

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 879 937**

51 Int. Cl.:

**G06K 9/62**

(2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.12.2012** **PCT/EP2012/076398**

87 Fecha y número de publicación internacional: **11.07.2013** **WO13102574**

96 Fecha de presentación y número de la solicitud europea: **20.12.2012** **E 12816273 (2)**

97 Fecha y número de publicación de la concesión europea: **14.04.2021** **EP 2801056**

54 Título: **Análisis de imágenes**

30 Prioridad:

**02.01.2012 IT MI20120003**

**16.02.2012 US 201261599537 P**

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

**23.11.2021**

73 Titular/es:

**TELECOM ITALIA S.P.A. (100.0%)**

**Via Gaetano Negri, 1**

**20123 Milano, IT**

72 Inventor/es:

**BALESTRI, MASSIMO;**

**FRANCINI, GIANLUCA y**

**LEPSOY, SKJALG**

74 Agente/Representante:

**PONTI & PARTNERS, S.L.P.**

ES 2 879 937 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

Análisis de imágenes

5 Antecedentes de la invención

Campo de la invención

10 **[0001]** La presente invención se refiere al campo del análisis de imágenes.

Descripción de la técnica relacionada

**[0002]** En el campo del análisis de imágenes, una operación común proporciona comparar dos imágenes para encontrar la relación que ocurre entre ellas en caso de que ambas imágenes incluyan al menos una porción de una misma escena o de un mismo objeto.

**[0003]** Entre un gran número de aplicaciones, la comparación de imágenes es de suma importancia para calibrar las cámaras de vídeo pertenecientes a un sistema multicámara, para evaluar el movimiento que ocurre entre dos fotogramas de una toma de vídeo y para el reconocimiento de un objeto dentro de una imagen (por ejemplo, una imagen). Esta última aplicación está adquiriendo cada vez más importancia debido al reciente desarrollo de algoritmos de reconocimiento de objetos diseñados específicamente para ser empleados en los llamados motores de búsqueda visual, es decir, servicios automatizados que, a partir de una imagen, son capaces de identificar el objeto u objetos representados en ellos y ofrecer información relacionada con el objeto u objetos identificados. Ejemplos de servicios conocidos de este tipo incluyen Google Goggles, Nokia Point&Find y Kooaba Smart Visuals. Una aplicación de reconocimiento de objetos típicamente proporciona comparar una primera imagen, en la jerga, denominada «imagen de consulta», que representa un objeto a reconocer con una pluralidad de imágenes modelo, cada una de las cuales representa un objeto conocido respectivo; esto permite realizar una comparación entre el objeto representado en la imagen de consulta y los objetos representados en las imágenes modelo.

**[0004]** Las imágenes modelo se disponen típicamente en una base de datos modelo adecuada. Por ejemplo, en caso de que el reconocimiento de objetos se explote en un escenario de compra en línea, cada imagen modelo corresponde a un elemento ofrecido por una tienda en línea (por ejemplo, la imagen de una cubierta de libro, una cubierta de DVD y/o una cubierta de CD). El número de imágenes modelo incluidas en una base de datos de este tipo es bastante alto; por ejemplo, una base de datos modelo de un servicio de compras en línea puede incluir varios millones de imágenes modelo diferentes.

**[0005]** Una forma muy eficiente de realizar operaciones de comparación entre dos imágenes proporciona seleccionar un conjunto de puntos, en la jerga, denominados puntos clave, en la primera imagen y a continuación hacer coincidir cada punto clave del conjunto con un punto clave correspondiente en la segunda imagen. La selección de qué punto de la primera imagen tiene que convertirse en un punto clave se lleva a cabo ventajosamente mediante la extracción de características locales del área de la imagen que rodea el punto en sí, tal como, por ejemplo, la escala de extracción de puntos, la orientación privilegiada del área y el denominado «descriptor». En el campo del análisis de imágenes, un descriptor de un punto clave es un operador matemático que describe el gradiente de luminancia de un área de la imagen (llamada parche) centrada en el punto clave, con dicho parche que está orientado de acuerdo con el gradiente de luminancia principal del parche en sí.

**[0006]** En «Distinctive image features from scale-invariant keypoints» de David G. Lowe, International Journal of computer vision, 2004, se ha propuesto un descriptor de Transformación de Características Invariantes de Escala (SIFT); brevemente, para permitir un reconocimiento de imagen confiable, los descriptores de SIFT se generan teniendo en cuenta que las características locales extraídas de la imagen correspondiente a cada punto clave deben ser detectables incluso bajo cambios en la escala de la imagen, el ruido y la iluminación. Por lo tanto, los descriptores de SIFT son invariantes para la escala uniforme, la orientación y parcialmente invariantes para afectar la distorsión y los cambios de iluminación.

**[0007]** El descriptor de SIFT es una herramienta bastante poderosa, que permite seleccionar puntos clave para realizar comparaciones de imágenes precisas. Sin embargo, esta precisión solo puede lograrse con el uso de una cantidad bastante grande de datos; por ejemplo, un descriptor de SIFT típico es una matriz de 128 bytes de datos. Dado que el número de puntos clave en cada imagen es relativamente alto (por ejemplo, 1000-1500 puntos clave para una imagen VGA estándar), y dado que cada punto clave está asociado con un descriptor de SIFT correspondiente, la cantidad total de datos a procesar puede volverse excesiva para gestionarse de manera eficiente.

**[0008]** Este inconveniente se exagera en caso de que el escenario implique el uso de terminales móviles (por ejemplo, identificación de objetos extraídos de imágenes tomadas por la cámara de un teléfono inteligente). De hecho, dado que las operaciones a realizar para llevar a cabo el análisis de imágenes son bastante complejas y exigentes en términos de carga computacional, en este caso, la mayoría de las operaciones se realizan generalmente en el lado

del servidor; con el fin de tener toda la información requerida para realizar el análisis, el servidor necesita recibir del terminal móvil todos los datos requeridos, incluidos los descriptores de SIFT para todos los puntos clave. Por lo tanto, la cantidad de datos que se transmitirán desde el terminal al servidor puede volverse excesiva para garantizar una buena eficiencia del servicio.

**[0009]** De acuerdo con una solución conocida en la técnica, como por ejemplo la empleada por Google Goggles, este inconveniente se resuelve en la raíz transmitiendo directamente la imagen, y no los descriptores, desde el terminal móvil al servidor. De hecho, debido al número bastante alto de puntos clave, la cantidad de datos de los descriptores de SIFT correspondientes puede exceder el tamaño (en términos de bytes) de una imagen VGA estándar en sí.

**[0010]** Hongping Cai et al «Learning weights for codebook in image classification and retrieval», the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 de junio de 2010, IEEE Piscataway, NJ, USA, 13 de junio de 2010, páginas 2320-2327, ISBN 978-1-4244-6984-0, presenta una estrategia de aprendizaje del libro de códigos para la clasificación y recuperación de imágenes. Corresponde a aprender una métrica de similitud ponderada para satisfacer que la similitud ponderada entre las mismas imágenes etiquetadas es mayor que la que existe entre las imágenes etiquetadas de manera diferente con mayor margen. Los autores formulan el problema de aprendizaje como una programación cuadrática convexa y adoptan la optimización alterna para resolverlo de manera eficiente. Los experimentos en conjuntos de datos sintéticos y reales validan la estrategia. El aprendizaje del libro de códigos mejora el rendimiento, en particular en el caso de que el número de ejemplos de capacitación no sea suficiente para un libro de códigos de gran tamaño.

#### Resumen de la invención

**[0011]** El Solicitante ha encontrado que las estrategias conocidas en la técnica no son eficientes, lo que aún requiere la gestión de una gran cantidad de datos y/o la concentración de una gran porción de las operaciones en el lado del servidor, lo que limita la escalabilidad del sistema y el rendimiento general.

**[0012]** Por ejemplo, la solución empleada por Google Goggles, que permite transmitir directamente la imagen, y no los descriptores, desde el terminal móvil al servidor requiere que toda la carga computacional se mueva hacia el servidor, lo que puede convertirse en una sobrecarga. Además, la transmisión de la imagen comprimida todavía requiere una cantidad considerable de datos (por ejemplo, decenas de Kbytes para una imagen VGA).

**[0013]** El Solicitante ha abordado el problema de cómo mejorar estas estrategias en términos de cantidad de datos a procesar.

**[0014]** En particular, el Solicitante ha abordado el problema para proporcionar un procedimiento para el procesamiento de una imagen que requiere una cantidad reducida de datos para ser gestionada.

**[0015]** El Solicitante ha descubierto que la cantidad de datos que deben procesarse para realizar procedimientos de análisis de imágenes puede reducirse ventajosamente seleccionando un subconjunto óptimo de puntos clave, entre los puntos clave identificados en la imagen, de acuerdo con las probabilidades de relevancia de características locales calculadas sobre la base de distribuciones estadísticas de referencia.

**[0016]** De acuerdo con un aspecto de la presente invención, se propone un procedimiento para procesar una imagen. El procedimiento comprende identificar un primer grupo de puntos clave en la imagen. Para cada punto clave del primer grupo, el procedimiento proporciona identificar al menos una característica local de punto clave correspondiente relacionada con dicho punto clave; para dicha al menos una característica local de punto clave, calcular una probabilidad de relevancia de característica local correspondiente; calcular una probabilidad de relevancia de puntos clave en función de las probabilidades de relevancia de características locales de dicha al menos una característica local. El procedimiento comprende además seleccionar puntos clave, entre los puntos clave del primer grupo, que tienen las mayores probabilidades de relevancia de puntos clave para formar un segundo grupo de puntos clave, y explotar los puntos clave del segundo grupo para analizar la imagen. La probabilidad de relevancia de característica local calculada para una característica local de un punto clave se obtiene comparando el valor asumido por dicha característica local con una distribución estadística de referencia correspondiente de los valores de dicha característica local.

**[0017]** De acuerdo con la presente invención, cada una de dicha distribución estadística de referencia correspondiente es estadísticamente equivalente a una distribución estadística correspondiente generada al recopilar, entre una pluralidad de puntos clave de referencia identificados en una pluralidad de pares de imágenes de referencia, los valores de características locales correspondientes a esos puntos clave de referencia de cada par de imágenes de referencia que se ha determinado que implican una coincidencia de características correcta entre las imágenes de referencia de dicho par.

**[0018]** Preferentemente, dicha al menos una característica local de punto clave relacionada con dicho cada

punto clave comprende al menos un elemento entre las coordenadas del punto clave, la escala en la que se ha identificado el punto clave, la orientación dominante del punto clave, el pico del punto clave y un descriptor del punto clave.

- 5 **[0019]** De acuerdo con la presente invención, cada distribución estadística de referencia está dispuesta en forma de un histograma correspondiente que tiene una pluralidad de cajas. Cada caja corresponde a un intervalo predefinido de valores de la característica local correspondiente. La frecuencia de cada caja corresponde a una relación entre:
- 10 a) el número de puntos clave de referencia que se ha determinado que implican una coincidencia de características correcta y que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y  
b) el número total de puntos clave de referencia que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja.
- 15 **[0020]** De acuerdo con la presente invención, dicho cálculo de la probabilidad de relevancia de característica local para una característica local de un punto clave comprende inspeccionar el histograma correspondiente a dicha característica local para identificar la caja del mismo que se ajusta al valor asumido por la característica local del punto clave, y establecer la probabilidad de relevancia de característica local a la frecuencia de la caja identificada.
- 20 **[0021]** Ventajosamente, dicho cálculo de una probabilidad de relevancia de puntos clave de un punto clave del primer grupo incluye combinar las probabilidades de relevancia de características locales de cada una de dicha al menos una característica local del punto clave correspondiente.
- [0022]** Preferentemente, dicho cálculo de una probabilidad de relevancia de puntos clave de un punto clave del  
25 primer grupo incluye multiplicar entre sí las probabilidades de relevancia de características locales de cada una de dicha al menos una característica local del punto clave correspondiente.
- [0023]** Otro aspecto de la presente invención proporciona un sistema para procesar una imagen, con el fin de  
30 compararla dicha imagen con al menos otra imagen. Dicho sistema comprende una primera unidad de procesamiento configurada para identificar un primer grupo de puntos clave en la imagen, y una segunda unidad de procesamiento configurada para realizar las siguientes operaciones para cada punto clave del primer grupo:
- a) identificar al menos una característica local de punto clave correspondiente relacionada con dicho punto clave;  
35 b) para dicha al menos una característica local de punto clave, calcular una probabilidad de relevancia de característica local correspondiente mediante la comparación del valor asumido por dicha característica local con una distribución estadística de referencia correspondiente de los valores de dicha característica local, dicha distribución estadística de referencia es estadísticamente equivalente a una distribución estadística correspondiente generada al recopilar, entre una pluralidad de puntos clave de referencia identificados en una pluralidad de pares de imágenes de referencia, los valores de características locales correspondientes a esos  
40 puntos clave de referencia de cada par de imágenes de referencia que se ha determinado que implican una coincidencia de características correcta entre las imágenes de referencia de dicho par;  
c) calcular una probabilidad de relevancia de puntos clave en función de las probabilidades de relevancia de características locales de dicha al menos una característica local.
- 45 **[0024]** El sistema comprende además una tercera unidad de procesamiento configurada para seleccionar puntos clave, entre los puntos clave del primer grupo, que tiene las mayores probabilidades de relevancia de puntos clave para formar un segundo grupo de puntos clave, y una cuarta unidad de procesamiento configurada para explotar los puntos clave del segundo grupo para comparar dicha imagen con dicha al menos otra imagen.
- 50 **[0025]** Cada distribución estadística de referencia está dispuesta en forma de un histograma correspondiente que tiene una pluralidad de cajas, cada caja corresponde a un intervalo predefinido de valores de la característica local correspondiente, y la frecuencia de cada caja corresponde a una relación entre:
- a) el número de puntos clave de referencia que se ha determinado que implican una coincidencia de características  
55 correcta y que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y  
b) el número total de puntos clave de referencia que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja.
- [0026]** La segunda unidad de procesamiento está configurada para calcular la probabilidad de relevancia de  
60 características locales para una característica local de un punto clave al:
- c) inspeccionar el histograma correspondiente a dicha característica local para identificar la caja del mismo que se ajusta al valor asumido por la característica local del punto clave, y  
d) establecer la probabilidad de relevancia de característica local a la frecuencia de la caja identificada.
- 65

**[0027]** Otro aspecto adicional de la presente invención se refiere a un procedimiento de generación de una distribución estadística de referencia de valores de una característica local de punto clave.

#### Breve descripción de los dibujos

5

**[0028]** Estas y otras características y ventajas de la presente invención se harán evidentes mediante la siguiente descripción de algunas realizaciones ejemplares y no taxativas de la misma, que se leerán junto con los dibujos adjuntos, en los que:

- 10 La **Figura 1** ilustra en términos de bloques funcionales un procedimiento de extracción dirigido a extraer de una imagen de consulta un conjunto óptimo de puntos clave y generar un conjunto comprimido de descriptores de acuerdo con una realización de la presente invención;
- Las **Figuras 2A-2F** son distribuciones estadísticas de las características locales seleccionadas correspondientes de puntos clave de acuerdo con algunas realizaciones ejemplares de la presente invención;
- 15 La **Figura 2G** es una imagen ejemplar procesada de acuerdo con el procedimiento de extracción de la **Figura 1**;
- La **Figura 3A** ilustra un descriptor ejemplar del tipo SIFT;
- La **Figura 3B** ilustra una matriz de descriptores ejemplar del descriptor de la **Figura 3A**;
- La **Figura 4A** ilustra una compresión de matriz de descriptores ejemplar de acuerdo con una solución conocida en la técnica; La **Figura 4B** ilustra una compresión de matriz de descriptores ejemplar de acuerdo con otra solución conocida en la técnica;
- 20 La **Figura 5** ilustra una disposición de subhistogramas de un descriptor en familias de correlación de acuerdo con una realización de la presente invención; Las **Figuras 6A-6D** muestran cómo se comprime la matriz de descriptores de acuerdo con realizaciones ejemplares de la presente invención;
- La **Figura 7A** ilustra una distribución ejemplar de puntos clave **KP**;
- 25 La **Figura 7B** ilustra cómo se puede superponer una cuadrícula sobre la imagen de consulta para cuantificar las coordenadas de los puntos clave de la **Figura 7A**;
- La **Figura 7C** es una representación gráfica ejemplar de un histograma obtenido al superponer la cuadrícula de la **Figura 7B** sobre el conjunto de puntos clave **KP** de la **Figura 7A**; La **Figura 7D** identifica las columnas y filas de la cuadrícula de la **Figura 7B** que están completamente formadas por celdas que no incluyen ningún punto clave;
- 30 La **Figura 7E** ilustra un histograma ejemplar sobre un *soporte de rango 1*;
- La **Figura 7F** ilustra un mapa de histogramas correspondiente al histograma sobre el *soporte de rango 1* de la **Figura 7E**;
- La **Figura 8A** ilustra un ejemplo de un histograma de palabras;
- La **Figura 8B** ilustra un ejemplo de un mapa de histogramas;
- 35 La **Figura 9** ilustra en términos de bloques funcionales un procedimiento de coincidencia dirigido a realizar la comparación entre dos imágenes de acuerdo con una realización de la presente invención, y
- La **Figura 10** ilustra en términos de bloques funcionales un procedimiento de recuperación dirigido a recuperar de una base de datos modelo una imagen modelo que representa el mismo objeto/escena representado en la imagen de consulta de acuerdo con una realización de la presente invención.

