

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 973 497**

51 Int. Cl.:

G06V 40/12 (2012.01)

G06V 10/24 (2012.01)

G06V 10/82 (2012.01)

G06V 10/44 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.10.2019** **E 19306348 (4)**

97 Fecha y número de publicación de la concesión europea: **20.12.2023** **EP 3640843**

54 Título: **Método de extracción de características de una huella dactilar representada por una imagen de entrada**

30 Prioridad:

19.10.2018 FR 1859682

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

20.06.2024

73 Titular/es:

IDEMIA IDENTITY & SECURITY FRANCE (100.0%)
2 Place Samuel de Champlain
92400 Courbevoie, FR

72 Inventor/es:

MABYLAHT, GUY y
KAZDAGHLI, LAURENT

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 973 497 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

Método de extracción de características de una huella dactilar representada por una imagen de entrada

5 CAMPO TÉCNICO GENERAL

La presente invención se refiere al campo de la biometría, y propone en particular un método para extraer características buscadas de una huella dactilar representada por una imagen de entrada, con vistas a un tratamiento biométrico de la imagen de entrada.

10 ESTADO DE LA TÉCNICA

La autenticación/identificación biométrica consiste en reconocer a un individuo basándose en sus rasgos biométricos, tales como las huellas dactilares (reconocimiento digital), el iris o el rostro (reconocimiento facial).

Los enfoques biométricos convencionales utilizan información característica del rasgo biométrico extraídas a partir de los datos biométricos adquiridos, denominada "rasgos", y el aprendizaje/clasificación se lleva a cabo a partir de la comparación de dichos características.

En particular, en el caso del reconocimiento digital, las imágenes de las yemas de los dedos se procesan para extraer características de una huella dactilar que pueden clasificarse en tres categorías:

- El nivel 1 define el patrón general de esta huella (una de las cuatro clases: bucle derecho, bucle izquierdo, arco y espiral), y la disposición general de las crestas (en particular, se obtiene un mapa de orientación conocido como "Matriz de Flujo de Crestas", o mapa RFM, que representa la dirección general de la cresta en cada punto de la huella).
- El nivel 2 define los puntos específicos de las huellas, conocidos como puntos característicos, que constituyen "eventos" a lo largo de las crestas (final de una cresta, bifurcación, etc.). Los enfoques clásicos de reconocimiento utilizan esencialmente estas características.
- El nivel 3 define informaciones más complejas, tales como la forma de las crestas, los poros de la piel, las cicatrices, etc.

El proceso de extracción de las características de una huella (en forma de mapas de rasgos) se denomina "codificación". Estos mapas de rasgos permiten componer una firma denominada "plantilla" que codifica la información útil para la fase final de clasificación. Más concretamente, la clasificación se realiza comparando los mapas de características obtenidos con uno o varios mapas de características de referencia asociados a individuos conocidos.

Actualmente existen "codificadores" que realizan, de manera eficaz, esta operación de extracción de características, es decir, algoritmos que llevan a cabo una serie de operaciones de procesamiento:

- Mejora de la imagen (mejora del contraste, reducción del ruido, etc.);
- Utilización de filtros específicos (Gabor a distintas resoluciones, derivadores, etc.);
- Utilización de métodos de decisiones (umbral de binarización, extracción de puntos, etc.).

Una posibilidad es el uso de redes neuronales, que ya se utilizan ampliamente para la clasificación de datos.

Tras una fase de aprendizaje automático (por lo general supervisado, es decir, sobre una base de datos de referencia ya clasificados), una red neuronal "aprende" y llega a ser capaz de aplicar por sí misma la misma clasificación a datos desconocidos.

Las redes neuronales convolucionales, o CNN (Convolutional Neural Networks) son un tipo de red neuronal en donde el patrón de conexión entre neuronas se inspira en la corteza visual de los animales. Por lo tanto, son especialmente adecuadas para un tipo concreto de clasificación, que es el análisis de imágenes, lo que permite reconocer de manera eficaz objetos o personas en imágenes o vídeos, sobre todo en aplicaciones de seguridad (vigilancia automática, detección de amenazas, etc.).

Y, en el campo de la autenticación/identificación biométrica, una red CNN puede entrenarse para reconocer a un individuo basándose en sus rasgos biométricos, en la medida en que estos datos se manipulen en forma de imágenes.

Sin embargo, aunque estos enfoques han dado lugar a importantes avances en el reconocimiento facial, por ejemplo, su aplicación al reconocimiento de huellas dactilares se topa con las características específicas inherentes a las huellas dactilares. Mientras que la mayoría de las huellas dactilares se adquieren en un entorno controlado y con la

orientación correcta, en algunos casos (imágenes de huellas dactilares latentes procedentes de la escena de un crimen, imágenes adquiridas desde un teléfono inteligente, etc.) la adquisición y, en concreto, la orientación, no están controladas. Y aunque sigue siendo posible extraer características de una huella dactilar mal orientada, el cotejo (es decir, la comparación con las características de una huella dactilar de referencia) es difícil o incluso imposible si las huellas dactilares representadas por la imagen de entrada y las huellas dactilares de referencia no están en el mismo marco de referencia.

Por ejemplo, en una evaluación basada en huellas latentes, utilizando un extractor/matcher desarrollado para una huella perfectamente registrada, se observó un deterioro del rendimiento, del 36% al 0%, a medida que aumentaban las diferencias de orientación entre la colocación de las huellas latentes y la orientación de la imagen de la huella de referencia.

Conviene señalar que es conocido el registro de una imagen de huella dactilar como "preprocesamiento" mediante algoritmos de procesamiento de imágenes, pero este enfoque no es muy eficaz y requiere el uso de una imagen de referencia (o plantilla).

También conviene señalar que es conocido el uso de las redes CNN de "detección", tales como R-CNN, su derivado Faster R-CNN (véase el documento Fingerprint Pose Estimation Based on Faster R-CNN, Jiahong Ouyang et al) o incluso YOLO, para estimar la colocación de una huella, es decir, la posición de su centro y su orientación. Más concretamente, una red CNN de detección es capaz de detectar elementos de interés (objetos) de varias categorías en una imagen en forma de "cuadro delimitador": un primer módulo (del tipo VGG-16 en Faster R-CNN), de nuevo un módulo de "preprocesamiento", extrae zonas candidatas de la imagen de entrada como posibles cuadros delimitadores de un elemento de interés, y un módulo de "postprocesamiento" selecciona/clasifica los cuadros delimitadores más probables (es decir, determina la categoría del elemento de interés en el cuadro). Para utilizar una CNN de detección para estimar una orientación, basta con considerar varias orientaciones de la huella dactilar como diferentes categorías de elementos de interés. De este modo, una vez identificado el mejor cuadro delimitador para una huella en la imagen de entrada, se extraen el centro del cuadro delimitador predicho y la categoría asociada para derivar la colocación correspondiente.

