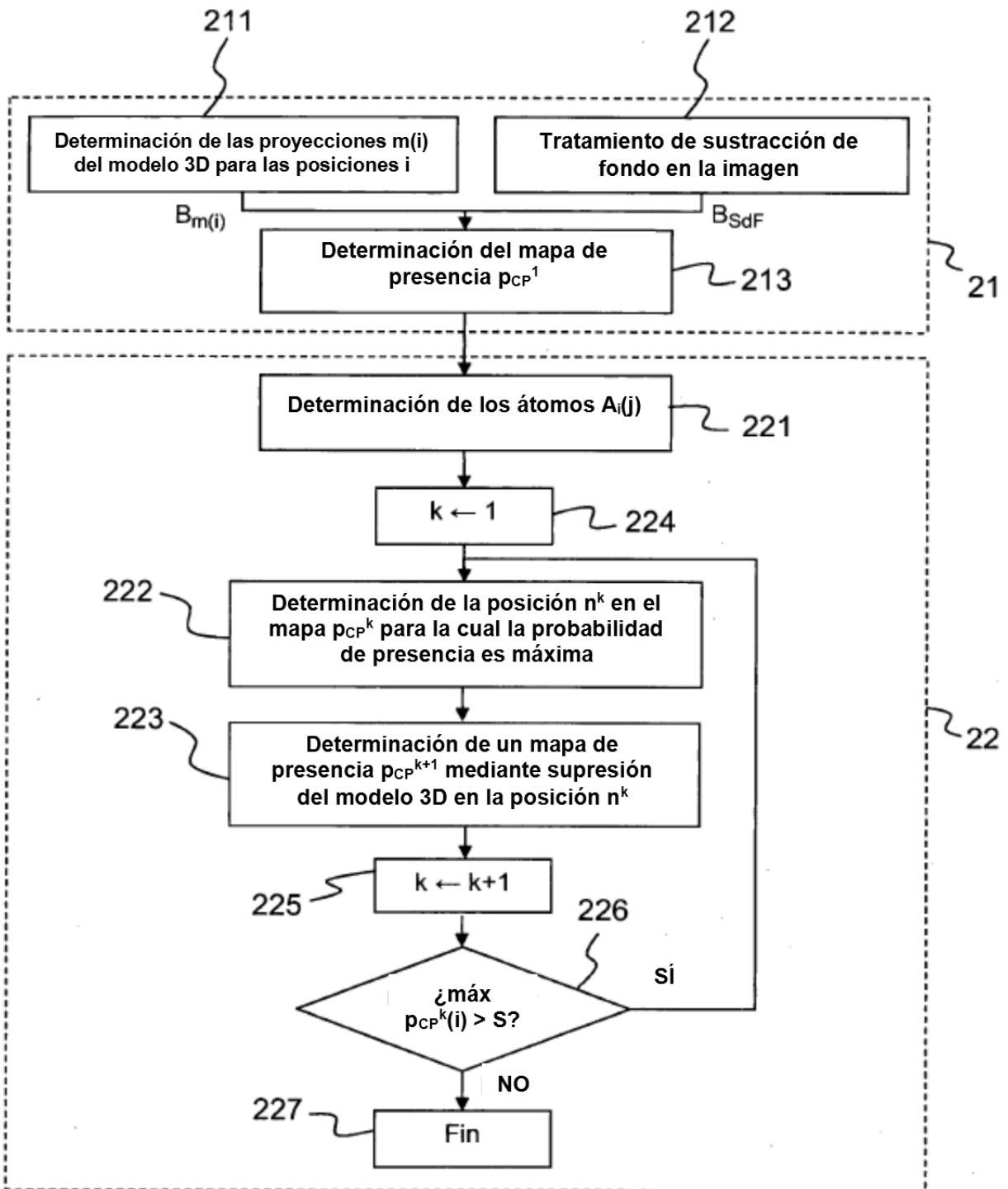


FIG.2