40

#### Descripción detallada de realizaciones ejemplares de la invención

#### **PROCEDIMIENTO DE EXTRACCIÓN (FIGURA 1)**

- 45 **[0029]** La **Figura 1** ilustra en términos de bloques funcionales un procedimiento, en lo sucesivo denominado «procedimiento de extracción» e identificado con la referencia **100**, dirigido a procesar una imagen de entrada con el fin de obtener un conjunto óptimo de puntos clave y generar un conjunto correspondiente de descriptores de acuerdo con una realización de la presente invención. Los puntos clave y los descriptores se explotarán a continuación con fines de análisis de imágenes. En lo que sigue de la presente descripción, las expresiones genéricas «análisis de imágenes» y «análisis de una imagen» tienen por objeto comprender todas aquellas operaciones que proporcionan comparar una imagen con al menos otra imagen. Estas operaciones pueden llevarse a cabo en una amplia variedad de aplicaciones, tal como, por ejemplo, en una aplicación de reconocimiento de objetos, así como en una aplicación que proporciona la creación de una única imagen panorámica a partir de una pluralidad de imágenes diferentes.
- 50
- 55 **[0030]** Como se describirá más adelante, los procedimientos de extracción de acuerdo con una realización de la presente invención proporcionan además seleccionar un subconjunto óptimo de puntos clave y comprimir los descriptores de dichos puntos clave en una medida tal que mejore en gran medida la eficiencia de los procedimientos posteriores.
- 60 **[0031]** Las etapas del procedimiento de extracción **100** descritas en esta sección pueden llevarse a cabo por unidades de procesamiento adecuadas, cuya estructura y función depende del campo de aplicación específico al que están destinadas. Por ejemplo, cada unidad de procesamiento puede ser una unidad de hardware diseñada específicamente para realizar una o más etapas del procedimiento. Además, las etapas del procedimiento pueden llevarse a cabo mediante una máquina programable (por ejemplo, una computadora) bajo el control de un conjunto
- 65 correspondiente de instrucciones.

**Extracción de puntos clave (fase 110)**

5 **[0032]** La primera fase **110** del procedimiento de extracción **100** proporciona recibir una imagen de consulta **115** y extraer de esta un primer conjunto de puntos clave **KP**, cada uno asociado con un par correspondiente de coordenadas espaciales **C** que identifican la ubicación de dicho punto clave **KP** dentro de la imagen de consulta **115**.

10 **[0033]** Esta operación puede llevarse a cabo explotando el algoritmo de extracción de puntos clave de Diferencia de Gaussianos (DoG) conocido; sin embargo, se aplican consideraciones similares en caso de que se empleen diferentes algoritmos de extracción de puntos clave, tales como, por ejemplo, el algoritmo de extracción de puntos clave de Determinante de Hessianos (DoH). Haciendo referencia al algoritmo de extracción de puntos clave de DoG, la imagen de consulta **115** se convoluciona con filtros gaussianos en una secuencia a diferentes escalas. A continuación, se lleva a cabo una operación de diferencia entre pares de imágenes adyacentes por desenfoque gaussiano en la secuencia. Los puntos clave **KP** se eligen entonces como los puntos que tienen valores  
15 máximos/mínimos de Diferencia de Gaussianos (DoG) en múltiples escalas. Particularmente, cada píxel en una imagen de DoG se compara con sus ocho vecinos en la misma escala y con nueve píxeles vecinos en cada una de las escalas vecinas (es decir, las escalas posteriores y anteriores en la secuencia). Si el valor de píxel es el máximo o mínimo entre todos los píxeles comparados, ese punto se considera un punto clave **KP** candidato.

20 **[0034]** La fase **110** también proporciona que cada punto clave **KP** se asigne a una o más orientaciones en función de las direcciones de gradiente de luminancia de imagen local. Por ejemplo, se forma un histograma de orientación con una pluralidad de cajas, donde cada caja cubre un intervalo de grado correspondiente. Cada muestra en la ventana vecina añadida a una caja de histograma se pondera por su magnitud de gradiente y por una ventana circular ponderada por gaussianos. Los picos en el histograma resultante corresponden a orientaciones dominantes.  
25 Una vez rellenado el histograma, las orientaciones correspondientes al pico más alto y los picos locales que se encuentran dentro del 80 % de los picos más altos se asignan al punto clave **KP**. En caso de que se hayan asignado múltiples orientaciones, se crea un punto clave **KP** adicional que tiene la misma ubicación y escala que el punto clave original para cada orientación adicional.

30 **[0035]** Al final de la fase **110** se genera así un conjunto de puntos clave **KP**, junto con las coordenadas **C** correspondientes, la escala **S** en la que se extrae el punto clave, su orientación dominante **O** y el pico **P**, es decir, el valor absoluto del DoG correspondiente a dicho punto clave (que es indicativo del contraste del mismo).

**Generación de descriptores (fase 120)**

35 **[0036]** La siguiente fase **120** proporciona procesar la imagen de consulta **115** para calcular para cada punto clave **KP** un descriptor correspondiente **D**. En el ejemplo en cuestión, los descriptores **D** calculados en la fase **120** son un descriptor del tipo SIFT. Si bien los puntos clave **KP** se han extraído de tal manera que garantizan la invariancia a la ubicación, escala y rotación de la imagen, los descriptores de SIFT **D** se calculan de tal manera que son altamente  
40 distintivos y parcialmente invariantes para la iluminación y el punto de vista. Específicamente, para cada punto clave **KP** se calcula un conjunto de 16 subhistogramas en una cuadrícula 4x4 que está centrada en la ubicación de punto clave **KP** y orientada de acuerdo con la orientación dominante del punto clave **KP**. Cada subhistograma incluye 8 cajas, cada una de las cuales corresponde a una orientación que tiene un ángulo  $n \cdot \pi/4$  ( $n = 0, 1, \dots, 7$ ) con respecto a la orientación dominante; la frecuencia de cada caja de un subhistograma es proporcional al gradiente de luminancia  
45 de la celda de la cuadrícula (en lo sucesivo denominada subregión) correspondiente a dicho subhistograma, considerado a lo largo de la dirección identificada por dicha caja. Los valores de dichos histogramas de orientación se disponen en una matriz, formando el descriptor **D** del punto clave **KP**. Dado que hay  $4 \times 4 = 16$  subhistogramas cada uno con 8 cajas, el descriptor **D** es una matriz que tiene 128 elementos.

50 **[0037]** Los conceptos de la presente invención también son aplicables si el descriptor de SIFT se calcula en una cuadrícula que incluye un número diferente de celdas y/o con un número diferente de cajas por histograma.

**[0038]** Además, incluso si en el ejemplo en cuestión se ha hecho referencia a descriptores del tipo SIFT, se aplican consideraciones similares en caso de que se empleen diferentes tipos de descriptores, como por ejemplo la  
55 Característica Robusta Acelerada (SURF) y el Histograma de Gradientes Orientados (HOG), o posiblemente otros. Además, incluso si se ha hecho referencia y se hará a continuación a descriptores que comprenden datos relacionados con gradientes de luminancia, se aplican consideraciones similares si se consideran gradientes de diferentes parámetros. De hecho, como es bien conocido por los expertos en la materia, la luminancia es solo una de las propiedades físicas del color. Por lo tanto, incluso si se ha determinado que la luminancia es la mejor propiedad física  
60 (es decir, la más robusta) que se debe considerar para fines de análisis de imágenes, también se pueden considerar diferentes tipos de descriptores, por ejemplo, que comprenden datos relacionados con gradientes de crominancia, gradientes de saturación o incluso gradientes de color (que incluyen tanto luminancia, saturación y crominancia).

**[0039]** Como ya se mencionó anteriormente, la realización de operaciones de análisis de imágenes implica la  
65 gestión de una cantidad bastante grande de datos: de hecho, cada punto clave **KP** está asociado con una pluralidad

de características locales (en lo sucesivo identificadas globalmente con referencia **LFkp**), que incluyen las coordenadas **C**, la escala **S**, la orientación dominante **O** y el pico **P**, así como un descriptor correspondiente **D** formado por una matriz de 128 elementos. Para este propósito, con el fin de reducir la cantidad total de datos que se gestionarán (por ejemplo, que se memorizarán y/o transmitirán), el procedimiento de extracción **100** de acuerdo con una realización de la presente invención proporciona dos procedimientos, es decir:

- 1) reducir el número de puntos clave **KP** generados previamente seleccionando los puntos clave **KP** más relevantes (desde el punto de vista de comparación de imágenes), con el fin de obtener un subconjunto óptimo **SUB** de puntos clave **KP**, y
- 2) comprimir adecuadamente tanto las coordenadas **C** como los descriptores **D**.

**[0040]** La fase **130** del procedimiento de extracción **100** se dedica a la selección del subconjunto óptimo **SUB**, la fase **140** se dedica a la compresión de los descriptores **D** y la fase **150** se dedica a la compresión de las coordenadas **C**.

#### **Selección del subconjunto óptimo de puntos clave (fase 130)**

**[0041]** De acuerdo con una realización de la presente invención, la selección del subconjunto óptimo **SUB** se lleva a cabo calculando para al menos una característica local **LFkp**, las coordenadas **C**, la escala **S**, la orientación dominante **O**, el pico **P** y el descriptor **D**, de cada punto clave **KP** de la imagen de consulta **115** al menos una probabilidad de relevancia de característica correspondiente **FRP**, ordenando los puntos clave **KP** de acuerdo con una probabilidad de relevancia de puntos clave **KRP** en función de las probabilidades de relevancia de característica **FRP** de sus características locales **LFkp**, y luego seleccionando los puntos clave **KP** que tienen las probabilidades de relevancia de puntos clave más altas **KRP**.

**[0042]** De acuerdo con una realización de la presente invención, la probabilidad de relevancia de la característica **FRP** de cada característica local **LFkp** del punto clave genérico **KP** se calcula explotando una distribución estadística de referencia correspondiente **Rsd**, que ya ha sido predeterminada de antemano después de haber realizado evaluaciones estadísticas en una base de datos de imágenes de referencia.

**[0043]** Las distribuciones estadísticas de referencia **Rsd** se realizan de tal manera que reflejen el comportamiento estadístico de las características locales **LFkp** de los puntos clave **KP** considerados útiles para los fines del análisis de imágenes.

**[0044]** Por ejemplo, en el caso de procedimientos de reconocimiento de objetos, la base de datos de imágenes de referencia es una base de datos que comprende una pluralidad de pares de imágenes, con cada par de imágenes que consiste en dos imágenes que representan un mismo objeto/escena. De acuerdo con una realización de la presente invención, las distribuciones estadísticas de referencia se generan de la siguiente manera.

**[0045]** Los puntos clave se extraen en primer lugar de todas las imágenes de la base de datos de referencia. A continuación, se lleva a cabo un primer análisis estadístico sobre una o más características locales seleccionadas de todos los puntos clave extraídos, para generar primeras distribuciones estadísticas de dichas características locales seleccionadas. Cada primera distribución estadística de una característica local está dispuesta en forma de un histograma, obtenido contando el número de puntos clave (frecuencia de puntos clave), entre la totalidad de puntos clave extraídos de las imágenes de la base de datos de referencia, que tiene un valor de dicha característica local que se encuentra dentro de cada uno de una pluralidad de intervalos (cajas) de valor de características locales predefinidos. A continuación, para cada par de imágenes, los puntos clave de una imagen se emparejan con los puntos clave de la otra imagen. Las coincidencias entre dichos puntos clave se procesan utilizando un procedimiento de comparación de imágenes (tal como uno cualquiera de los procedimientos de comparación de imágenes conocidos basados en la coincidencia de características de imagen) para identificar qué coincidencia es correcta (inlier) y cuál es incorrecta (outlier). A continuación, se lleva a cabo un segundo análisis estadístico sobre la misma característica o características consideradas previamente con el fin de generar las distribuciones estadísticas de referencia **Rsd** que se utilizarán para calcular las probabilidades de relevancia de característica **FRP**. Esta vez, la generación de las distribuciones estadísticas de referencia **Rsd** se lleva a cabo calculando para cada caja una relación entre el número de puntos clave pertenecientes a inliers y que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y el número total de puntos clave (tanto pertenecientes a inliers como outliers) que tienen un valor de la característica local correspondiente que se encuentra dentro de la misma caja. El Solicitante ha observado que las primeras distribuciones estadísticas y las distribuciones estadísticas de referencia **Rsd** son bastante diferentes entre sí. Dado que las distribuciones estadísticas de referencia **Rsd** se generan teniendo en cuenta los puntos clave que implican una coincidencia de características correcta (inlier), el Solicitante ha encontrado que dichas distribuciones estadísticas son buenos representantes del comportamiento estadístico de los puntos clave (en adelante, «puntos clave relevantes») que son relevantes para fines de análisis de imágenes, y particularmente adecuados para ser empleados eficientemente en un procedimiento de comparación de imágenes.

**[0046]** Las Figuras **2A-2F** ilustran algunas distribuciones estadísticas **Rsd** de características locales

seleccionadas correspondientes **LF<sub>kp</sub>** de puntos clave **KP** de acuerdo con algunas realizaciones ejemplares de la presente invención. En particular, las distribuciones estadísticas **R<sub>sd</sub>** de las **Figuras 2A-2F** se han generado a partir de imágenes de una base de datos de referencia dispuesta específicamente para aplicaciones de reconocimiento de objetos. Si se considerara una aplicación de análisis de imágenes diferente, como por ejemplo la creación de una

5 única imagen panorámica a partir de una pluralidad de imágenes diferentes, las imágenes del punto de referencia y, por lo tanto, las distribuciones estadísticas resultantes **R<sub>sd</sub>** serían diferentes.

**[0047]** La **Figura 2A** es una distribución estadística **R<sub>sd</sub>** relacionada con las coordenadas **C** de los puntos clave **KP**. Cada caja del histograma correspondiente representa la distancia (en píxeles) del punto clave **KP** genérico

10 desde el centro de la imagen. En el ejemplo en cuestión, la imagen considerada es del tipo VGA (es decir, tiene una resolución de 640 x 480), por lo tanto, el centro corresponde a la coordenada (320, 240). De acuerdo con el histograma ilustrado en la **Figura 2A**, la caja que tiene la frecuencia **KP** de puntos clave más alta es la que corresponde al centro de la imagen. Esto significa que cuanto más cerca esté un punto clave **KP** del centro, mayor será la probabilidad de que dicho punto clave **KP** sea un punto clave relevante; la tendencia de las frecuencias del histograma disminuye

15 monotónicamente a medida que aumenta la distancia desde el centro. Esto podría explicarse fácilmente por el hecho de que cuando se fotografía un objeto, es muy probable que dicho objeto esté enmarcado en el centro de la imagen. Se debe apreciar que en este caso las cajas del histograma no tienen todos los mismos anchos; esto se debe al hecho de que el ancho de cada caja ha sido determinado adecuadamente por un cuantificador (escalar y/o vectorial) de tal manera que se calculan pocas cajas, evitando así la aparición de ocurrencias de fenómenos de sobreajuste. Los

20 conceptos de la presente invención también se aplican en caso de que se emplee una cuantificación uniforme (escalar y/o vectorial), es decir, con todas las cajas del histograma que tienen un mismo ancho.

**[0048]** La **Figura 2B** es una distribución estadística **R<sub>sd</sub>** relacionada con la orientación dominante **O** de los puntos clave **KP**. Cada caja del histograma correspondiente representa el ángulo (en radianes) de la dirección

25 dominante del punto clave **KP** genérico con respecto al horizonte (correspondiente a 0 radianes). De acuerdo con el histograma ilustrado en la **Figura 2B**, las cajas que tienen las frecuencias de puntos clave **KP** más altas son las que corresponden a las orientaciones que son paralelas o perpendiculares a la orientación del horizonte (es decir, que corresponden a  $\pi/2$ , 0,  $-\pi/2$ ,  $-\pi$ ). Esto significa que cuanto más cercana sea la orientación de un punto clave **KP** a una de dichas orientaciones, mayor será la probabilidad de que dicho punto clave **KP** sea un punto clave relevante.

30 Esto podría explicarse por el hecho de que cuando se fotografía un objeto, es altamente probable que dicho objeto esté enmarcado de manera que se extienda principalmente paralelo y/o perpendicular a la línea del horizonte. En este caso también, la anchura de las cajas se determina por medio de un cuantificador.