Sin embargo, este enfoque "indirecto" sigue siendo muy pesado en cuanto a la potencia necesaria, y los bloques de procesamiento de imágenes tales como VGG-16 o VGG-19 son especialmente masivos. Además, los resultados de un enfoque de este tipo pueden mejorarse, sobre todo si la huella dactilar no es de buena calidad.

Schuch, "Unsupervised Learning of Fingerprint Rotations", 2018, esbozó una red CNN para la estimación de la rotación de huellas dactilares, pero que requiere 19 capas, por lo que es a nivel de cálculos costosa.

Por lo tanto, sería deseable disponer de una solución más potente para extraer características de una huella dactilar.

DESCRIPCIÓN DE LA INVENCION

Según un primer aspecto, la presente invención se refiere a un método de extracción de características buscadas a partir de una huella dactilar representada por una imagen de entrada, caracterizándose el método porque comprende la forma de realización, por medios de procesamiento de datos de un equipo cliente, de las etapas de:

- (a) estimación de al menos una desviación angular candidata de una orientación de dicha imagen de entrada con respecto a una orientación de referencia, mediante una red neuronal de convolución, CNN;
- (b) registro de dicha imagen de entrada en función de dicha desviación angular candidata estimada, de modo que la orientación de la imagen registrada coincida con dicha orientación de referencia;
- (c) procesamiento de dicha imagen recalibrada para extraer las características buscadas de la huella dactilar representada por dicha imagen de entrada.

Según otras características ventajosas y no limitativas:

- dicha red CNN comprende un conjunto de capas de convolución sucesivas que tienen un tamaño de filtro decreciente y un número de filtros también decreciente;
- la etapa (a) comprende la identificación de al menos una clase potencial de desviación angular de la orientación de dicha imagen de entrada con respecto a la orientación de referencia a partir de una pluralidad de clases potenciales predeterminadas de desviación angular, estando cada clase potencial de desviación angular asociada a un valor de desviación angular representativo de la clase, teniendo la(s) desviación(es) angular(es) candidata(s) estimada(s) como valores el(los) valor(es) representativo(s) de la(s) clase(s) identificada(s);
- cada clase está definida por un intervalo de valores de desviación angular;

- dichos intervalos son continuos y forman una partición de un margen dado de posibles desviaciones angulares de orientación;
- dichos intervalos son todos de la misma magnitud;
- la etapa (a) comprende la determinación de un vector de orientación de dicha imagen de entrada asociando a cada una de dicha pluralidad de clases potenciales de desviación angular, con una puntuación representativa de la probabilidad de que dichos datos de entrada pertenezcan a dicha clase de desviación angular potencial;
- el método comprende una etapa previa (a0) de aprendizaje, por medios de procesamiento de datos de un servidor, a partir de una base de imágenes de huellas dactilares, cada una ya asociada a una clase de dicha pluralidad de clases potenciales de desviación angular predeterminada, de parámetros de dicha red CNN;
- dicho aprendizaje utiliza una función de coste de tipo Sigmoide;
- en la etapa (b) se genera una imagen recalibrada para cada desviación angular candidata estimada, realizándose la etapa (c) para cada imagen recalibrada;
- dicha red CNN es del tipo red residual;
- la etapa (c) se realiza también sobre la imagen de entrada no registrada;
- dichas características que se desea extraer de la huella dactilar representada por dicha imagen de entrada comprenden la posición y/o la orientación de los puntos característicos;
- dicha huella dactilar representada por la imagen de entrada es la de un individuo, y el método comprende, además, una etapa (d) de identificación o autenticación de dicho individuo mediante la comparación de las características buscadas extraídas de la huella dactilar representada por dicha imagen de entrada con las características de la huella dactilar de referencia.

Según un segundo y tercer aspecto, la invención propone un producto de programa informático que comprende instrucciones de código para ejecutar un método según el primer aspecto de extracción de características buscadas a partir de una huella dactilar representada por una imagen de entrada; y un medio de almacenamiento legible por un equipo informático en donde un producto de programa informático comprende instrucciones de código para ejecutar un método según el primer aspecto de extracción de características buscadas a partir de una huella dactilar representada por una imagen de entrada.

PRESENTACIÓN DE LAS FIGURAS

Otras características y ventajas de la presente invención se pondrán de manifiesto a partir de la siguiente descripción de una forma de realización preferida. Esta descripción se hará con referencia a los dibujos adjuntos en los que:

- la Figura 1 es un diagrama esquemático de una arquitectura para poner en práctica el proceso según la invención;
- la Figura 2 muestra un ejemplo de red neuronal de convolución para aplicar el método según la invención;
- la Figura 3 muestra un ejemplo de vector de orientación obtenido mediante dicha red neuronal de convolución;

las Figuras 4a-4b ilustran dos pruebas de rendimiento para diferentes formas de realización de un método según la invención.

DESCRIPCIÓN DETALLADA

Principio y arquitectura

El presente método propone un proceso para extraer las características buscadas de una huella dactilar representada por una imagen de entrada. Este método consiste típicamente en "codificar" la huella dactilar, es decir, dichas características buscadas que deben extraerse son por lo general características "biométricas", es decir, las características "finales" que permiten componer una plantilla de la huella dactilar con vistas a la clasificación (identificación/autenticación de un individuo por comparación de las características buscadas extraídas de la huella dactilar representada por dicha imagen de entrada, con las características de las huellas dactilares de referencia, véase más adelante). A este respecto, dichas características buscadas describen por lo general los puntos característicos, es decir, incluyen la posición y/o la orientación de los puntos característicos. Sin embargo, se entenderá que el presente método no se limita a esta forma de realización, y que todas las características de posible interés en biometría pueden extraerse como resultado de este método.

El presente método se distingue en que propone una etapa (a) de estimación directa (a partir de la imagen de entrada) de al menos una desviación angular candidata de una orientación de dicha imagen de entrada con respecto a una orientación de referencia, mediante una red neuronal de convolución, CNN. Se volverá sobre la noción de estimación "directa" más adelante; se entiende sin ningún preprocesamiento de la imagen de entrada.

Tal como ya se ha mencionado con anterioridad, el cotejo (es decir, la correspondencia con las características de una huella dactilar de referencia) resulta difícil en huellas dactilares mal orientadas.

La idea del presente método es permitir, en una etapa (b), la recalibración de dicha imagen de entrada en función de dicha desviación angular candidata estimada, de modo que la orientación de la imagen recalibrada coincida con dicha orientación de referencia.

De este modo, la extracción de características puede realizarse sobre una imagen recalibrada "bien orientada", para obtener un proceso sólido a las variaciones de orientación.

Se entiende que la orientación de "referencia" corresponde a una orientación elegida de manera arbitraria, tal como aquella en donde el dedo está vertical y apuntando hacia arriba, conocida como "Norte-Sur", es decir, la orientación natural cuando el dedo se presiona contra un sensor en buenas condiciones, corresponde a la orientación habitual de las bases de huellas dactilares de referencia. Dicho de otro modo, la orientación de referencia se fija preferiblemente en relación con la orientación del dedo, en particular igual a la orientación del dedo, pero se entenderá que puede utilizarse cualquier orientación de referencia como punto de partida.

La desviación angular de la orientación puede expresarse como un valor angular, por ejemplo, en la dirección trigonométrica (positiva cuando la imagen gira en sentido antihorario, negativa en caso contrario).

Por ejemplo, puede elegirse en un margen $[0^\circ, 360^\circ]$ o en un margen $[-180^\circ, +180^\circ]$ de forma equivalente. Además, se prefiere un intervalo de 360° en la medida en que corresponde a todas las orientaciones posibles, pero se entenderá que se puede trabajar con un intervalo más pequeño que corresponda a la mayoría de las orientaciones observadas, por ejemplo, $[-45^\circ, +45^\circ]$. En efecto, en una aplicación tal como el sensor de huellas dactilares de un teléfono inteligente, es extremadamente raro tener una desviación de orientación de 180° , es decir, una huella dactilar completamente al revés, pero es muy común tener desviaciones de $10^\circ - 20^\circ$ en valor absoluto respecto a la orientación de referencia. De manera alternativa, en el análisis de huellas latentes (escena de un crimen, por ejemplo), puede encontrarse cualquier orientación de la huella.

Tal como se ha explicado con anterioridad, es conocido el registro de una imagen de huella dactilar como "preprocesamiento" mediante algoritmos de procesamiento de la imagen, pero se ha descubierto que es posible realizar este registro de forma muy eficiente con redes neuronales, pero sobre todo sin necesidad de utilizar una imagen (o modelo) de referencia.

También es conocido el uso de las redes CNN de "detección", tales como Faster R-CNN (véase el documento Fingerprint Pose Estimation Based on Faster R-CNN, Jiahong Ouyang et al.) o YOLO, para estimar la colocación de una huella, es decir, la posición de su centro y su orientación. Sin embargo, este enfoque "indirecto" sigue siendo muy engorroso en cuanto a la potencia necesaria y sus resultados son mejorables, sobre todo si la huella no es de buena calidad. Tal como se verá más adelante, otra ventaja del presente método es que las desviaciones de orientación pueden extraerse de forma automática, sin ninguna información a priori sobre su naturaleza, y sobre todo incluso cuando la huella está degradada o tiene un tipo de "whorl" (es decir, un verticilo, también denominado espiral o vórtice) que no permite distinguir la izquierda de la derecha, a diferencia de un tipo de "arco".

El presente método se pone en práctica dentro de una arquitectura tal como la mostrada en la Figura 1, utilizando un servidor 1 y un cliente 2. El servidor 1 es el equipo de aprendizaje (que pone en práctica el aprendizaje de la red CNN) y el cliente 2 es un equipo de clasificación (que pone en práctica el presente método para extraer las características buscadas de una huella dactilar), por ejemplo, un terminal de usuario.

Es muy posible que los dos equipos 1, 2 se confundan, pero preferiblemente el servidor 1 es el de un proveedor de soluciones de seguridad, y el cliente 2 es un equipo personal de consumo, en particular un teléfono inteligente, un ordenador personal, una tableta táctil, una caja fuerte, etc.; o un equipo de adquisición de huellas dactilares latentes.

En todos los casos, cada equipo 1, 2 suele ser un ordenador remoto conectado a una red de zona extensa 10, tal como la red Internet, para el intercambio de datos. Cada uno comprende medios de procesamiento de datos 11, 21 del tipo procesador, y medios de almacenamiento de datos 12, 22 tal como una memoria informática, por ejemplo, una memoria instantánea o un disco duro.

El servidor 1 almacena una base de datos de aprendizaje, es decir, un conjunto de imágenes de huellas dactilares de las que se conoce la orientación (véase más adelante cómo representarla), en contraposición a las denominadas imágenes de entrada que en realidad se trata de procesar.

El equipo cliente 2 comprende, de manera ventajosa, un escáner de huellas dactilares 23, para poder adquirir directamente dicha imagen de entrada, en general para que un usuario pueda autenticarse.

5 Red CNN

Una red CNN contiene por lo general cuatro tipos de capas que procesan de manera sucesiva la información:

- la capa de convolución, que procesa bloques de la entrada uno tras otro;
- la capa no lineal, que pone en práctica una función de activación que permite añadir no linealidad a la red y, por tanto, disponer de funciones de decisión mucho más complejas;
- la capa de puesta en común (denominada "pooling", que permite reagrupar varias neuronas en una sola neurona;
- la capa totalmente conectada, que enlaza todas las neuronas de una capa con todas las neuronas de la capa anterior (para la clasificación).

Las capas no lineales suelen ir precedidas de una capa de normalización en lotes ("capa *BN*") antes de cada capa no lineal *NL*, para acelerar el aprendizaje.

La función de activación de la capa no lineal *NL* suele ser la función *ReLU* (Unidad Lineal Rectificada), que es igual a $f(x) = \max(0, x)$, y la capa de agrupación (denominada *POOL*) más utilizada es la función *AvgPool*, que corresponde a una media entre los valores de un marco (varios valores se agrupan en un único valor).

La capa de convolución, *CONV*, y la capa totalmente conectada, *FC*, corresponden por lo general a un producto escalar entre las neuronas de la capa anterior y los pesos de la red CNN.

En general, una red CNN suele comprender un conjunto de capas de convolución sucesivas. De forma conocida y tal como ya se ha explicado con anterioridad, cada una de dichas capas de convolución puede ir seguida de una capa de normalización en lotes *BN* y/o de una capa no lineal, en particular *ReLU*, preferiblemente ambas en este orden.

Las arquitecturas CNN típicas apilan unos cuantos pares de capas *CONV* → *NL*, a continuación añaden una capa *POOL* y repiten este patrón $[(CONV \rightarrow NL)^p \rightarrow POOL]$ hasta obtener un vector de salida suficientemente pequeño, y luego terminan con dos capas *FC* totalmente conectadas.