**[0049]** La **Figura 2C** es una distribución estadística **R<sub>sd</sub>** relacionada con el pico **P** de los puntos clave **KP**.

35 Cada caja del histograma correspondiente representa el contraste entre el punto clave **KP** genérico y el punto más similar entre los vecinos. De acuerdo con el histograma ilustrado en la **Figura 2C**, la caja que tiene la frecuencia de puntos clave **KP** más alta es la que corresponde a los valores de pico más altos. Esto significa que cuanto mayor sea el contraste de un punto clave **KP**, mayor será la probabilidad de que dicho punto clave **KP** sea un punto clave relevante; la tendencia de las frecuencias del histograma aumenta monotónicamente a medida que aumenta el

40 contraste. Esto podría explicarse fácilmente por el hecho de que un punto de una imagen que tiene un alto contraste es fácilmente reconocible e identificable. En este caso también, la anchura de las cajas se determina por medio de un cuantificador.

**[0050]** La **Figura 2D** es una distribución estadística **R<sub>sd</sub>** relacionada con la escala **S** de los puntos clave **KP**.

45 Cada caja del histograma correspondiente representa una escala particular **S** en la que se puede extraer el punto clave **KP**. De acuerdo con el histograma ilustrado en la **Figura 2D**, la caja que tiene la frecuencia de puntos clave **KP** más alta corresponde a una escala media-baja. En este caso también, la anchura de las cajas se determina por medio de un cuantificador.

**[0051]** La **Figura 2E** es una primera distribución estadística **R<sub>sd</sub>** relacionada con los descriptores **D** de los puntos clave **KP**. En este caso, el histograma correspondiente es tridimensional, y cada caja de este corresponde a

50 valores de intervalo de dos parámetros del descriptor **D** del punto clave **KP** genérico, es decir, la media (eje x) y la varianza (eje y) del descriptor **D**. Los valores de mayor frecuencia se indican mediante círculos de mayor diámetro. La media y la varianza se han considerado en conjunto para formar un mismo histograma, ya que están vinculadas entre

55 sí. De acuerdo con dicho histograma, la caja que tiene la frecuencia de puntos clave **KP** más alta, representada por círculos más grandes, es la que corresponde a la media más alta y la varianza más baja. Esto puede explicarse por el hecho de que cuanto mayor es la media del descriptor **D** de un punto clave **KP**, mayor es el gradiente de luminancia correspondiente a dicho punto clave **KP**, y cuanto menor es la varianza del descriptor **D** de un punto clave **KP**, menor es el ruido no deseado que afecta a dicho punto clave **KP**.

60

**[0052]** La **Figura 2F** es una segunda distribución estadística **R<sub>sd</sub>** relacionada con los descriptores **D** de los puntos clave **KP**. En este caso, cada caja corresponde a una distancia máxima particular entre el descriptor **D** de un

65 punto clave **KP** y los descriptores **D** de los otros puntos clave **KP** de la misma imagen. Por ejemplo, dicha distancia máxima se puede calcular en función de la distancia euclidiana entre descriptores. También se puede contemplar otro procedimiento conocido, tal como, por ejemplo, explotar la divergencia simétrica de Kullback-Leibler.

**[0053]** Volviendo a la **Figura 1**, de acuerdo con una realización de la presente invención, la fase **130** del procedimiento de extracción **100** proporciona el cálculo, para cada punto clave **KP** extraído en la fase **110**:

- 5 - Una primera probabilidad de relevancia de la característica **FRP1**, obtenida a partir de la distribución estadística **Rsd** relacionada con las coordenadas **C** de dicho punto clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta a las coordenadas **C** de dicho punto clave **KP**; entonces, la probabilidad de relevancia de la característica **FRP1** se establece igual a la frecuencia de puntos clave de la caja identificada.
- 10 - Una segunda probabilidad de relevancia de la característica **FRP2**, obtenida a partir de la distribución estadística **Rsd** relacionada con la orientación dominante **O** de dicho punto clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta a la orientación dominante **O** de dicho punto clave **KP**; entonces, la probabilidad de relevancia de la característica **FRP2** se establece igual a la frecuencia de puntos clave de la caja identificada.
- 15 - Una tercera probabilidad de relevancia de la característica **FRP3**, obtenida a partir de la distribución estadística **Rsd** relacionada con el pico **P** de dicho punto clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta al pico **P** de dicho punto clave **KP**; entonces, la probabilidad de relevancia de la característica **FRP3** se establece igual a la frecuencia de puntos clave de la caja identificada.
- Una cuarta probabilidad de relevancia de la característica **FRP4**, obtenida a partir de la distribución estadística **Rsd** relacionada con la escala **S** de dicho punto clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta a la escala **S** de dicho punto clave **KP**; entonces, la probabilidad de relevancia de la característica **FRP4** se establece igual a la frecuencia de puntos clave de la caja identificada.
- 20 - Una quinta probabilidad de relevancia de la característica **FRP5**, obtenida a partir de la distribución estadística **Rsd** relacionada con la media y la varianza del descriptor **D** de dicho punto clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta a la media y la varianza de los elementos del descriptor **D** de dicho punto clave **KP**; entonces, la probabilidad de relevancia de la característica **FRP5** se establece igual a la frecuencia de puntos clave de la caja identificada.
- 25 - Una sexta probabilidad de relevancia de la característica **FRP6**, obtenida a partir de la distribución estadística **Rsd** relacionada con la distancia máxima (por ejemplo, la distancia euclidiana) entre el descriptor **D** de dicho punto clave **KP** y los descriptores **D** de los otros puntos clave **KP**. Se inspecciona el histograma correspondiente a dicha distribución con el fin de identificar la caja del mismo que se ajusta a dicha distancia; entonces, la probabilidad de relevancia de la característica **FRP6** se establece igual a la frecuencia de puntos clave de la caja identificada.

**[0054]** Por lo tanto, para cada punto clave **KP**, se obtiene una probabilidad de relevancia de punto clave **KRP** mediante al menos uno de, o combinando entre ellos, las probabilidades de relevancia de la característica **FRP** de las características locales de estos. Por ejemplo, comenzando con la suposición de que las probabilidades de relevancia de la característica **FRP** son independientes entre sí, la probabilidad de relevancia de punto clave **KRP** del punto clave genérico **KP** se calcula multiplicando entre sí sus probabilidades de relevancia de la característica **FRP** correspondientes. Generalmente, cuanto mayor sea el número de diferentes probabilidades de relevancia de la característica **FRP** utilizadas para calcular la probabilidad de relevancia de punto clave **KRP**, mejores serán los resultados que se pueden obtener mediante el empleo de dicho procedimiento. Al considerar el ejemplo de descriptores de SIFT para aplicaciones de búsqueda visual, es preferible que las probabilidades de relevancia de características consideradas para calcular la probabilidad de relevancia de puntos clave incluyan al menos aquellas correspondientes a la escala, el pico y la distancia desde el centro.

**[0055]** La **Figura 2G** es una imagen ejemplar en la que se identifican una pluralidad de puntos clave mediante puntos circulares correspondientes, cada uno con un diámetro que es proporcional a la probabilidad de relevancia **KRP** del punto clave.

**[0056]** Una vez calculadas las probabilidades de relevancia de punto clave **KRP** de todos los puntos clave **KP** extraídos en la fase **110**, dichos puntos clave **KP** se clasifican en una secuencia de acuerdo con un orden de probabilidad de relevancia de punto clave **KRP** decreciente. A continuación, el subconjunto óptimo **SUB** se forma tomando un número (basado en la reducción deseada en la cantidad de datos a gestionar) de puntos clave **KP** de los primeros de la secuencia ordenada. Los puntos clave **KP** seleccionados que pertenecen al subconjunto óptimo **SUB** resultan ser los puntos clave **KP** más relevantes (desde el punto de vista de comparación de imágenes) entre la totalidad de puntos clave **KP** extraídos en la fase **110**. De esta manera, la reducción de la cantidad total de datos se lleva a cabo de manera inteligente y eficiente, teniendo en cuenta solo los puntos clave **KP** relevantes, y descartando aquellos que son menos útiles.

**[0057]** Se subraya que aunque la selección del subconjunto óptimo de puntos clave de acuerdo con la realización de la invención descrita anteriormente proporciona el cálculo de la probabilidad de relevancia de cada característica explotando una distribución estadística correspondiente **Rsd** obtenida mediante el cálculo para cada caja de la misma de una relación entre los inliers de puntos clave que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y el número total de puntos clave que tienen un valor de la característica local correspondiente que se encuentra dentro de la misma caja, los conceptos de la presente invención

también son aplicables en caso de que se empleen distribuciones estadísticas diferentes, estadísticamente equivalentes, obtenidas con procedimientos diferentes, incluso manuales. En la siguiente descripción, dos distribuciones estadísticas se consideran estadísticamente equivalentes entre sí si permiten obtener probabilidades de relevancia de características similares a partir de un mismo conjunto de puntos clave.

5

#### **Compresión de los descriptores (fase 140)**

[0058] De acuerdo con una realización de la presente invención, la compresión de los descriptores **D** se lleva a cabo mediante cuantificación vectorial, mediante la explotación de un número reducido de libros de códigos optimizados.

[0059] La **Figura 3A** ilustra un descriptor **D** ejemplar del tipo SIFT (uno de los descriptores **D** generados en la fase **120** del procedimiento de extracción **100** de la **Figura 1** que se ha seleccionado para formar parte del subconjunto óptimo **SUB**) correspondiente a un punto clave **KP** genérico. Como ya se mencionó anteriormente, el descriptor **D** comprende dieciséis subhistogramas **shi** ( $i = 1, 2, \dots, 16$ ), cada uno mostrando cómo el gradiente de luminancia de una subregión respectiva de la imagen cercana al punto clave **KP** se distribuye a lo largo de ocho direcciones. Específicamente, cada subhistograma **shi** está asociado con una subregión correspondiente a una de las 16 celdas de una cuadrícula 4x4 que está centrada en la ubicación del punto clave **KP** y orientada de acuerdo con la orientación dominante **O** del punto clave **KP**; cada subhistograma **shi** incluye ocho cajas, cada una correspondiente a una orientación que tiene un ángulo  $n \cdot \pi/4$  ( $n = 0, 1, \dots, 7$ ) con respecto a la orientación dominante **O**.

[0060] Como se ilustra en la **Figura 3B**, los valores de todos los histogramas de orientación **shi** de un descriptor **D** están dispuestos en una matriz de descriptores correspondiente, identificada en la figura con la referencia **DA**. La matriz de descriptores **DA** comprende dieciséis elementos **ai** ( $i = 1, 2, \dots, 16$ ), cada uno almacenando los valores tomados por un subhistograma **shi** correspondiente ( $i = 1, 2, \dots, 16$ ); cada elemento **ai** comprende a su vez ocho subelementos respectivos, cada uno de los cuales almacena un valor de frecuencia correspondiente a una de las ocho cajas respectivas del subhistograma **shi**. Por lo tanto, cada matriz de descriptores **DA** incluye  $16 \cdot 8 = 128$  subelementos. Al considerar que en un descriptor de **SIFT D** un valor de frecuencia típico puede variar de 0 a 255, cada subelemento de la matriz de descriptores **DA** puede representarse con un byte; por lo tanto, la ocupación de memoria de la matriz de descriptores **DA** es igual a 128 bytes. Por lo tanto, haciendo referencia nuevamente al procedimiento de extracción **100** de la **Figura 1**, la cantidad de datos (en bytes) correspondientes a todos los descriptores **D** de los puntos clave **KP** que pertenecen al subconjunto óptimo **SUB** seleccionado es igual a 128 multiplicado por el número de puntos clave **KP** del subconjunto óptimo **SUB**.

[0061] Para reducir esta cantidad de datos, las matrices de descriptores **DA** correspondientes a dichos descriptores **D** se comprimen a través de la cuantificación de vectores.

[0062] Como es bien sabido por los expertos en la materia, comprimir una matriz de datos formada por  $n$  elementos ( $n$ -tupla) mediante la explotación de la cuantificación de vectores proporciona cuantificar conjuntamente el conjunto de todos los posibles valores de  $n$ -tupla que la matriz de datos puede asumir en un conjunto reducido que comprende un menor número de  $n$ -tupla (cuyos valores pueden incluso diferir de los valores del conjunto a cuantificar). Dado que el conjunto reducido comprende un menor número de valores de  $n$ -tupla, requiere menos espacio de almacenamiento. Los valores de  $n$ -tupla que forman el conjunto reducido también se denominan «palabras de código». Cada palabra de código está asociada con un conjunto correspondiente de diferentes valores de  $n$ -tupla que la matriz puede asumir. Las relaciones de asociación entre los valores de  $n$ -tupla de la matriz de datos y las palabras de código se determinan por medio de un libro de códigos correspondiente.

[0063] Haciendo referencia en particular a la matriz de descriptores **DA**, que incluye 16 elementos **ai** formados a su vez por ocho subelementos, cada uno con valores que varían de 0 a 255, la matriz de descriptores **DA** puede tomar un número  $N = 256^{128}$  de diferentes valores de 16 tuplas. Mediante la aplicación de compresión a través de cuantificación vectorial, dichos  $N$  valores diferentes de 16 tuplas se aproximan con un número  $N1 < N$  de palabras de código de un libro de códigos. El libro de códigos determina las relaciones de asociación entre cada palabra de código y un conjunto correspondiente de valores de 16 tuplas de la matriz de descriptores **DA**. Por lo tanto, cada palabra de código del libro de códigos es un valor de 16 tuplas que se usa para «aproximar» un conjunto correspondiente de valores de 16 tuplas de la matriz de descriptores **DA**. La cuantificación del vector es una compresión de datos con pérdida, cuya precisión se puede medir a través de un parámetro llamado distorsión. La distorsión se puede calcular, por ejemplo, como la distancia euclidiana entre una palabra de código genérica del libro de códigos y el conjunto de valores de  $n$ -tupla de la matriz que se aproximan por dicha palabra de código. Consideraciones similares se aplican incluso si la distorsión se calcula con un procedimiento diferente. En cualquier caso, en términos generales, cuanto mayor sea el número  $N1$  de palabras de código de un libro de códigos, menor será la distorsión de la compresión.

[0064] Como es bien conocido por los expertos en la materia, la generación de las palabras de código de un libro de códigos se lleva a cabo típicamente mediante la realización de operaciones estadísticas (denominadas operaciones de capacitación) en una base de datos de capacitación que incluye una colección de un número muy alto de matrices de capacitación. Haciendo referencia en particular a la matriz de descriptores **DA**, la base de datos de

capacitación puede incluir varios millones de matrices de descriptores de capacitación, donde cada matriz de descriptores de capacitación es una de los  $N = 256^{128}$  posibles valores de 16 tuplas que la matriz de descriptores **DA** puede asumir.

- 5 **[0065]** De acuerdo con una solución ilustrada en la **Figura 4A**, toda la matriz de descriptores **DA** se comprime usando un único libro de códigos **CBK** que comprende  $N1$  palabras de código de valor **CWj** de 16 tuplas ( $j = 1, 2, \dots, N1$ ). Por lo tanto, con  $N1$  diferentes palabras de código **CWj**, el número mínimo de bits necesarios para identificar las palabras de código es igual a  $\log_2 N1$ . Como ya se mencionó anteriormente, la generación de las  $N1$  diferentes palabras de código **CWj** de dicho libro de códigos **CBK** único se lleva a cabo mediante la realización de operaciones de capacitación en una pluralidad de matrices de descriptores de capacitación, donde cada matriz de descriptores de capacitación es uno de los  $N = 256^{128}$  posibles valores de 16 tuplas que puede asumir la matriz de descriptores **DA**.

- [0066]** Con el fin de mantener la distorsión de compresión por debajo de un umbral lo suficientemente reducido como para no perjudicar el resultado de las operaciones posteriores de análisis de imágenes, el número  $N1$  de palabras de código requeridas puede llegar a ser muy alto. Tener un libro de códigos formado por un número demasiado alto  $N1$  de palabras de código es desventajoso bajo diferentes puntos de vista. De hecho, el número de matrices de capacitación que se utilizarían para generar las palabras de código sería excesivo y los tiempos de procesamiento serían demasiado largos. Además, para llevar a cabo operaciones de compresión mediante la explotación de un libro de códigos, todas las  $N1$  palabras de código que forman este último tienen que memorizarse en algún lugar, ocupando una cantidad no despreciable de espacio de memoria. Este último inconveniente es bastante crítico, ya que el hardware empleado para las aplicaciones de análisis de imágenes (por ejemplo, Unidades de Procesamiento Gráfico, GPU) pueden estar equipadas con memorias no tan potentes.

- [0067]** Haciendo referencia a la **Figura 4B**, con el fin de reducir la cantidad total de palabras de código **CWj** que se gestionarán sin aumentar la distorsión, la matriz de descriptores **DA** se puede subdividir en una pluralidad de submatrices **SDAk** ( $k = 1, 2, \dots$ ), donde cada una comprende un número respectivo  $mk$  de elementos **ai** de la matriz de descriptores **DA**, y a continuación cada submatriz **SDAk** se comprime individualmente usando un libro de códigos **CBKk** respectivo que comprende palabras de código **CWj** de valor  $N2$   $mk$  tupla ( $j = 1, 2, \dots, N2$ ).

- 30 **[0068]** En el ejemplo ilustrado en la **Figura 4B**, la matriz de descriptores **DA** se subdivide en cuatro submatrices **SDAk** ( $k = 1, 2, 3, 4$ ), cada una comprende  $mk = 4$  elementos **ai** de la matriz de descriptores **DA**:

- la primera submatriz **SDA1** está formada por la secuencia de elementos **a1, a2, a3, a4**;
- la segunda submatriz **SDA2** está formada por la secuencia de elementos **a5, a6, a7, a8**;
- 35 - la tercera submatriz **SDA3** está formada por la secuencia de elementos **a9, a10, a11, a12**, y
- la cuarta submatriz **SDA4** está formada por la secuencia de elementos **a13, a14, a15, a16**.

- [0069]** La compresión de cada submatriz **SDAk** se lleva a cabo utilizando un libro de códigos **CBKy** respectivo ( $y = k$ ) que comprende  $N2$  palabras de código **CWj** de valor de 4 tuplas ( $j = 1, 2, \dots, N2$ ). Por lo tanto, con  $4 \cdot N2$  palabras de código diferentes **CWj**, el número mínimo de bits requeridos para identificar todas las palabras de código es igual a  $4 \cdot \log_2 N2$ . Incluso si en el caso considerado cada submatriz **SDAk** se ha comprimido usando un libro de códigos **CBKy** que comprende un mismo número  $N2$  de palabras de código **CWj**, se aplican consideraciones similares en caso de que cada submatriz **SDAk** se comprima usando un número respectivo, diferente, de palabras de código **CWj**.

- 45 **[0070]** En el caso ilustrado en la **Figura 4B**, la generación de las  $N2$  diferentes palabras de código **CWj** de cada libro de códigos **CBKy** se lleva a cabo realizando operaciones de capacitación en un subconjunto respectivo de matrices de descriptores de capacitación. Cada subconjunto de matrices de descriptores de capacitación de un libro de códigos **CBKk** corresponde a una de las cuatro submatrices **SDAk**, y puede obtenerse considerando de cada matriz de descriptores de capacitación utilizada para generar el libro de códigos **CBK** único de la **Figura 4A** solo la porción de la misma correspondiente a la submatriz **SDAk**. Por ejemplo, con el fin de generar el libro de códigos **CBK1**, solo se emplean los primeros cuatro elementos **a1, a2, a3, a4** de las matrices de descriptores de capacitación de 16 tuplas utilizadas para generar el único libro de códigos **CBK** de la **Figura 4A**.

- [0071]** En comparación con el caso de la **Figura 4A**, en el que toda la matriz de descriptores **DA** se comprime utilizando un único libro de códigos **CBK** formado por las palabras de código **CWj** que tienen la misma dimensión que la propia matriz de descriptores **DA** (16 elementos), el uso de los libros de códigos **CBKy** formados por las palabras de código **CWj** que tienen una dimensión (más pequeña)  $mk$  de una submatriz **SDAk** del mismo (por ejemplo,  $mk = 4$  elementos) permite obtener, con un mismo número de palabras de código **CWj**, una distorsión menor.

- 60 **[0072]** Una vez fijado el número total de palabras de código **CWj**, cuanto mayor sea el número de submatrices **SDAk** en las que se subdivide la matriz de descriptores **DA**, menor será la distorsión, pero, al mismo tiempo, mayor será el número mínimo de bits necesarios para identificar todas las palabras de código **CWj**.

- [0073]** De acuerdo con una realización de la presente invención, la subdivisión de la matriz de descriptores **DA** en submatrices **SDAk** para fines de compresión se lleva a cabo teniendo en cuenta la aparición de relaciones de

correlación entre los elementos **ai** de la matriz de descriptores **DA**.

**[0074]** Como ya se describió con referencia a las **Figuras 3A y 3B**, cada elemento **ai** de la matriz de descriptores **DA** almacena los valores tomados por el subhistograma **shi** asociado con una subregión respectiva, cuya subregión corresponde a su vez a una celda de la cuadrícula 4x4 centrada en el punto clave **KP** correspondiente a dicha matriz de descriptores **DA**.

**[0075]** De acuerdo con una realización de la presente invención ilustrada en la **Figura 5**, después de haber llevado a cabo un análisis estadístico del comportamiento en una gran cantidad de matrices de descriptores **DA** (por ejemplo, explotando las matrices de descriptores de capacitación de la base de datos de capacitación), se ha encontrado que los subhistogramas **shi** de un punto clave **KP** genérico se pueden disponer en familias de correlación **CFx** ( $x = 1, 2, 3, 4$ ), donde cada familia de correlación **CFx** comprende un conjunto de subhistogramas **shi** correlacionados con un comportamiento estadístico similar, es decir, con una tendencia similar de las frecuencias de caja. Por ejemplo, dos subhistogramas **shi** que pertenecen a una misma familia de correlación **CFx** pueden tener un número similar de picos de frecuencia en las mismas cajas (o similares).

**[0076]** El análisis estadístico de comportamiento empleado para formar las familias de correlación **CFx** mostró que, habiendo fijado el número máximo de palabras de código **CWj** que se utilizarán para comprimir la matriz de descriptores **DA**, si la disposición de los subhistogramas **shi** en las familias de correlación **CFx** es variada (asignando los subhistogramas **shi** a diferentes familias de correlación **CFx**), la distorsión resultante varía en consecuencia. Por lo tanto, las familias de correlación **CFx** se forman considerando, entre todas las subdivisiones **shi** de subhistogramas posibles, la que corresponde a la distorsión más baja.

**[0077]** Después de haber realizado dicho análisis estadístico conductual, también se ha encontrado que la correlación entre el comportamiento estadístico de dos subhistogramas **shi** depende de dos parámetros principales, es decir, la distancia de las subregiones asociadas a los subhistogramas **shi** desde el punto clave **KP** y la orientación dominante de los mismos.

**[0078]** Haciendo referencia a la **Figura 5**, los dieciséis subhistogramas **shi** de un punto clave **KP** se disponen en cuatro familias de correlación, es decir:

- una primera familia de correlación **CF1** que comprende los subhistogramas **sh1, sh4, sh13 y sh16**;
- una segunda familia de correlación **CF2** que comprende los subhistogramas **sh2, sh3, sh14 y sh15**;
- una tercera familia de correlación **CF3** que comprende los subhistogramas **sh5, sh8, sh9 y sh12**, y
- una cuarta familia de correlación **CF4** que comprende los subhistogramas **sh6, sh7, sh10 y sh11**.

**[0079]** De acuerdo con una realización de la presente invención, las familias de correlación identificadas anteriormente **CFx** se explotan ventajosamente para comprimir la matriz de descriptores **DA** usando un número reducido de libros de códigos optimizados **CBKy**. La subdivisión de la matriz de descriptores **DA** en submatrices **SDAk** se lleva a cabo de tal manera que al menos dos submatrices **SDAk** tengan el mismo comportamiento estadístico global (es decir, considerando todos los elementos de este); de esta manera, es posible usar un solo libro de códigos **CBKy** para comprimir más de una submatriz **SDAk**. A tal efecto, la subdivisión de la matriz de descriptores **DA** se lleva a cabo de tal manera que se obtiene(n) un grupo o grupos de submatrices **SDAk** en el(los) que para cada grupo los elementos **ai** que ocupan la misma posición en todas las submatrices **SDAk** del grupo pertenecen a una misma familia de correlación **CFx**. Por lo tanto, todas las submatrices **SDAk** que pertenecen a un mismo grupo se pueden comprimir ventajosamente utilizando un mismo libro de códigos correspondiente **CBKy**, cuyas palabras de código **CWj** se obtienen considerando, de cada matriz de descriptores de capacitación utilizada para generar el único libro de códigos **CBK** de la **Figura 4A**, solo los elementos de los mismos que pertenecen a las familias de correlación **CFx** a las que pertenecen los elementos **ai** de las submatrices **SDAk** del grupo.

**[0080]** De acuerdo con una realización ejemplar de la presente invención ilustrada en la **Figura 6A**, la matriz de descriptores **DA** se subdivide en cuatro submatrices **SDA1-SDA4** que están dispuestas en un solo grupo. Por lo tanto, todas las submatrices **SDAk** se comprimen utilizando un mismo libro de códigos **CBK1**. Específicamente:

- la primera submatriz **SDA1** está formada por la secuencia de elementos **a1, a2, a6, a5**;
- la segunda submatriz **SDA2** está formada por la secuencia de elementos **a4, a3, a7, a8**;
- la tercera submatriz **SDA3** está formada por la secuencia de elementos **a16, a15, a11, a12**, y
- la cuarta submatriz **SDA4** está formada por la secuencia de elementos **a13, a14, a10, a9**.

**[0081]** En este caso:

- los primeros elementos **ai** de cada submatriz **SDAk** pertenecen a la primera familia de correlación **CF1**;
- los segundos elementos **ai** de cada submatriz **SDAk** pertenecen a la segunda familia de correlación **CF2**;
- los terceros elementos **ai** de cada submatriz **SDAk** pertenecen a la cuarta familia de correlación **CF4**, y
- los cuartos elementos **ai** de cada submatriz **SDAk** pertenecen a la tercera familia de correlación **CF3**.

**[0082]** El libro de códigos **CBK1** para comprimir la submatriz genérica **SDA1-SDA4** incluye  $N3$  palabras de código **CWj**, donde cada palabra de código **CWj** tiene el primer elemento que pertenece a la primera familia de correlación **CF1**, el segundo elemento que pertenece a la segunda familia de correlación **CF2**, el tercer elemento que pertenece a la cuarta familia de correlación **CF4** y el cuarto elemento que pertenece a la tercera familia de correlación **CF3**.

**[0083]** Con  $N3$  palabras de código **CWj** diferentes, el número mínimo de bits requeridos para identificar todas las palabras de código es igual a  $4^*(\log_2 N3)$ .

**[0084]** De acuerdo con otra realización ejemplar de la presente invención ilustrada en la **Figura 6B**, la matriz de descriptores **DA** está subdividida en dos submatrices **SDA1**, **SDA2** que están dispuestas en un solo grupo. Por lo tanto, todas las submatrices **SDAk** se comprimen utilizando un mismo libro de códigos **CBK1**. Específicamente:

- la primera submatriz **SDA1** está formada por la secuencia de elementos **a1, a2, a3, a4, a5, a6, a7, a8**, y
- la segunda submatriz **SDA2** está formada por la secuencia de elementos **a13, a14, a15, a16, a9, a10, a11, a12**.

**[0085]** En este caso:

- el primer y el cuarto elementos **ai** de cada submatriz **SDAk** pertenecen a la primera familia de correlación **CF1**;
- el segundo y el tercer elementos **ai** de cada submatriz **SDAk** pertenecen a la segunda familia de correlación **CF2**;
- el quinto y el octavo elementos **ai** de cada submatriz **SDAk** pertenecen a la tercera familia de correlación **CF3**, y
- el sexto y el séptimo elementos **ai** de cada submatriz **SDAk** pertenecen a la cuarta familia de correlación **CF4**.

**[0086]** El libro de códigos **CBK1** para comprimir la submatriz genérica **SDA1, SDA2** incluye  $N4$  palabras de código **CWj**, donde cada palabra de código **CWj** tiene el primer y el cuarto elementos que pertenecen a la primera familia de correlación **CF1**, el segundo y el tercer elementos que pertenecen a la segunda familia de correlación **CF2**, el quinto y el octavo elementos que pertenecen a la tercera familia de correlación **CF3**, y el sexto y el séptimo elementos que pertenecen a la tercera familia de correlación **CF3**.

**[0087]** Con  $N4$  palabras de código **CWj** diferentes, el número mínimo de bits requeridos para identificar todas las palabras de código es igual a  $2^*(\log_2 N4)$ .

**[0088]** De acuerdo con otra realización ejemplar de la presente invención ilustrada en la **Figura 6C**, la matriz de descriptores **DA** se subdivide en seis submatrices **SDA1-SDA6**, cuatro de las cuales (**SDA1-SDA4**) están dispuestas en un primer grupo, y dos de las cuales (**SDA5, SDA6**) están dispuestas en un segundo grupo. Por lo tanto, las submatrices **SDA1-SDA4** se comprimen usando un mismo primer libro de códigos **CBK1**, mientras que las submatrices **SDA5-SDA6** se comprimen usando un mismo segundo libro de códigos **CBK2**. Específicamente:

- la primera submatriz **SDA1** está formada por la secuencia de elementos **a5, a1, a2**;
- la segunda submatriz **SDA2** está formada por la secuencia de elementos **a8, a4, a3**;
- la tercera submatriz **SDA3** está formada por la secuencia de elementos **a9, a13, a14**;
- la cuarta submatriz **SDA4** está formada por la secuencia de elementos **a12, a16, a15**;
- la quinta submatriz **SDA5** está formada por la secuencia de elementos **a6, a7** y
- la sexta submatriz **SDA6** está formada por la secuencia de elementos **a10, a11**. En este caso:
- los primeros elementos **ai** de cada submatriz **SDA1-SDA4** del primer grupo pertenecen a la tercera familia de correlación **CF3**;
- los segundos elementos **ai** de cada submatriz **SDA1-SDA4** del primer grupo pertenecen a la primera familia de correlación **CF1**;
- los terceros elementos **ai** de cada submatriz **SDA1-SDA4** del primer grupo pertenecen a la segunda familia de correlación **CF2**, y
- los elementos primero y segundo **ai** de cada submatriz **SDA5-SDA6** del segundo grupo pertenecen a la cuarta familia de correlación **CF4**.

**[0089]** El libro de códigos **CBK1** para comprimir la submatriz genérica **SDA1-SDA4** que pertenece al primer grupo incluye  $N5$  palabras de código **CWj**, donde cada palabra de código **CWj** tiene el primer elemento que pertenece a la tercera familia de correlación **CF3**, el segundo elemento que pertenece a la primera familia de correlación **CF1**, y el tercer elemento que pertenece a la segunda familia de correlación **CF2**. El libro de códigos **CBK2** para comprimir la submatriz genérica **SDA5-SDA6** que pertenece al segundo grupo incluye  $N6$  palabras de código **CWj**, donde cada palabra de código **CWj** tiene los primer y segundo elementos que pertenecen a la cuarta familia de correlación **CF4**.

**[0090]** Con  $N5 + N6$  diferentes palabras de código **CWj**, el número mínimo de bits necesarios para identificar todas las palabras de código es igual a  $4^*(\log_2 N5) + 2^*(\log_2 N6)$ .

[0091] De acuerdo con otra realización ejemplar de la presente invención ilustrada en la **Figura 6D**, la matriz de descriptores **DA** está subdividida en ocho submatrices **SDA1-SDA8**, cuatro de las cuales (**SDA1-SDA4**) están dispuestas en un primer grupo, y cuatro de las cuales (**SDA5-SDA8**) están dispuestas en un segundo grupo. Por lo tanto, las submatrices **SDA1-SDA4** se comprimen usando un mismo primer libro de códigos **CBK1**, mientras que las submatrices **SDA5-SDA8** se comprimen usando un mismo segundo libro de códigos **CBK2**. Específicamente:

- la primera submatriz **SDA1** está formada por la secuencia de elementos **a5, a1**;
- la segunda submatriz **SDA2** está formada por la secuencia de elementos **a8, a4**;
- la tercera submatriz **SDA3** está formada por la secuencia de elementos **a9, a13**;
- la cuarta submatriz **SDA4** está formada por la secuencia de elementos **a12, a16**;
- la quinta submatriz **SDA5** está formada por la secuencia de elementos **a6, a2**;
- la sexta submatriz **SDA6** está formada por la secuencia de elementos **a7, a3**;
- la séptima submatriz **SDA7** está formada por la secuencia de elementos **a10, a14** y
- la octava submatriz **SDA8** está formada por la secuencia de elementos **a11, a15**.

[0092] En este caso:

- los primeros elementos **ai** de cada submatriz **SDA1-SDA4** del primer grupo pertenecen a la tercera familia de correlación **CF3**;
- los segundos elementos **ai** de cada submatriz **SDA1-SDA4** del primer grupo pertenecen a la primera familia de correlación **CF1**;
- los primeros elementos **ai** de cada submatriz **SDA5-SDA8** del segundo grupo pertenecen a la cuarta familia de correlación **CF4**, y
- los segundos elementos **ai** de cada submatriz **SDA5-SDA8** del segundo grupo pertenecen a la segunda familia de correlación **CF2**.