He aquí una arquitectura típica de CNN:

$$INPUT \rightarrow [(CONV \rightarrow NL)^p \rightarrow POOL]^n \rightarrow FC \rightarrow FC$$

De manera ventajosa, la determinación directa de la desviación angular de orientación se considera una clasificación "pura" de la imagen de entrada, es decir, se cuantifica el problema, sin que intervenga ninguna tarea de detección de objetos.

Dicho de otro modo, esta red CNN de clasificación no predice ningún cuadro delimitador ni estima directamente valores candidatos precisos de desviación angular de la orientación, sino que identifica directamente al menos una clase potencial de desviación angular de la orientación de dicha imagen de entrada con respecto a la orientación de referencia entre una pluralidad de clases potenciales predeterminadas de desviación angular, asociándose a cada clase potencial de desviación angular un valor de desviación angular representativo de la clase (que será el valor que se tendrá en cuenta cuando se identifique una clase). Por "directamente" se entiende, tal como se mencionado, que la presente red CNN asocia a la imagen de entrada una o varias clasificaciones sin procesamiento previo. Lo que antecede contrasta con las mencionadas de tipo Faster R-CNN en las que el bloque de detección (tal como VGG-16) propone cajas englobantes candidatas, cada una de las cuales se selecciona y clasifica, es decir, la clasificación se aplica a las cajas candidatas (es decir, de manera indirecta) y no directamente a la imagen en su conjunto. Naturalmente, cualquier red CNN produce un cierto número de estados internos que son mapas de características (Feature Maps), incluso en el presente caso de clasificación denominada "directa", pero dichos mapas de características forman parte del funcionamiento normal de cualquier red CNN, a diferencia de los resultados intermedios que son las cajas delimitadoras candidatas de una red CNN de detección.

También se entenderá que la presente red CNN solamente asocia a la imagen de entrada las clasificaciones (tal como se verá por lo general en forma de puntuación de confianza), pero no otras salidas tales como las coordenadas de los cuadros delimitadores o incluso una imagen anotada.

En particular, cada clase se define por un intervalo (preferiblemente continuo, pero posiblemente una unión discontinua de subintervalos continuos) de desviaciones angulares, y el valor de desviación angular representativo de la clase es entonces por lo general el valor medio del intervalo. Por ejemplo, un intervalo $[10^\circ, 20^\circ]$ puede asociarse al valor 15° .

El conjunto de intervalos forma preferentemente una partición de un margen de desviaciones angulares de orientaciones posibles (que repetimos puede ser de 360° o menos).

Según una primera posibilidad, los intervalos que definen las clases tienen todos el mismo tamaño. Por ejemplo, pueden considerarse 360 clases de 1 grado cada una (por ejemplo, $[2^\circ, 3^\circ]$ es un intervalo), o 60 clases de 6 grados cada una (por ejemplo, $[6^\circ, 12^\circ]$ que es un intervalo).

Según una segunda posibilidad, los intervalos que definen las clases tienen diferentes tamaños, en particular según una densidad de probabilidad. Por ejemplo, para una aplicación de tipo sensor de huellas dactilares de un teléfono inteligente es, tal como se ha explicado, mucho más probable tener una desviación de orientación del orden de magnitud de $10^\circ - 20^\circ$ que una desviación de $90^\circ - 100^\circ$. Por tanto, en una aplicación de este tipo, pueden preverse intervalos más estrechos para desviaciones pequeñas que para desviaciones grandes, por ejemplo, en un lado $[2^\circ, 4^\circ]$ y en el otro $[90^\circ, 105^\circ]$. El experto en esta técnica podrá construir cualquier conjunto de clases de desviaciones de orientación de su elección.

Se entiende que los valores de desviación angular candidatos estimados son valores teóricos que son funciones de las clases, y más concretamente, valores de desviación angular que son representativos de las clases. De este modo, para una imagen de huella dada, la red CNN identifica una o varias clases candidatas, es decir, definidas por un intervalo susceptible de contener el valor real de la desviación angular de orientación de esta imagen con respecto a la orientación de referencia (conviene señalar que pueden identificarse varias clases, véase más adelante, aunque en la práctica la desviación angular real tiene un único valor), y para cada clase identificada, el valor de desviación angular representativo de la clase se considera una desviación angular candidata estimada en la etapa (a).

Por ejemplo, en una forma de realización con 60 clases de igual tamaño, con una imagen de entrada que tiene una desviación angular de orientación real de 11° , entonces se identifica la clase correspondiente al intervalo $[6^\circ, 12^\circ]$, y si el valor representativo de esta clase es 9° (caso de la media), entonces el valor estimado de la desviación angular candidata es 9° .

Se puede identificar una pluralidad de clases si la red CNN emite un vector de orientación, es decir, una puntuación para cada una de las clases. Más concretamente, el vector de orientación de una imagen de entrada asocia a cada una de la pluralidad de clases una puntuación representativa de la probabilidad de que dichos datos de entrada pertenezcan a la clase de desviación angular potencial, es decir, la probabilidad de que la desviación angular de orientación real pertenezca al intervalo correspondiente. Por tanto, el vector tiene una dimensión igual al número de clases, es decir, 60 en nuestro ejemplo. Se entiende que una red CNN de clasificación tal como la definida con anterioridad devuelve preferentemente solamente el vector de orientación, a diferencia de una red CNN de detección.

Tal como se verá más adelante, la clase o clases identificadas pueden ser las k (por ejemplo, dos) con la puntuación más alta (es decir, se identifica un número fijo de clases), o todas aquellas con una puntuación superior a un umbral predeterminado (por ejemplo, probabilidad superior al 10%, o incluso superior al 20%, en función de la flexibilidad deseada y del "riesgo" de tener imágenes de entrada mal orientadas. Por ejemplo, para un escáner de huellas dactilares latentes, el umbral puede reducirse para maximizar el número de picos tomados en consideración debido a la posible multiplicidad de orientaciones plausibles).

Dicha red CNN puede ser de muchos tipos y, en particular, puede ser una red CNN convencional (sucesión directa de capas de convolución *CONV*, capas de normalización de lotes *BN* y capas no lineales *NL*).

Según una forma de realización preferida, dicha red CNN es del tipo red residual. Se puede observar que dichas redes residuales son muy eficaces para la única tarea de clasificar la imagen de entrada que se les asigna, y mucho más ligeras que las CNN de detección cuyos bloques de procesamiento de imágenes, tal como VGG-16 o VGG-19, son especialmente masivos.