[0093] El libro de códigos **CBK1** para comprimir la submatriz genérica **SDA1-SDA4** que pertenece al primer grupo incluye **N7** palabras de código **CWj**, donde cada palabra de código **CWj** tiene el primer elemento que pertenece a la tercera familia de correlación **CF3**, y el segundo elemento que pertenece a la primera familia de correlación **CF1**. El libro de códigos **CBK2** para comprimir la submatriz genérica **SDA5-SDA8** que pertenece al segundo grupo incluye **N8** palabras de código **CWj**, donde cada palabra de código **CWj** tiene los primeros elementos que pertenecen a la cuarta familia de correlación **CF4** y los segundos elementos que pertenecen a la segunda familia de correlación **CF2**.

[0094] Por lo tanto, con **N7 + N8** diferentes palabras de código **CWj**, el número mínimo de bits necesarios para identificar todas las palabras de código es igual a  $4^{(\log_2 N7)} + 4^{(\log_2 N8)}$ .

[0095] Naturalmente, los conceptos de la presente invención también son aplicables con subdivisiones en un número diferente de submatrices y/o con un número diferente de libros de códigos. Además, incluso si en la presente descripción se ha hecho referencia a la compresión de un descriptor de SIF calculado en una cuadrícula que incluye celdas 4x4 con ocho cajas por histograma, se aplica una consideración similar si la cantidad de celdas y/o la cantidad de cajas por histograma es diferente, así como también se consideran descriptores de otros tipos.

[0096] En comparación con las soluciones conocidas, con una misma distorsión de compresión, el uso combinado de subdividir la matriz de descriptores **DA** en submatrices **SDAk** y emplear un mismo libro de códigos **CBKy** para más de una submatriz **SDAk** permite reducir drásticamente el espacio de memoria necesario para almacenar el o los libros de códigos **CBKy** utilizados para comprimir la matriz de descriptores **DA**. Esta es una gran ventaja, ya que, como ya se mencionó anteriormente, el hardware empleado para las aplicaciones de análisis de imágenes (por ejemplo, Unidades de Procesamiento Gráfico, GPU) puede estar equipado con memorias no tan potentes. Otra ventaja dada por el uso combinado de subdividir la matriz de descriptores **DA** en submatrices **SDAk** y emplear un mismo libro de códigos **CBKy** para más de una submatriz **SDAk** consiste en que el procedimiento de capacitación para la generación del o de los libros de códigos **CBKy** resulta ser más rápido.

[0097] Las operaciones de compresión llevadas a cabo en la fase **140** del procedimiento de extracción **100** (véase la **Figura 1**) en cada descriptor **D** recibido generan como resultado una matriz de descriptores comprimida correspondiente **CDA**, que se aproxima al valor tomado por la respectiva matriz de descriptores **DA**. Más específicamente, para cada libro de códigos **CBKy** utilizado para comprimir la matriz de descriptores **DA**, cada palabra de código **CWj** de dicho libro de códigos **CBKy** se identifica mediante un índice de compresión correspondiente **Cy**; si el libro de códigos **CBKy** está formado por un número **N** de diferentes palabras de código **CWj**, el índice de compresión **Cy** está formado por al menos  $\log_2 N$  bits. Para una matriz de descriptores **DA** que se ha subdividido en un conjunto de submatrices **SDAk**, la matriz de descriptores comprimida correspondiente **CDA** comprende un índice de compresión **Cy** para cada submatriz **SDAk** del conjunto, donde cada índice de compresión **Cy** identifica la palabra de código **CWj** del libro de códigos **CBKy** utilizado para aproximarse a dicha submatriz **SDAk**.

#### Compresión de las coordenadas (fase 150)

**[0098]** De acuerdo con una realización de la presente invención, la cantidad de datos a gestionar (por ejemplo, a memorizar y/o transmitir) para realizar operaciones de análisis de imágenes se reduce adicionalmente al comprimir las coordenadas **C** de los puntos clave **KP** que pertenecen al subconjunto óptimo **SUB** calculado en la fase **130** del procedimiento de extracción **100** (véase la **Figura 1**).

**[0099]** La **Figura 7A** ilustra una distribución ejemplar de los puntos clave **KP** del subconjunto óptimo **SUB** dentro de un espacio bidimensional correspondiente a la imagen de consulta **115**; cada punto clave **KP** está asociado con un par correspondiente de coordenadas espaciales **C** que identifican la ubicación de dicho punto clave **KP** dentro de la imagen de consulta **115**.

**[0100]** En primer lugar, se cuantifican las coordenadas **C** de todos los puntos clave **KP** del subconjunto **SUB**. Para este propósito, se superpone una cuadrícula  $n \times m$  sobre la imagen de consulta **115**. En el ejemplo ilustrado en la **Figura 7B**, la cuadrícula tiene  $n = 10$  filas y  $m = 15$  columnas.

**[0101]** A continuación, se genera un histograma bidimensional al contar para cada celda de la cuadrícula (correspondiente a una caja del histograma) el número de puntos clave **KP** que se encuentran dentro de esta. La **Figura 7C** es una representación gráfica ejemplar del histograma obtenido mediante la superposición de la cuadrícula de la **Figura 7B** sobre el conjunto de puntos clave **KP** de la **Figura 7A**. En la representación gráfica de la **Figura 7C**, las celdas vacías de puntos clave **KP** están coloreadas en negro, mientras que las celdas que incluyen al menos un punto clave **KP** están coloreadas en gris. En el ejemplo en cuestión (donde las celdas que incluyen el mayor número de puntos clave incluyen dos puntos clave), las celdas que incluyen un único punto clave **KP** están coloreadas en gris oscuro, mientras que las que incluyen dos puntos clave **KP** están coloreadas en un gris más claro.

**[0102]** El histograma obtenido del recuento de puntos clave tiene una gran cantidad de cajas cuya frecuencia es igual a cero, es decir, con la celda correspondiente que no incluye ningún punto clave **KP** (las celdas negras representadas en la **Figura 7C**).

**[0103]** Los datos que representan el histograma pueden comprimirse ventajosamente teniendo en cuenta que las porciones de estos correspondientes a las cajas de frecuencia cero solo proporcionan la información de que su celda correspondiente no incluye ningún punto clave.

**[0104]** Para este propósito, las filas y las columnas de la cuadrícula que están completamente formadas por celdas que no incluyen ningún punto clave **KP** pueden eliminarse ventajosamente. Sin embargo, dado que la eliminación de tales filas y/o columnas alteraría las posiciones absolutas y relativas de los puntos clave **KP**, se debe registrar una indicación de las posiciones de todas las filas y columnas vacías de puntos clave **KP** (que comprenden las correspondientes a las filas y/o columnas que se eliminarán).

**[0105]** Para ello, se definen dos matrices  $r$  y  $c$  de la siguiente manera:

- la matriz  $r$  es una matriz que incluye un elemento para cada fila de la cuadrícula, donde el elemento genérico de la matriz se establece en un primer valor (por ejemplo, 0) si la celda correspondiente de la cuadrícula no incluye ningún punto clave **KP**, y se establece en un segundo valor (por ejemplo, 1) si la celda correspondiente de la cuadrícula incluye al menos un punto clave **KP**, y

- la matriz  $c$  es una matriz que incluye un elemento para cada columna de la cuadrícula, donde el elemento genérico de la matriz se establece en un primer valor (por ejemplo, 0) si la celda correspondiente de la cuadrícula no incluye ningún punto clave **KP**, y se establece en un segundo valor (por ejemplo, 1) si la celda correspondiente de la cuadrícula incluye al menos un punto clave **KP**.

**[0106]** Una vez que se han generado las matrices  $r$  y  $c$ , la siguiente etapa proporciona identificar las filas y/o las columnas que están formadas en su totalidad por celdas que no incluyen ningún punto clave **KP**. Haciendo referencia al ejemplo en cuestión, dichas filas y columnas se representan en negro en la **Figura 7D**.

**[0107]** Las filas y/o las columnas de la cuadrícula que están completamente formadas por celdas que no incluyen ningún punto clave **KP** se eliminan a continuación, y las porciones resultantes de la cuadrícula se compactan para llenar los espacios vacíos dejados por las eliminaciones. Por lo tanto, en la cuadrícula resultante (compactada), denominada *soporte de rango 1*, todas las filas y todas las columnas incluyen al menos una celda que comprende al menos un punto clave **KP**. El histograma sobre el *soporte de rango 1* correspondiente al ejemplo en cuestión se ilustra en la **Figura 7E**.

**[0108]** De dicho histograma se pueden extraer dos fragmentos de información diferentes, es decir:

- 1) las posiciones de las celdas del *soporte de rango 1* que incluyen al menos un punto clave **KP**, y
- 2) para cada celda del *soporte de rango 1* identificada en el punto 1), el número de puntos clave **KP** incluidos en el mismo.

[0109] Ventajosamente, como proponen S. Tsai, D. Chen, G. Takacs, V. Chandrasekhar, J. P. Singh y B. Girod en «Location coding for mobile image retrieval», Proc. Int. Mobile Multimedia Conference (MobiMedia), 2009, la información correspondiente al punto 1) puede extraerse explotando un denominado «mapa de histogramas», mientras que la información correspondiente al punto 2) puede disponerse en un denominado «recuento de histogramas».

[0110] El mapa de histogramas es un mapeo bidimensional del histograma sobre el *soporte de rango 1* que identifica las cajas del mismo que tienen una frecuencia igual o superior a 1. El mapa de histogramas correspondiente al histograma sobre el *soporte de rango 1* de la **Figura 7E** se ilustra en la **Figura 7F**.

10 [0111] El mapa de histogramas se puede representar con una matriz correspondiente, cuyo elemento genérico es igual a cero si la celda correspondiente del *soporte de rango 1* no incluye ningún punto clave **KP**, y es igual a uno si la celda correspondiente del *soporte de rango 1* incluye al menos un punto clave **KP**. La matriz del mapa de histogramas ilustrado en la **Figura 7F** es la siguiente:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

[0112] De acuerdo con una realización de la presente invención, la información proporcionada por el mapa de histogramas se puede comprimir ventajosamente usando una codificación entrópica optimizada en función del comportamiento estadístico de histogramas de *soporte de rango 1* ejemplares aprendidos del análisis de una gran cantidad de imágenes de capacitación.

[0113] A partir de dicho análisis se ha encontrado que las ubicaciones de los puntos clave **KP** dentro de la imagen genérica son tales que implican una distribución estadística común del «1» dentro de la matriz del mapa de histogramas.

[0114] La codificación entrópica se lleva a cabo de la siguiente manera.

[0115] La matriz del mapa de histogramas se escanea (por ejemplo, columna por columna) para subdividirla en una pluralidad de palabras, cada una con una misma longitud  $x$ . Con base en el análisis estadístico llevado a cabo en las imágenes de capacitación, se genera un histograma de palabras que incluye una caja para cada posible valor que la  $x$ -tupla de la palabra genérica puede tomar, con la frecuencia de cada caja que indica la probabilidad de que la  $x$ -tupla de la palabra tome el valor asociado con dicha caja. Brevemente, dicho análisis estadístico se ha llevado a cabo asumiendo que los elementos de la matriz del mapa de histogramas son independientes entre sí. Analizando un número muy alto de imágenes de capacitación, se puede identificar cuál es la probabilidad de que ocurra un «1» en la matriz cada  $n$  «0»; entonces, la palabra histograma se genera en función de dicha probabilidad.

[0116] La **Figura 8A** ilustra un ejemplo de un histograma de palabras en el que la longitud  $x$  de las palabras es igual a seis, y donde cada caja se identifica por el valor decimal del valor de  $x$ -tupla correspondiente. Como se esperaba, la frecuencia más alta corresponde a la  $x$ -tupla (0,0,0,0,0,0), ya que hay una probabilidad muy alta de que la celda genérica del *soporte de rango 1* no incluya ningún punto clave **KP**. La probabilidad más alta que se indica a continuación es la que corresponde a un único punto clave **KP** para la celda ( $x$ -tupla (1,0,0,0,0,0), (0,1,0,0,0,0), (0,0,1,0,0,0), (0,0,0,1,0,0), (0,0,0,0,1,0), (0,0,0,0,0,1)), entonces corresponde a dos puntos clave **KP** para la celda, y así sucesivamente.

[0117] Las palabras se codifican con una técnica de codificación entrópica (por ejemplo, la técnica de Huffman o la técnica de codificación aritmética) utilizando para cada palabra una palabra codificada  $bci$  ( $i = 1, 2, \dots$ ) que tiene un número de bits que depende de la probabilidad de la caja correspondiente en el histograma de palabras. Cuanto mayor sea la probabilidad de la palabra, menor será el número de bits de la palabra codificada  $bci$  utilizada para codificar dicha palabra.

[0118] La otra información que se puede extraer del histograma sobre el *soporte de rango 1* se refiere a la cantidad de puntos clave **KP** que se incluyen en cada celda del mapa de histogramas que comprende al menos un punto clave **KP**. Dicha información se organiza en un histograma correspondiente, denominado recuento de

histogramas. Cada caja del recuento de histogramas corresponde a uno correspondiente entre las células del *soporte de rango 1* que incluye al menos un punto clave **KP**. El recuento de histogramas enumera para cada caja el número de puntos clave **KP** incluidos en la celda correspondiente. El mapa de histogramas del ejemplo en cuestión se ilustra en la **Figura 8B**, donde 11 celdas incluyen un único punto clave **KP** cada una y dos celdas incluyen dos puntos clave **KP** cada una. Las cajas del mapa de histogramas de la **Figura 8B** se ordenan después de un escaneo por columnas del *soporte de rango 1*.

**[0119]** La información de recuento de puntos clave proporcionada por el recuento de histogramas se codifica en un conjunto de palabras codificadas  $w_j$  ( $j = 1, 2, \dots$ ) de diferentes longitudes, con cada palabra codificada  $w_j$  del conjunto que indica qué caja(s) de un conjunto respectivo de cajas de recuento de histogramas corresponden a un número de puntos clave **KP** mayor o igual a un valor determinado.

**[0120]** Más específicamente, si el número más alto de puntos clave **KP** contados dentro de cada caja es igual a  $N_{\text{máx}}$ , dicho conjunto de palabras codificadas  $w_j$  comprende un número de palabras codificadas  $w_j$  igual a  $N_{\text{máx}}-2$ . La generación de cada palabra codificada  $w_j$  se lleva a cabo realizando una correspondiente entre un conjunto de etapas de procedimiento  $N_{\text{máx}}-2$ . De acuerdo con una realización de la presente invención, dichas etapas de procedimiento se describen a continuación.

**[0121]** **Etapas 1** - Una primera palabra codificada  $w_1$  está configurada para incluir un elemento para cada caja del mapa de histogramas. Por lo tanto, la primera palabra codificada  $w_1$  incluye un número de elementos igual al número de cajas del mapa de histogramas. Cada elemento de la primera palabra codificada  $w_1$  se establece en un primer valor (por ejemplo, «1») si la caja correspondiente del recuento de histogramas corresponde a un número de puntos clave **KP** mayor que uno, de lo contrario se establece en un segundo valor (por ejemplo, «0»). Si  $N_{\text{máx}}$  es mayor que 2, se realiza una segunda etapa para generar una segunda palabra codificada  $w_2$ , de lo contrario se termina el procedimiento. En este último caso, toda la información proporcionada por el recuento de histogramas resulta codificada con la primera palabra codificada  $w_1$  solamente.

**[0122]** **Etapas  $j$  ( $j > 1$ )** - Se genera una  $j$ -ésima palabra codificada  $w_j$ . La  $j$ -ésima palabra codificada  $w_j$  está configurada para incluir un elemento para cada caja del mapa de histogramas que incluye más de  $j$  puntos clave **KP**. Por lo tanto, la  $j$ -ésima palabra codificada  $w_j$  incluye un número de elementos igual o inferior a la  $j-1$  palabra codificada  $w_{j-1}$ . Cada elemento de la  $j$ -ésima palabra codificada  $w_j$  se establece en el primer valor si la caja correspondiente del recuento de histogramas corresponde a un número de puntos clave **KP** mayor que  $j$ , de lo contrario se establece en el segundo valor. Si  $N_{\text{máx}}$  es mayor que  $j+1$ , se realiza una etapa  $(j+1)$ -ésima, para generar una palabra codificada  $w_{j+1}$ , de lo contrario el procedimiento se termina. En este último caso, toda la información proporcionada por el recuento de histogramas se codifica con las palabras codificadas  $w_1 - w_j$ .

**[0123]** Las operaciones de compresión llevadas a cabo en la fase **150** del procedimiento de extracción **100** (véase la **Figura 1**) permiten obtener para las coordenadas **C** de los puntos clave **KP** que pertenecen al subconjunto **SUB** un conjunto de coordenadas comprimidas **CC** correspondiente que comprende:

- la matriz  $r$  y la matriz  $c$ ;
- las palabras codificadas  $bci$ , y
- las palabras codificadas  $w_j$ .

**[0124]** La cantidad de datos necesarios para gestionar (memorizar y/o transmitir) el conjunto de coordenadas comprimidas **CC** es sensiblemente menor que la cantidad de datos necesarios para gestionar el conjunto de coordenadas (no comprimidas) **C**.

## **PROCEDIMIENTO DE COINCIDENCIA (Figura 9)**

**[0125]** La **Figura 9** ilustra en términos de bloques funcionales un procedimiento de análisis de imágenes de acuerdo con una realización de la presente invención, en lo sucesivo denominado «procedimiento de coincidencia» e identificado con la referencia **900**, dirigido a realizar la comparación entre dos imágenes **I1**, **I2**, mediante la explotación para cada imagen de un subconjunto óptimo respectivo de puntos clave y los correspondientes descriptores y coordenadas comprimidos generados con el procedimiento de extracción **100** de la **Figura 1**.

**[0126]** Las etapas del procedimiento de coincidencia **900** pueden llevarse a cabo por unidades de procesamiento adecuadas; por ejemplo, cada unidad de procesamiento puede ser una unidad de hardware diseñada específicamente para realizar una o más etapas del procedimiento. Un posible escenario puede proporcionar un usuario (lado del cliente) que desea explotar un servicio de comparación de imágenes (lado del servidor) para comparar la imagen **I1** con la imagen **I2**. En este caso, las imágenes **I1** e **I2** pueden procesarse en el cliente de acuerdo con el procedimiento de extracción **100** de la **Figura 1** para la generación del subconjunto óptimo de puntos clave y los descriptores y coordenadas comprimidos correspondientes; a continuación, el subconjunto óptimo de puntos clave y los correspondientes descriptores y coordenadas comprimidos se envían al servidor, que realiza el procedimiento de coincidencia **900** explotando los datos recibidos y, a continuación, proporciona los resultados al

cliente. En este caso, el procedimiento de extracción **100** puede llevarse a cabo mediante unidades de procesamiento ubicadas en el cliente, por ejemplo, mediante el teléfono inteligente de un usuario, mientras que el procedimiento de coincidencia **900** puede llevarse a cabo mediante unidades de procesamiento ubicadas en el servidor, por ejemplo, mediante una o más unidades de servidor adaptadas para ofrecer servicios de comparación de imágenes. Otro escenario posible puede proporcionar en cambio que el procedimiento de coincidencia **900** se realice directamente en el cliente. También se contemplan escenarios mixtos, en los que el procedimiento de coincidencia **900** se lleva a cabo en el cliente con los descriptores y las coordenadas comprimidos enviados por el servidor.

[0127] Las coordenadas comprimidas de la imagen **I1** se identifican con la referencia **CC1**, mientras que los descriptores comprimidos de la imagen **I1** se identifican con la referencia **CDA1**. De manera similar, las coordenadas comprimidas de la imagen **I2** se identifican con la referencia **CC2**, mientras que los descriptores comprimidos de la imagen **I2** se identifican con la referencia **CDA2**.

[0128] Los descriptores comprimidos **CDA1** de la primera imagen **I1** se descomprimen para recuperar los descriptores correspondientes (descomprimidos) **D1** (fase **902**). De manera similar, los descriptores comprimidos **CDA2** de la segunda imagen **I2** se descomprimen para recuperar los descriptores correspondientes (descomprimidos) **D2** (fase **904**). La descompresión de los descriptores puede llevarse a cabo mediante versiones inversas de las operaciones de compresión realizadas en la fase **140** del procedimiento de extracción **100**. Haciendo referencia a los descriptores del tipo SIFT, después de las fases **902** y **904**, los descriptores **D1** y **D2** están representados por lo tanto por matrices de descriptores correspondientes formadas por 128 subelementos.

[0129] En la fase **906**, las coincidencias entre los descriptores **D1** de la primera imagen **I1** y los descriptores **D2** de la segunda imagen **I2** se forman explotando uno cualquiera de los algoritmos de coincidencia de características conocidos en la técnica, tales como, por ejemplo, la prueba de relación de distancia euclidiana.

[0130] A continuación, en la fase **908**, se realizan operaciones de verificación geométrica para determinar qué coincidencias entre las formadas en la fase **906** son correctas (inliers) y qué coincidencias no están corregidas (outliers). Como saben los expertos en la materia, una operación de este tipo requiere, además de los descriptores, las coordenadas de cada punto clave cuyo descriptor correspondiente se ha emparejado con el descriptor de otro punto clave. Para este propósito, las coordenadas comprimidas **CC1** de la imagen **I1** y las coordenadas comprimidas **CC2** de la imagen **I2** también deben descomprimirse, por ejemplo, por medio de versiones inversas de las operaciones de compresión realizadas en la fase **150** del procedimiento de extracción **100**. La fase dedicada a la descompresión de las coordenadas comprimidas **CC1** se identifica en la **Figura 9** con referencia **910**, mientras que la fase dedicada a la descompresión de las coordenadas comprimidas **CC2** se identifica en la **Figura 9** con referencia **912**. Una vez que se han identificado los inliers, la verificación geométrica puede proporcionar como resultado un parámetro **DOM** indicativo del grado de coincidencia entre la imagen **I1** y **I2**. Por ejemplo, si dicho parámetro **DOM** es mayor que un umbral predeterminado, las imágenes **I1** y **I2** son consideradas por representar un mismo objeto/escena.

[0131] Además, las operaciones de localización (fase **914**) se pueden llevar a cabo adicionalmente para recuperar la(s) ubicación(es) **L** de dicho(s) mismo(s) objeto(s)/escena(s) dentro de las dos imágenes **I1**, **I2**.

[0132] Haciendo referencia al escenario de comparación de imágenes cliente-servidor mencionado anteriormente, dado que el procedimiento de coincidencia **900** está configurado para operar con un número reducido de puntos clave (solo los que pertenecen al subconjunto **SUB** generado por medio del procedimiento de extracción **100**), y dado que los descriptores y las coordenadas de dicho número reducido de puntos clave se reciben de manera comprimida, con la solución propuesta la cantidad total de datos que se enviarán del cliente al servidor se reduce drásticamente en comparación con las soluciones conocidas.

#### **PROCEDIMIENTO DE RECUPERACIÓN (Figura 10)**

[0133] La **Figura 10** ilustra en términos de bloques funcionales un procedimiento de análisis de imágenes de acuerdo con una realización de la presente invención, en lo sucesivo denominado «procedimiento de recuperación» e identificado con la referencia **1000**, en el que una imagen de consulta, tal como la imagen de consulta **115** de la **Figura 1**, que representa un objeto/escena a reconocer se compara con una pluralidad de imágenes modelo, cada una de las cuales representa un objeto/escena conocido respectivo, almacenadas en una base de datos modelo, con el fin de recuperar la o las imágenes modelo que representan el mismo objeto/escena representado en la imagen de consulta.

[0134] Al igual que el procedimiento de coincidencia **900** de la **Figura 9**, las etapas del procedimiento de recuperación **1000** pueden llevarse a cabo mediante unidades de procesamiento adecuadas; por ejemplo, cada unidad de procesamiento puede ser una unidad de hardware diseñada específicamente para realizar una o más etapas del procedimiento. Un escenario típico puede proporcionar un usuario (lado del cliente) que desea explotar un servicio de reconocimiento de imágenes (lado del servidor) para reconocer automáticamente un objeto/escena representado en una imagen de consulta **115**. En este caso, la imagen de consulta **115** puede procesarse en el cliente de acuerdo con el procedimiento de extracción **100** de la **Figura 1** para la generación del subconjunto óptimo **SUB** de puntos clave y los correspondientes descriptores **CDA** y coordenadas **CC** comprimidos; a continuación, el subconjunto óptimo de

puntos clave y los correspondientes descriptores y coordenadas comprimidos se envían al servidor, que realiza el procedimiento de recuperación **1000** explotando los datos recibidos y a continuación proporciona los resultados al cliente. La pluralidad de imágenes modelo que se utilizarán para el reconocimiento del objeto/escena representado en la imagen de consulta **115** se almacenan en una base de datos modelo **1002**, que se encuentra en el lado del servidor.

**[0135]** Los descriptores comprimidos **CDA** se descomprimen para recuperar los descriptores correspondientes (descomprimidos) **DD** (fase **1004**). La descompresión de los descriptores puede llevarse a cabo mediante versiones inversas de las operaciones de compresión realizadas en la fase **140** del procedimiento de extracción **100**. Nuevamente, haciendo modelo a los descriptores del tipo SIFT, después de la fase **1004**, los descriptores **DD** están representados por matrices de descriptores correspondientes formadas por 128 subelementos.

**[0136]** Dado que un procedimiento de reconocimiento de objetos estándar típicamente requiere la ejecución de operaciones de comparación entre la imagen de consulta y un número muy alto de imágenes modelo (por ejemplo, las imágenes modelo incluidas en la base de datos modelo **1002** pueden ser unos pocos millones), dicho procedimiento consume tanto tiempo como memoria. Para este propósito, una solución conocida proporciona la realización de tales operaciones de comparación en dos fases distintas. En lugar de comparar directamente los descriptores de la imagen de consulta con los descriptores de todas las imágenes modelo, se realiza una comparación rápida, aproximada, preliminar entre las palabras visuales extraídas de la imagen de consulta y las palabras visuales extraídas de las imágenes modelo; entonces, la comparación (refinada) de los descriptores se lleva a cabo solo entre los descriptores de la imagen de consulta y los descriptores de un conjunto reducido de imágenes modelo seleccionadas en función de la comparación preliminar. Una palabra visual es una matriz obtenida al realizar una cuantificación vectorial de un descriptor; en otras palabras, cada palabra visual es una palabra de código de un libro de códigos visual. La generación de las palabras visuales se lleva a cabo para cada descriptor de la imagen de consulta y cada descriptor de las imágenes modelo. Por ejemplo, la comparación preliminar se lleva a cabo contando el número de palabras visuales en común entre la imagen de consulta y cada imagen modelo. A continuación, para cada imagen modelo, se calcula un rango de similitud basado en los recuentos del número de palabras visuales en común. Consideraciones similares se aplican si el rango de similitud se genera comparando las palabras visuales utilizando procedimientos alternativos. De esta manera, la comparación refinada entre descriptores puede llevarse a cabo ventajosamente solo entre la imagen de consulta y las imágenes modelo que tienen los rangos de similitud más altos (es decir, las que tienen los números más altos de palabras visuales en común con la imagen de consulta). Esta estrategia, que se deriva del campo de análisis de texto, también se conoce como «clasificación por medio de Bolsa de Características (BoF)».

**[0137]** Haciendo referencia nuevamente a la **Figura 10**, con el fin de permitir la realización de la clasificación por medio de BoF, se deben generar palabras visuales **VD** para cada descriptor de la imagen de consulta y palabras visuales **VDR** para cada descriptor de cada imagen modelo.

**[0138]** Se señala que, para permitir la comparación entre palabras visuales, tanto las palabras visuales **VD** como las palabras visuales **VDR** deben generarse utilizando un mismo libro de códigos.

**[0139]** Si bien las palabras visuales **VD** de la imagen de consulta **115** tienen que generarse cada vez que se realiza el procedimiento de recuperación **1000** (fase **1006**), para reducir drásticamente los tiempos de operación, la generación de las palabras visuales **VDR** de las imágenes modelo puede llevarse a cabo ventajosamente solo una vez, y a continuación la pluralidad resultante de palabras visuales **VDR** puede almacenarse directamente en la base de datos modelo **1002**; alternativamente, las palabras visuales **VDR** pueden actualizarse periódicamente.

**[0140]** Habiendo generado para cada descriptor **DD** de la imagen de consulta una palabra visual correspondiente **VD**, en la fase **1008** se lleva a cabo la clasificación por medio del procedimiento BoF. De esta manera, para cada imagen modelo, se calcula un índice de rango contando el número de palabras visuales **VDR** de dicha imagen modelo que también son palabras visuales **VD** de la imagen de consulta. Dicho recuento puede llevarse a cabo utilizando la clasificación conocida mediante la implementación de BoF también conocida como Índice invertido. Sin embargo, se aplican consideraciones similares en caso de que se apliquen diferentes implementaciones. Una vez calculados todos los índices de rango, se genera una lista en la que las imágenes modelo de la base de datos se ordenan según un orden decreciente del índice de rango. A continuación, se selecciona un conjunto **SR** de imágenes modelo que tienen los valores de índice de rango más altos para someterse a las operaciones de comparación subsiguientes (refinadas).

**[0141]** Se señala que dado que según una realización de la presente invención el número de descriptores de cada imagen se reduce ventajosamente, correspondiendo solo al subconjunto óptimo **SUB** de puntos clave que se consideran relevantes (véase la fase 130 del procedimiento de extracción **100** de la **Figura 1**), la cantidad de datos necesarios para llevar a cabo la clasificación por medio del procedimiento BoF (fase **1008**) que tienen que cargarse en la memoria de trabajo (por ejemplo, en bancos de RAM ubicados en el lado del servidor) se reduce fuertemente, mejorando drásticamente la velocidad del procedimiento. Además, dado que las comparaciones se realizan teniendo en cuenta solo los descriptores de los puntos clave considerados relevantes, la precisión de la comparación aumenta, ya que el ruido se reduce. Con el fin de mejorar aún más la velocidad y la precisión, también se genera un subconjunto óptimo que incluye un número reducido de descriptores para cada imagen modelo incluida en la base de datos modelo

1002.

[0142] Se ha encontrado que el número de puntos clave que forman el subconjunto óptimo **SUB** influye fuertemente en el resultado de la clasificación por medio de BoF. De hecho, con una misma cantidad de imágenes consideradas, la probabilidad de que el objeto/escena representado en la imagen de consulta **115** también se represente en al menos una de las imágenes modelo que pertenecen al conjunto seleccionado **SR** de imágenes modelo aumenta a medida que disminuye la cantidad de puntos clave del subconjunto óptimo **SUB**. Sin embargo, si dicha cantidad de puntos clave del subconjunto óptimo **SUB** cae por debajo de un umbral inferior, los rendimientos del procedimiento disminuyen, ya que la cantidad de puntos clave incluidos en el subconjunto **SUB** se vuelve demasiado pequeña para representar satisfactoriamente cada imagen.

[0143] En este punto, se lleva a cabo una segunda comparación refinada entre la imagen de consulta **115** y el conjunto **SR** de imágenes modelo (fase **1010**). Se puede emplear uno de los procedimientos de coincidencia de características ya conocidos para hacer coincidir los descriptores **DD** de la imagen de consulta **115** con descriptores de las imágenes modelo del conjunto **SR** (subfase **1012**), por ejemplo, mediante el cálculo de distancias euclidianas entre los descriptores, y a continuación se realiza una verificación geométrica para determinar qué coincidencia son inliers y cuáles son outliers (subfase **1014**). De esta manera, si existe, la imagen modelo **RI** del conjunto **SR** que representa un objeto/escena representado también en la imagen de consulta **115** se recupera al final de la fase.

[0144] De acuerdo con una realización de la presente invención, en lugar de realizar directamente operaciones de coincidencia de características en los descriptores **DD** de la imagen de consulta **115** y en los descriptores de las imágenes modelo del conjunto **SR**, las operaciones de coincidencia de características se llevan a cabo en sus versiones comprimidas obtenidas subdividiendo las matrices de descriptores correspondientes en submatrices y comprimiendo cada submatriz mediante un libro de códigos basado en la cuantificación vectorial. Para este propósito, los descriptores **DD** de la imagen de consulta **115** se comprimen en la fase **1016**, por ejemplo, subdividiendo las matrices de descriptores correspondientes en cuatro submatrices y comprimiendo cada una de dichas cuatro submatrices con un libro de códigos respectivo. De manera similar a la generación de las palabras visuales, la base de datos modelo **1002** almacena para cada imagen modelo sus versiones comprimidas precalculadas correspondientes, que se han comprimido utilizando los mismos libros de códigos utilizados para comprimir los descriptores **DD** de la imagen de consulta **115**. De acuerdo con esta realización, la coincidencia de características (subfase **1012**) se puede realizar de una manera muy rápida y eficiente. De hecho, dado que la coincidencia de características se lleva a cabo en el espacio comprimido (tanto los descriptores de la imagen de consulta como de las imágenes modelo están comprimidos), y dado que se reduce el número de descriptores a considerar (que corresponden solo a los puntos clave del subconjunto óptimo), es posible cargar directamente en la memoria principal también los datos que representan las imágenes modelo de la base de datos modelo. Además, dado que la compresión de las matrices de descriptores se ha llevado a cabo subdividiendo las matrices de descriptores en submatrices, reduciendo así fuertemente la cantidad de palabras de código de los libros de códigos correspondientes, una lista que incluye todas las distancias euclidianas posibles entre cada palabra de código de cada libro de códigos puede calcularse previamente y cargarse en la memoria principal, aumentando aún más la velocidad de la subfase **1012**. Consideraciones similares se aplican si la coincidencia de características se lleva a cabo explotando un algoritmo diferente que no hace uso de las distancias euclidianas.