En referencia a la Figura 2, una red residual, o RESNET, es una red CNN que tiene al menos una "conexión residual" (en inglés "skip connection" o simplemente "short-cut"), es decir, una conexión que "cortocircuita" al menos una capa, por analogía con lo que se observa en las neuronas piramidales del cerebro.

De hecho, cuando un modelo se hace más complejo añadiendo capas, algunas de estas capas pueden tener un impacto negativo en el rendimiento del modelo. Las conexiones residuales garantizan que si una capa no consigue aprender una transformación útil, en el peor de los casos debe aprender la identidad, lo que evita degradar el rendimiento de las demás capas. El principio básico de funcionamiento de las redes residuales consiste en añadir la entrada y la salida de una capa de convolución punto por punto, lo que permite propagar la señal de las capas menos profundas a las capas más profundas. Tal como se ha explicado, esta red proporciona excelentes resultados en la determinación directa de la desviación angular de la orientación, y en particular en la clasificación propiamente dicha.

En el ejemplo RESNET de la Figura 2, la rama principal (cortocircuitada por la conexión residual) suele tener una pluralidad de capas de convolución, cada una seguida de una capa de normalización en lotes y/o una capa no lineal del tipo *ReLU*. La salida de esta rama suele sumarse punto por punto con la imagen de entrada (debido a la conexión

residual), y pasa por una última capa de activación *ReLU*, con una reducción de la dimensionalidad (por lo general MaxPool).

Se utiliza una capa totalmente conectada (IP/FC) para generar el vector de orientación (es decir, la clasificación propiamente dicha), y se puede utilizar una función de pérdida del tipo Sigmoide o Softmax para el aprendizaje, véase a continuación.

Aprendizaje y clasificación

De manera ventajosa, el método comienza con una etapa (a0) de aprendizaje, por los medios de tratamiento de datos 11 del servidor 1, a partir de una base de imágenes de huellas dactilares etiquetadas (es decir, para las que se conoce la clase de desviación angular de orientación correspondiente), de los parámetros de dicha red CNN.

Este aprendizaje puede realizarse de forma convencional, minimizando la función de coste.

La principal dificultad del presente método consiste en extraer las características relevantes para la determinación de la desviación angular esperada, penalizando al mismo tiempo las características que no permiten discriminar las clases. Para ello, una primera posibilidad es utilizar una entropía cruzada sobre un Softmax como función de coste, ya que permite extraer características que maximizan la probabilidad de la desviación angular esperada y penalizan la probabilidad de otras desviaciones angulares (es decir, otras clases).

La función Softmax solamente favorece a una única clase; de manera alternativa, si queremos aumentar las probabilidades de que se identifiquen varias clases, se puede utilizar una Sigmoide como función de coste para favorecer no solamente la desviación angular de orientación esperada, sino también, con menor probabilidad, las que más se aproximen.

De este modo, si la imagen de entrada es parcial y tiene, por ejemplo, una parte (por ejemplo, un bucle) que puede estar situada, por ejemplo, a la derecha o a la izquierda de una huella, el vector de orientación tendrá dos puntuaciones pico distintas, véase, por ejemplo, la Figura 3, para tener en cuenta los dos casos posibles, cada uno de los cuales sigue siendo probable.

Conviene señalar que dicha base de imágenes de huellas dactilares ya etiquetadas puede obtenerse mediante el aumento de datos. Más concretamente, se parte de una base de referencia en donde todas las imágenes están orientadas según la orientación de referencia (es decir, desviación angular de orientación nula). A continuación, se generan de manera artificial otras versiones de estas impresiones de referencia aplicándoles rotaciones de varios ángulos determinados (y etiquetándolas en consecuencia), para crear nuevas imágenes de aprendizaje con diversas desviaciones de orientación. Por ejemplo, se les aplican ángulos de $\pm [5^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 70^\circ, 120^\circ]$, con el fin de crear una nueva base 14 veces mayor, para garantizar la solidez de la red CNN frente a los fallos de adquisición habituales.

Si es necesario, la red CNN objeto de aprendizaje puede almacenarse en los medios de almacenamiento de datos 22 del cliente 2 para utilizarla en la estimación de la orientación. Conviene señalar que la misma CNN puede incorporarse en numerosos clientes 2, siendo necesario un único aprendizaje.

En una etapa principal (a), tal como se ha explicado con anterioridad, al menos una desviación angular candidata de una orientación de dicha imagen de entrada con respecto a una orientación de referencia es estimada por los medios de procesamiento de datos 21 del cliente 2 mediante la red CNN incorporada.

A continuación, en una etapa (b), dicha imagen de entrada se recalibra en función de dicha desviación angular candidata estimada, de modo que la orientación de la imagen recalibrada coincida con dicha orientación de referencia. Conviene señalar que en la etapa (b) se puede generar una imagen recalibrada para cada desviación angular candidata estimada. La idea es que si surgen varios picos de probabilidad no existe forma de saber cuál es el correcto, y se prueba con todos. De este modo, una de las imágenes registradas estará "bien registrada", es decir, orientada de conformidad con la orientación de referencia.

Por último, en una etapa (c), dicha imagen recalibrada puede procesarse para extraer las características buscadas de la huella dactilar representada por dicha imagen de entrada, que pueden incluir en particular la posición y/u orientación de los puntos característicos. Si se han generado varias imágenes de registro, la etapa (c) se realiza para cada una de ellas. En particular, esta etapa (c) puede realizarse mediante una segunda red CNN dedicada, tomando como entrada cada imagen recalibrada. La primera red CNN se denomina CNN de registro y la segunda red CNN de codificación. En este caso, puede utilizarse cualquier red CNN de codificación conocida.

Conviene señalar que la etapa (c) también puede realizarse de manera sistemática sobre la imagen de entrada "tal cual", es decir, sin recalibración (dicho de otro modo, la imagen de entrada se añade con las imágenes recalibradas). De hecho, si la imagen de entrada ya está bien orientada, la red CNN puede tender a realinearla de todos modos, lo que puede causar una ligera disminución del rendimiento.

De manera preferible, el método comprende, además, una etapa (d) de identificación o autenticación de dicho individuo mediante la comparación de las características buscadas extraídas de la huella dactilar representada por dicha imagen de entrada con las características de la huella dactilar de referencia, que puede ponerse en práctica de cualquier manera conocida por el experto en esta técnica.

Por ejemplo, el cliente 2 puede almacenar las características de las huellas dactilares de uno o varios usuarios autorizados como huellas dactilares de referencia, con el fin de gestionar el desbloqueo del equipo cliente 2 (en particular en el caso de una imagen de entrada adquirida directamente por un escáner integrado 23): si las características extraídas corresponden a las esperadas de un usuario autorizado, los medios de tratamiento de datos 21 consideran que el individuo que intenta autenticarse está autorizado, y proceden al desbloqueo.

De manera alternativa, el cliente 2 puede enviar las características extraídas a una base de datos remota de dichas características de huellas dactilares de referencia, para la identificación del individuo.

Se han realizado diversas pruebas del presente método. La Figura 4a muestra el rendimiento de la red CNN de registro en la predicción de la desviación angular de la orientación en función del número medio de intentos, es decir, del número medio de imágenes registradas obtenidas. Se observa un aumento de la precisión de hasta casi el 90% cuando se pasa de un intento (es decir, solamente se tiene en cuenta el pico más fuerte), a 2,8 intentos de media.

La Figura 4b muestra el rendimiento del emparejamiento, en función de la desviación angular de la orientación real: puede observarse que, en ausencia de registro (una única CNN de codificación utilizada directamente para la extracción de características), el rendimiento se desploma tras unos pocos grados de desviación.

En cambio, las distintas configuraciones probadas, que incluyen todas una primera red CNN de registro y una segunda red CNN de codificación, y que difieren en el número medio de intentos, muestran que el rendimiento se mantiene sea cual sea la desviación angular de la orientación. El método que incluye también el uso sistemático de la imagen de entrada (no recalibrada) es ligeramente menos eficaz para grandes desviaciones de orientación (ya que la imagen de entrada es entonces irrelevante), pero más eficaz si la desviación angular de orientación es pequeña, lo que sigue siendo un caso frecuente.

Producto de programa informático

Según un segundo y un tercer aspecto, la invención se refiere a un producto de programa informático que comprende instrucciones de código para ejecutar (en particular en los medios de procesamiento de datos 11, 21 del servidor 1 y/o del cliente 2) de un método para extraer características buscadas de una huella dactilar representada por una imagen de entrada, así como medios de almacenamiento legibles por un equipo informático (una memoria 12, 22 del servidor 1 y/o del cliente 2) en donde se encuentra este producto de programa informático.

REIVINDICACIONES

- 5 **1.** Método de extracción de características buscadas a partir de una huella dactilar representada por una imagen de entrada, estando caracterizado dicho método porque comprende la puesta en práctica, mediante medios de procesamiento de datos (21), de un equipo cliente (2), de las etapas de:
 - 10 (a) mediante una red neuronal de convolución, CNN, del tipo red residual, la estimación directa tal que dicha CNN no pone en práctica ninguna tarea de detección de objetos en la imagen de entrada, a partir de la imagen de entrada, de al menos una desviación angular candidata de una orientación de dicha imagen de entrada con respecto a una orientación de referencia;
 - (b) registro de dicha imagen de entrada en función de dicha desviación angular candidata estimada, de forma que la orientación de la imagen registrada coincida con dicha orientación de referencia;
 - 15 (c) procesar dicha imagen realineada para extraer las características buscadas de la huella dactilar representada por dicha imagen de entrada.
- 2.** Método según la reivindicación 1, en donde la etapa (a) comprende la identificación directa de al menos una clase potencial de desviación angular de la orientación de dicha imagen de entrada con respecto a la orientación de referencia a partir de una pluralidad de clases potenciales predeterminadas de desviación angular, estando cada clase potencial de desviación angular asociada a un valor de desviación angular representativo de la clase, teniendo la(s) desviación(es) angular(es) candidata(s) estimada(s) como valores el(los) valor(es) representativo(s) de la(s) clase(s) identificada(s).
- 25 **3.** Método según la reivindicación 2, en donde cada clase está definida por un intervalo de valores de desviación angular.
- 4.** Método según la reivindicación 3, en donde dichos intervalos son continuos y forman una partición de un margen dado de posibles desviaciones angulares de orientación.
- 30 **5.** Método según la reivindicación 4, en donde dichos intervalos son todos de la misma magnitud.
- 6.** Método según una de las reivindicaciones 2 a 5, en donde la etapa (a) comprende la determinación de un vector de orientación de dicha imagen de entrada que asocia a cada una de dicha pluralidad de clases potenciales de desviación angular una puntuación representativa de la probabilidad de que dichos datos de entrada pertenezcan a dicha clase de desviación angular potencial.
- 35 **7.** Método según una de las reivindicaciones 2 a 6, que comprende una etapa previa (a0) de aprendizaje, por los medios de tratamiento de datos (11) de un servidor (1), a partir de una base de imágenes de huellas dactilares ya asociadas cada una de ellas a una clase de dicha pluralidad de clases potenciales de desviación angular predeterminada, de los parámetros de dicha CNN.
- 8.** Método según la reivindicación 7, en donde dicho aprendizaje utiliza una función de coste de tipo Sigmoide.
- 45 **9.** Método según una de las reivindicaciones 1 a 8, en donde se genera una imagen recalibrada en la etapa (b) para cada desviación angular candidata estimada, realizándose la etapa (c) para cada imagen recalibrada.
- 10.** Método según una de las reivindicaciones 1 a 9, en donde la etapa (c) se realiza también en la imagen no recalibrada.
- 50 **11.** Método según una de las reivindicaciones 1 a 10, en donde dichas características buscadas a extraer de la huella dactilar representada por dicha imagen de entrada comprenden la posición y/u orientación de los puntos característicos.
- 55 **12.** Método según una de las reivindicaciones 1 a 11, en donde dicha huella dactilar representada por la imagen de entrada es la de un individuo, comprendiendo, además, el método una etapa (d) de identificación o de autenticación de dicho individuo mediante la comparación de las características buscadas extraídas de la huella dactilar representada por dicha imagen de entrada, con las características de huellas dactilares de referencia.
- 60 **13.** Producto de programa informático que comprende instrucciones de código para ejecutar un método según una de las reivindicaciones 1 a 12, para extraer características buscadas de una huella dactilar representada por una imagen de entrada, cuando dicho programa se ejecuta en un ordenador.
- 65 **14.** Medio de almacenamiento legible por un equipo informático en donde un producto de programa informático comprende instrucciones de código para ejecutar un método según una de las reivindicaciones 1 a 12, para extraer características buscadas de una huella dactilar representada por una imagen de entrada.