[0145] De acuerdo con una realización de la presente invención, la subfase **1012** puede mejorarse adicionalmente mediante la compresión de las submatrices de cada descriptor utilizando un mismo libro de códigos, utilizando una estrategia similar a la utilizada en la fase **140** del procedimiento de extracción **100** de la **Figura 1**.

[0146] Dado que la verificación geométrica (subfase **1014**) requiere, además de los descriptores, las coordenadas de los puntos clave cuyos descriptores correspondientes se han emparejado con descriptores de otros puntos clave, las coordenadas comprimidas **CC** de los puntos clave de la imagen de consulta **115** también deben descomprimirse (fase **1018**).

[0147] La descripción anterior presenta y analiza en detalle varias realizaciones de la presente invención; no obstante, son posibles varios cambios en las realizaciones descritas, así como diferentes realizaciones de invención, sin apartarse del alcance definido por las reivindicaciones adjuntas.

## REIVINDICACIONES

1. Un procedimiento para procesar una imagen con el fin de comparar dicha imagen con al menos otra imagen, comprendiendo el procedimiento:

- identificar un primer grupo de puntos clave en la imagen;
- para cada punto clave del primer grupo:
  - a) identificar al menos una característica local de punto clave correspondiente relacionada con dicho punto clave;
  - b) para dicha al menos una característica local de punto clave, calcular una probabilidad de relevancia de característica local correspondiente mediante la comparación del valor asumido por dicha característica local con una distribución estadística de referencia correspondiente de los valores de dicha característica local, dicha distribución estadística de referencia es estadísticamente equivalente a una distribución estadística correspondiente generada al recopilar, entre una pluralidad de puntos clave de referencia identificados en una pluralidad de pares de imágenes de referencia, los valores de características locales correspondientes a esos puntos clave de referencia de cada par de imágenes de referencia que se ha determinado que implican una coincidencia de características correcta entre las imágenes de referencia de dicho par;
  - c) calcular una probabilidad de relevancia de puntos clave en función de las probabilidades de relevancia de características locales de dicha al menos una característica local;
- seleccionar puntos clave, entre los puntos clave del primer grupo, que tienen las mayores probabilidades de relevancia de puntos clave para formar un segundo grupo de puntos clave, y
- explotar los puntos clave del segundo grupo para comparar dicha imagen con dicha al menos otra imagen, donde:
  - cada distribución estadística de referencia está dispuesta en forma de un histograma correspondiente que tiene una pluralidad de cajas, correspondiendo cada caja a un intervalo predefinido de valores de la característica local correspondiente, y la frecuencia de cada caja corresponde a una relación entre:
    - a) el número de puntos clave de referencia que se ha determinado que implican una coincidencia de características correcta y que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y
    - b) el número total de puntos clave de referencia que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja,
- dicho cálculo de la probabilidad de relevancia de característica local para una característica local de un punto clave comprende:
  - c) inspeccionar el histograma correspondiente a dicha característica local con el fin de identificar la caja del mismo que se ajusta al valor asumido por la característica local del punto clave, y
  - d) establecer la probabilidad de relevancia de característica local a la frecuencia de la caja identificada.

2. El procedimiento de la reivindicación 1, donde dicha al menos una de la característica local de punto clave relacionada con dicho punto clave comprende al menos uno de los elementos entre:

- las coordenadas del punto clave;
- la escala a la que se ha identificado el punto clave;
- la orientación dominante del punto clave;
- el pico del punto clave, y
- un descriptor del punto clave.

3. El procedimiento de una cualquiera de las reivindicaciones anteriores, donde dicho cálculo de una probabilidad de relevancia de puntos claves de un punto clave del primer grupo incluye combinar las probabilidades de relevancia de las características locales de cada una de dicha al menos una característica local del punto clave correspondiente.

4. El procedimiento de la reivindicación 3, donde dicho cálculo de una probabilidad de relevancia de puntos claves de un punto clave del primer grupo incluye multiplicar entre sí las probabilidades de relevancia de características locales de cada una de dicha al menos una característica local del punto clave correspondiente.

5. Un procedimiento de una cualquiera de las reivindicaciones anteriores, donde dicha distribución estadística de referencia de valores de una característica local de punto clave se genera al:

- recopilar, entre una pluralidad de puntos clave de referencia identificados en una pluralidad de pares de imágenes de referencia, los valores de características locales correspondientes a esos puntos clave de referencia de cada par de imágenes de referencia que se ha determinado que implican una coincidencia de características correcta

entre las imágenes de referencia de dicho par.

6. Un sistema para procesar una imagen con el fin de comparar dicha imagen con al menos otra imagen, comprendiendo el sistema:

5

- una primera unidad de procesamiento configurada para identificar un primer grupo de puntos clave en la imagen;
- una segunda unidad de procesamiento configurada para realizar las siguientes operaciones para cada punto clave del primer grupo:

10

a) identificar al menos una característica local de punto clave correspondiente relacionada con dicho punto clave;

15

b) para dicha al menos una característica local de punto clave, calcular una probabilidad de relevancia de característica local correspondiente mediante la comparación del valor asumido por dicha característica local con una distribución estadística de referencia correspondiente de los valores de dicha característica local, dicha distribución estadística de referencia es estadísticamente equivalente a una distribución estadística correspondiente generada al recopilar, entre una pluralidad de puntos clave de referencia identificados en una pluralidad de pares de imágenes de referencia, los valores de características locales correspondientes a esos puntos clave de referencia de cada par de imágenes de referencia que se ha determinado que implican una coincidencia de características correcta entre las imágenes de referencia de dicho par;

20

c) calcular una probabilidad de relevancia de puntos clave en función de las probabilidades de relevancia de características locales de dicha al menos una característica local;

25

- una tercera unidad de procesamiento configurada para seleccionar puntos clave, entre los puntos clave del primer grupo, que tiene las mayores probabilidades de relevancia de puntos clave para formar un segundo grupo de puntos clave, y

- una cuarta unidad de procesamiento configurada para explotar los puntos clave del segundo grupo para comparar dicha imagen con dicha al menos otra imagen, donde:

30

- cada distribución estadística de referencia está dispuesta en forma de un histograma correspondiente que tiene una pluralidad de cajas, correspondiendo cada caja a un intervalo predefinido de valores de la característica local correspondiente, y la frecuencia de cada caja corresponde a una relación entre:

35

a) el número de puntos clave de referencia que se ha determinado que implican una coincidencia de características correcta y que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja, y

b) el número total de puntos clave de referencia que tienen un valor de la característica local correspondiente que se encuentra dentro de dicha caja,

40

- la segunda unidad de procesamiento está configurada para calcular la probabilidad de relevancia de característica local para una característica local de un punto clave al:

c) inspeccionar el histograma correspondiente a dicha característica local con el fin de identificar la caja del mismo que se ajusta al valor asumido por la característica local del punto clave, y

d) establecer la probabilidad de relevancia de característica local a la frecuencia de la caja identificada.