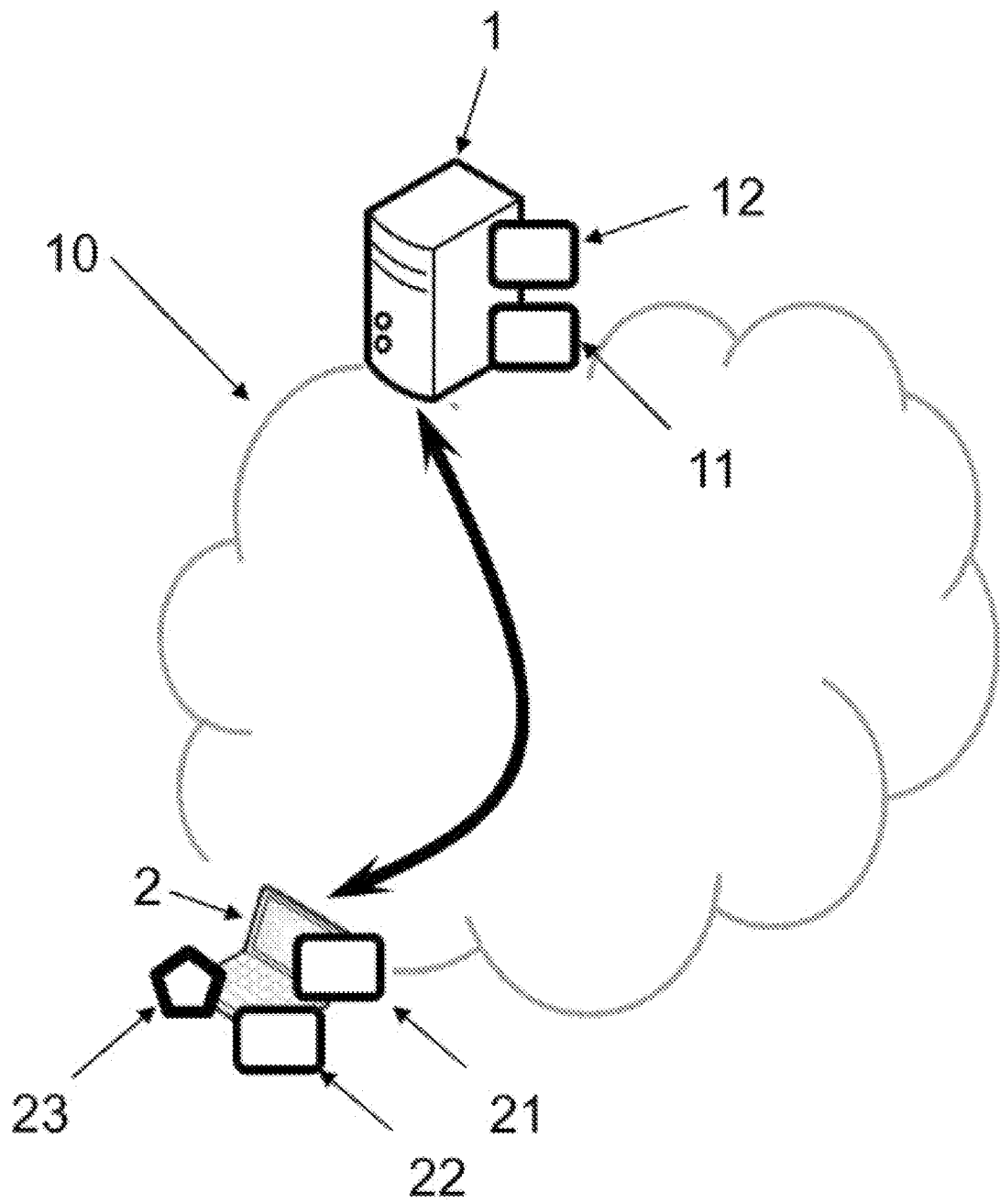


FIG. 1

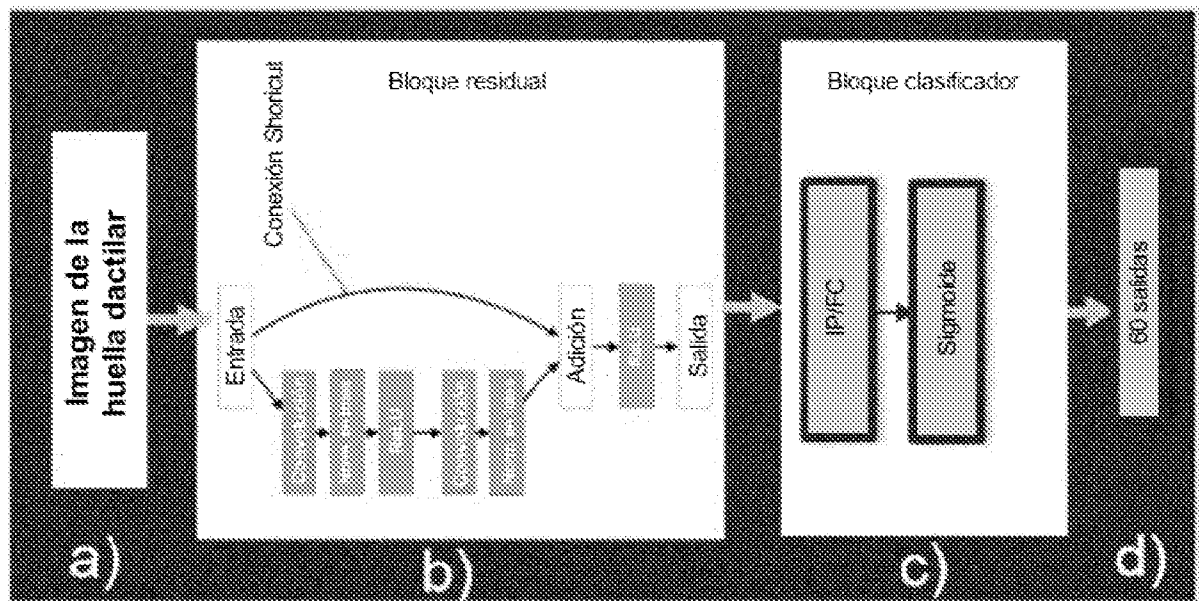


FIG. 2

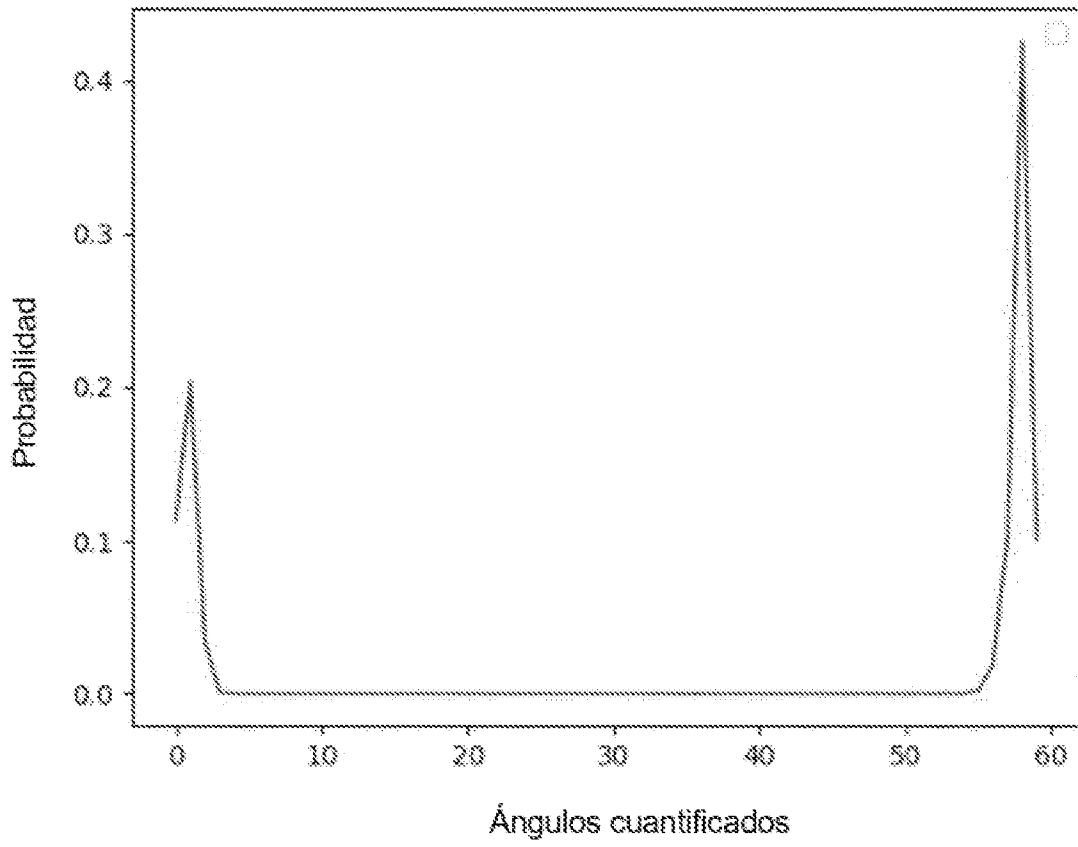


FIG. 3

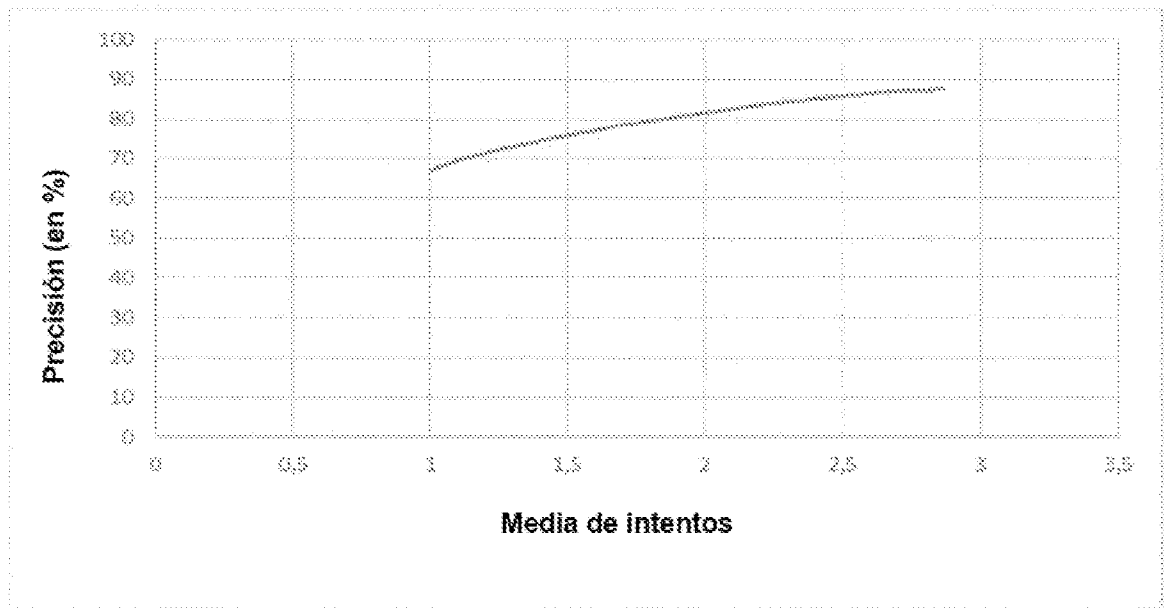


FIG. 4a

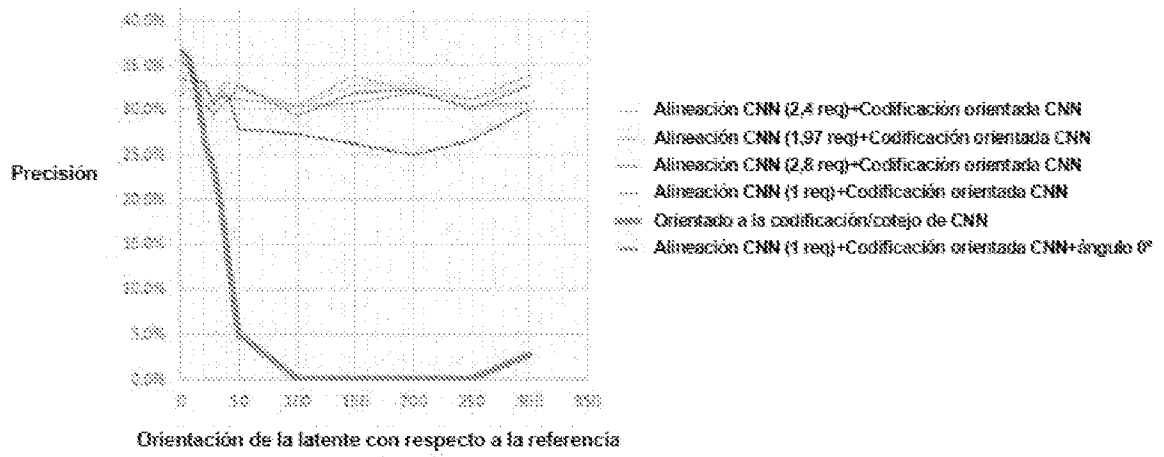


FIG. 4b