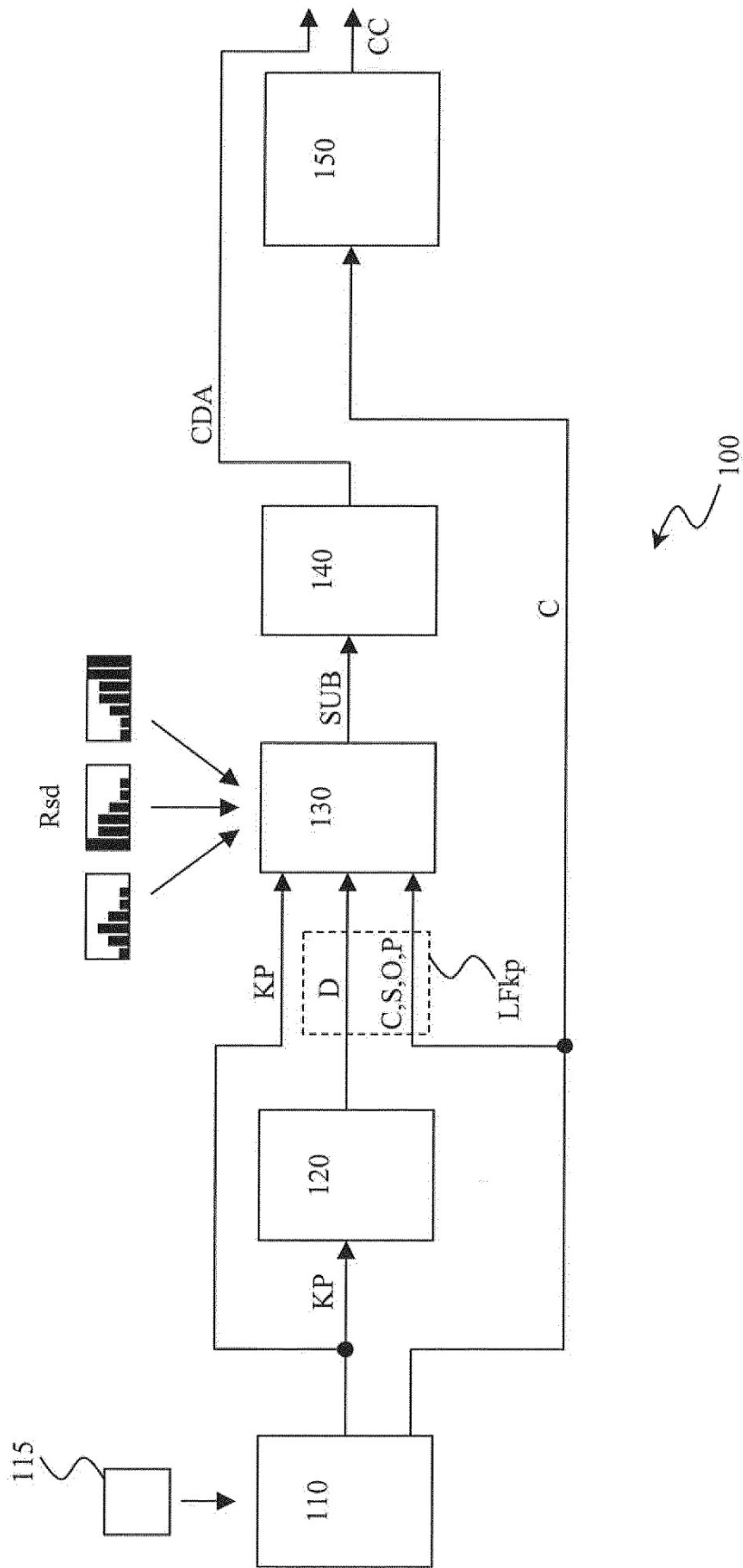


FIG.1

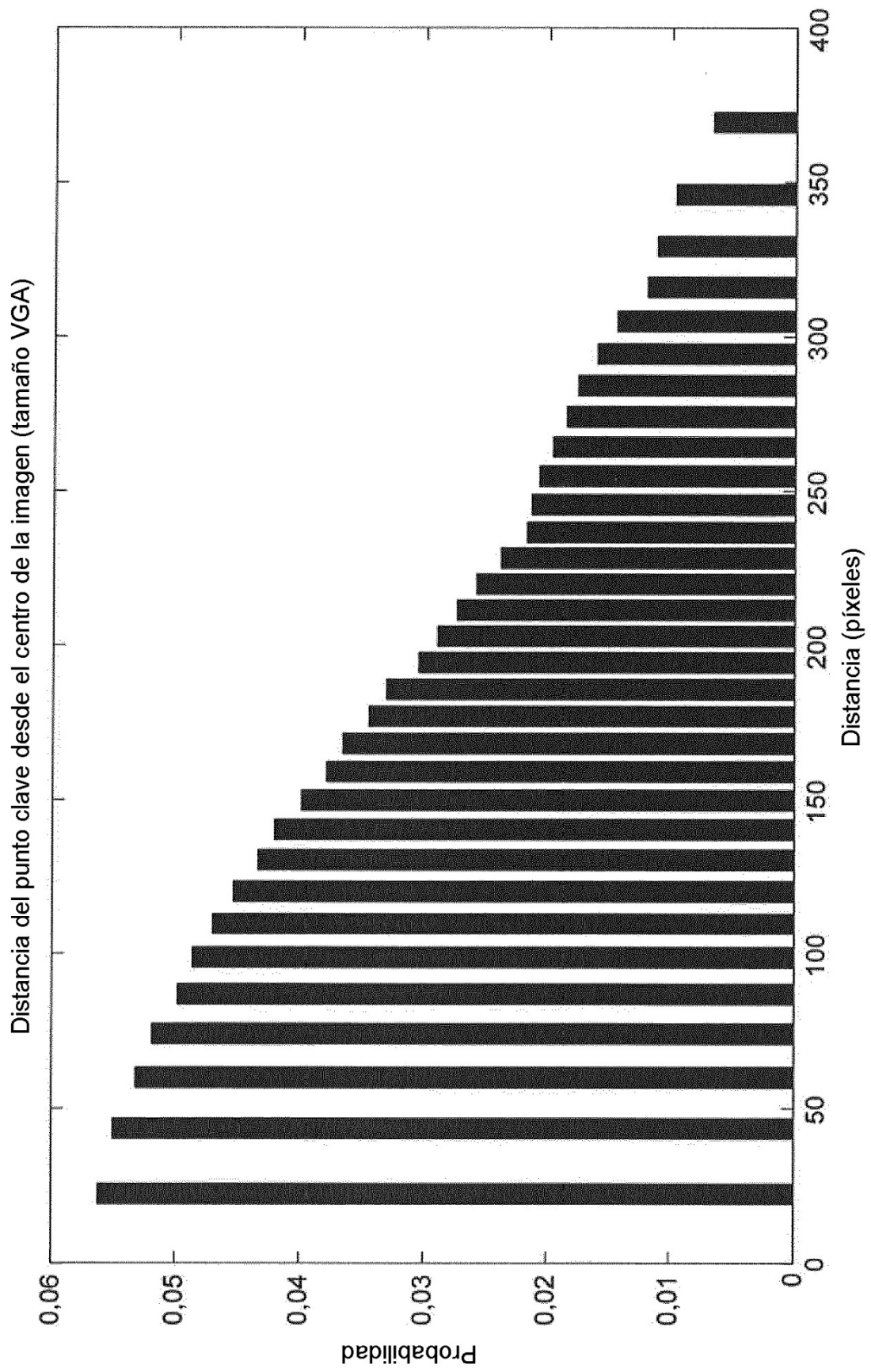


FIG.2A

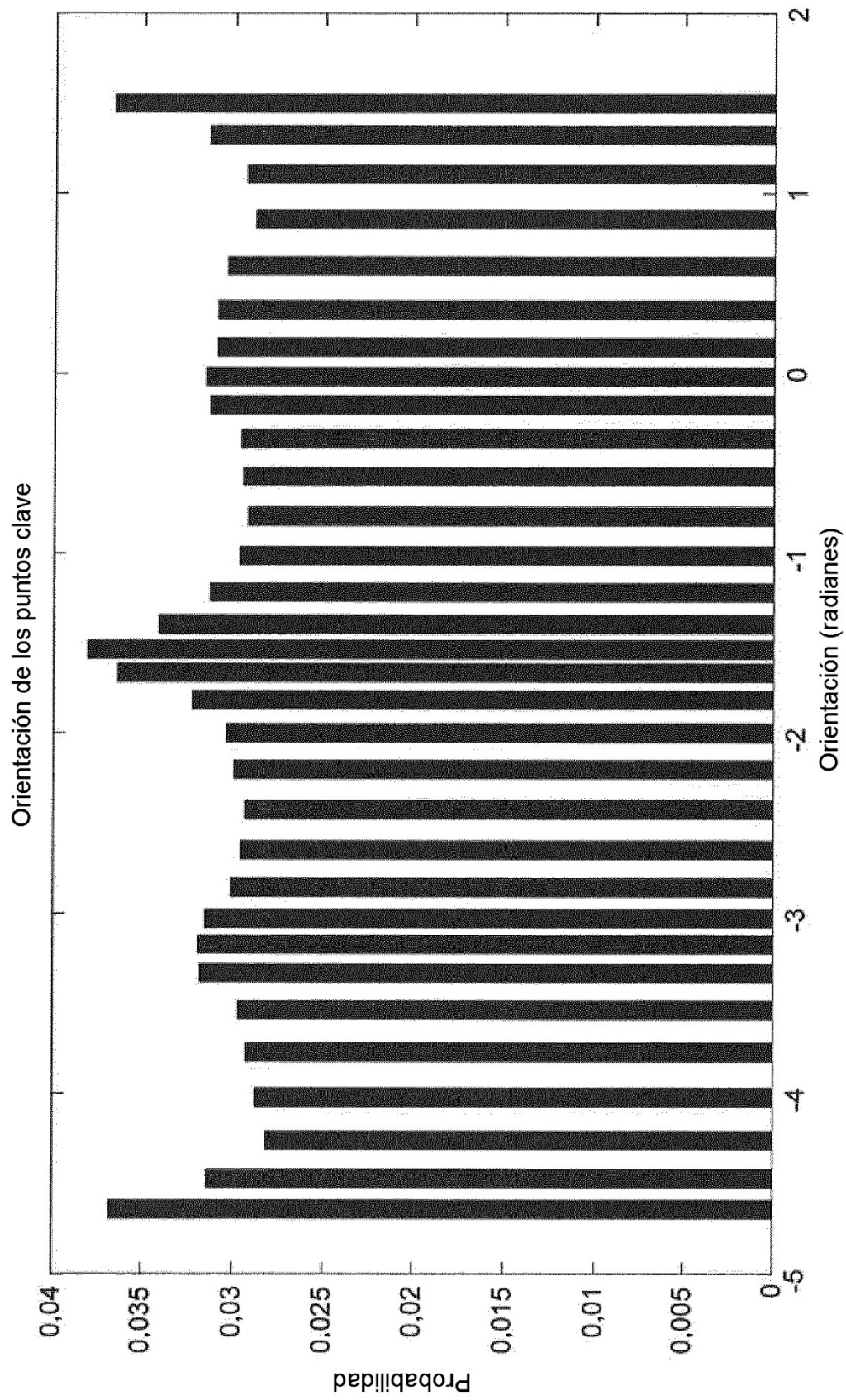


FIG.2B

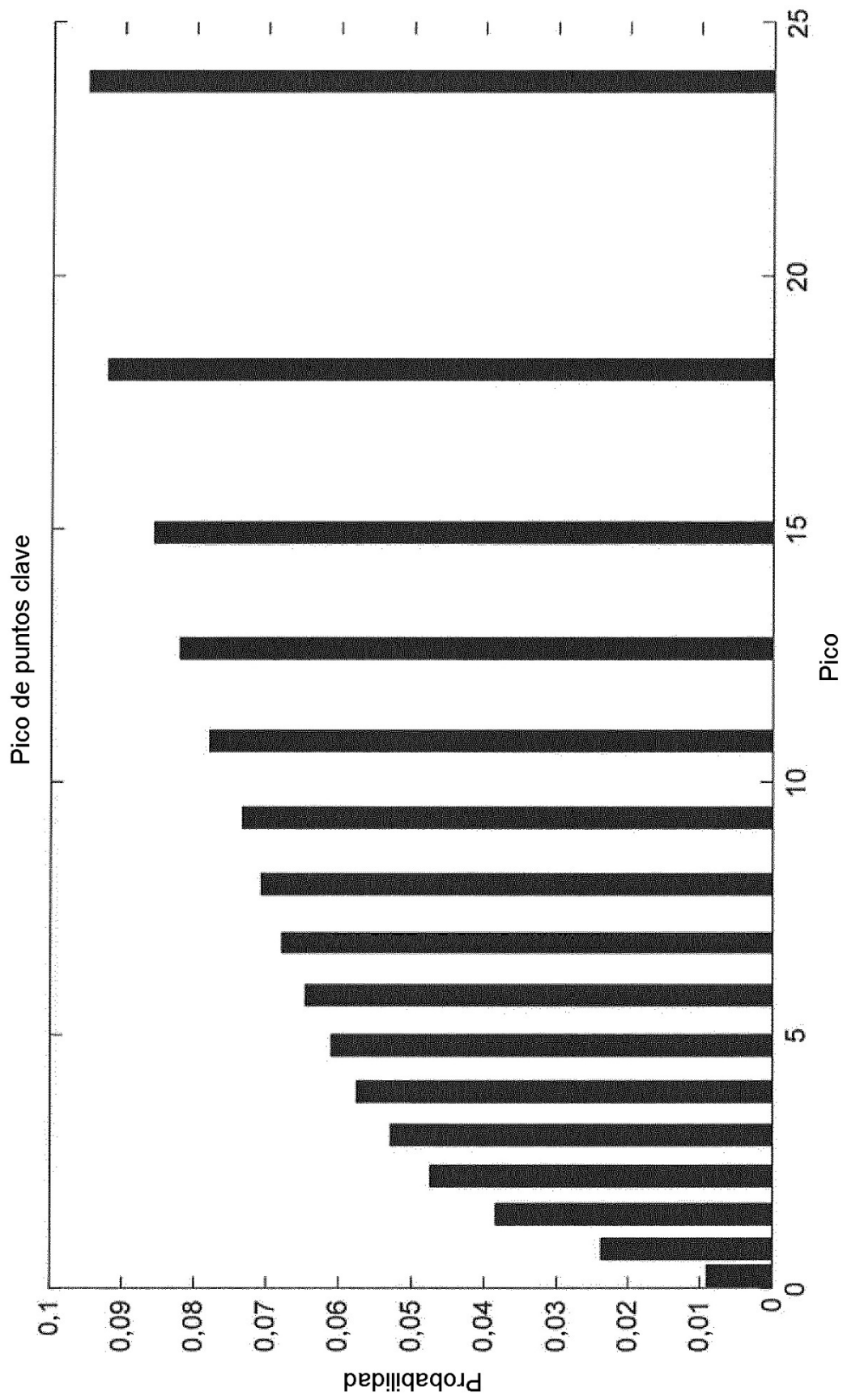


FIG.2C

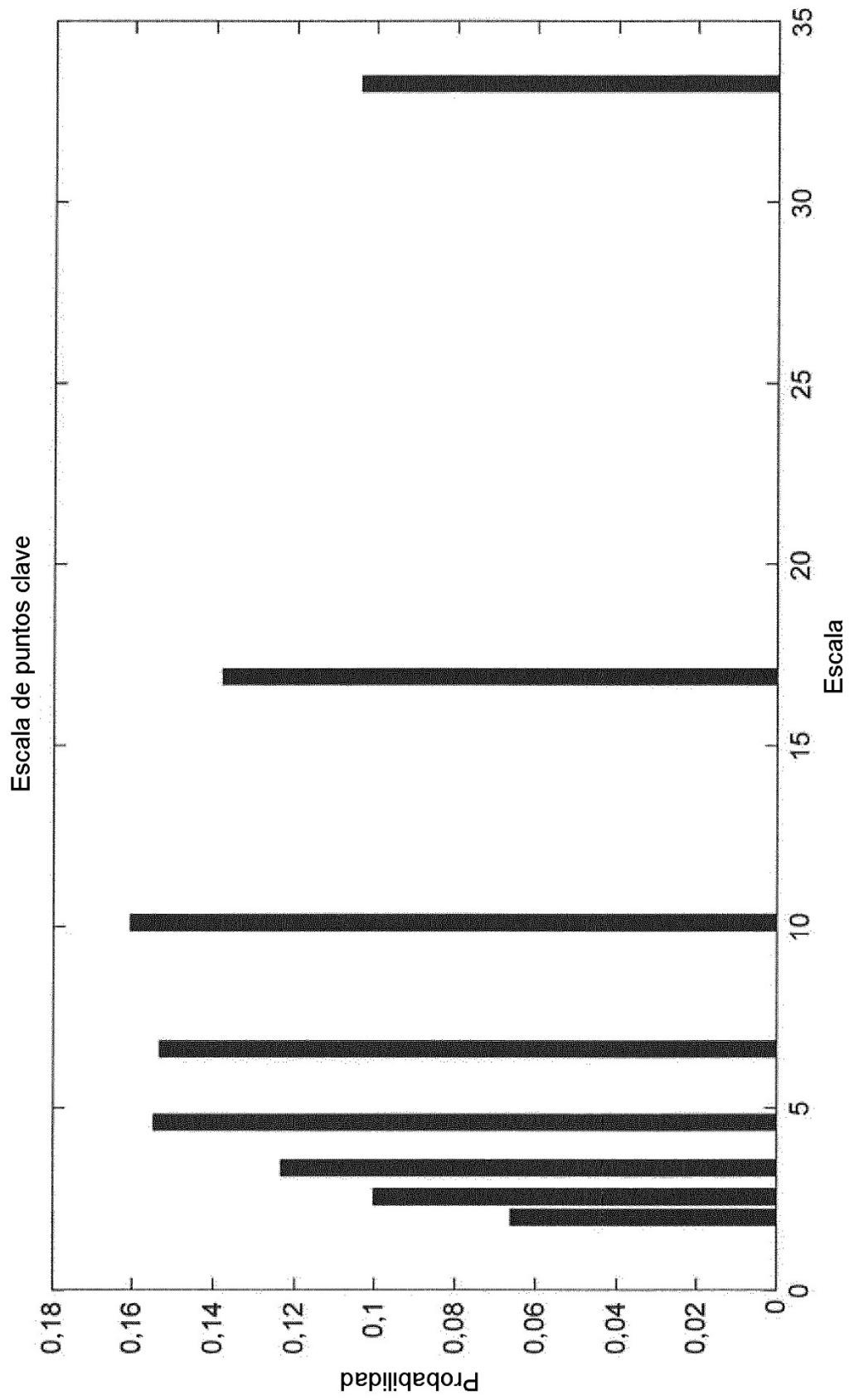


FIG.2D

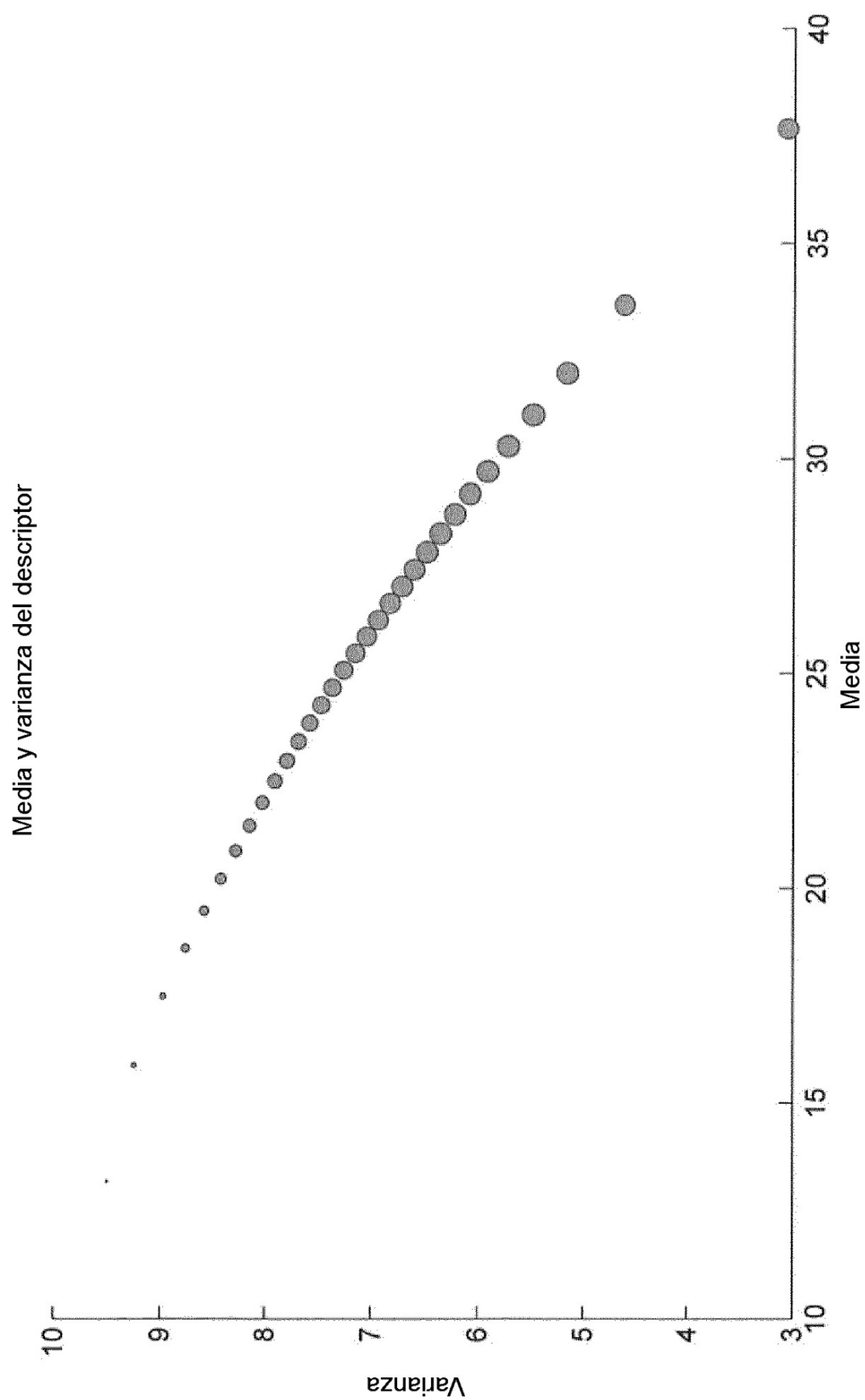


FIG.2E

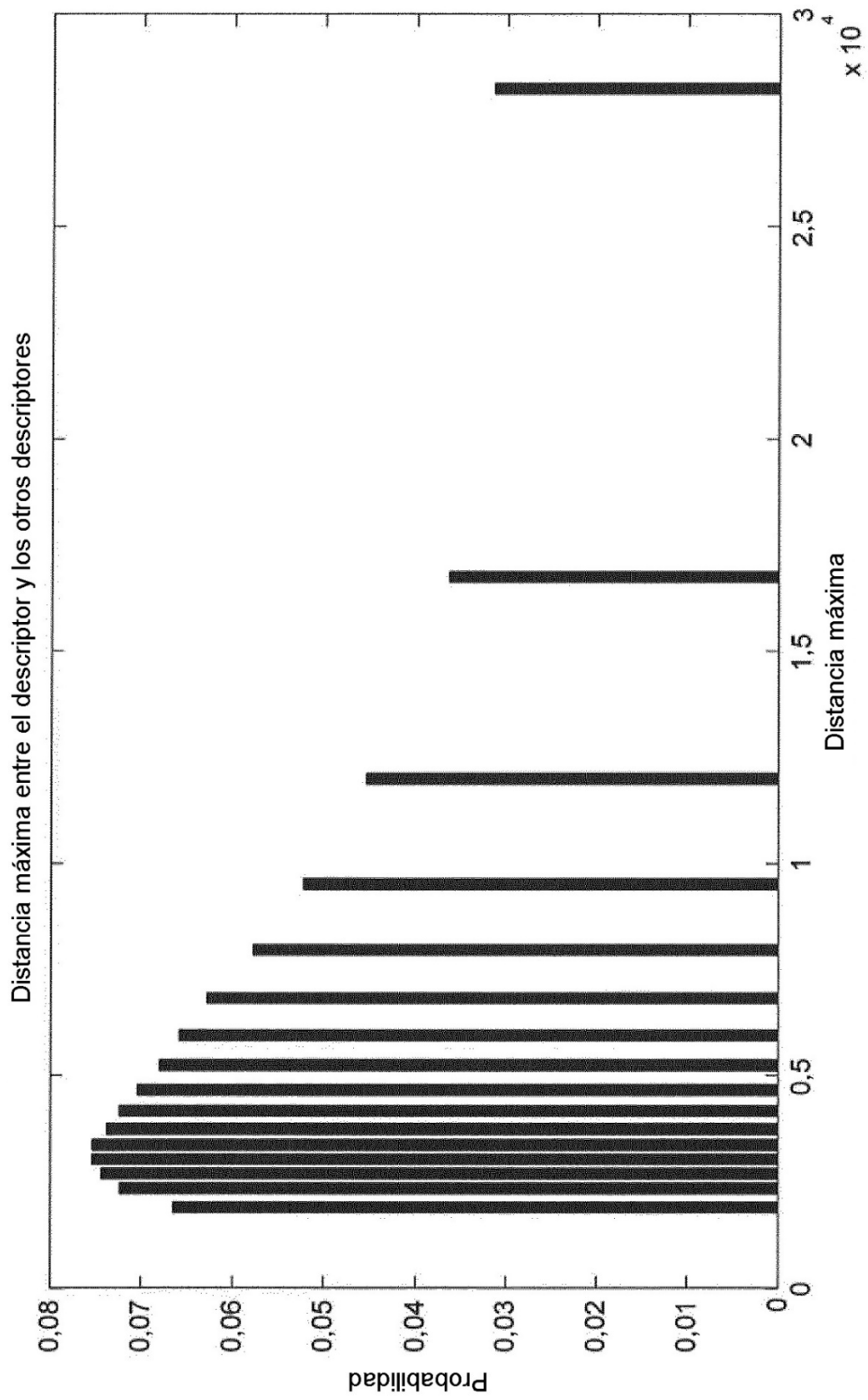


FIG.2F

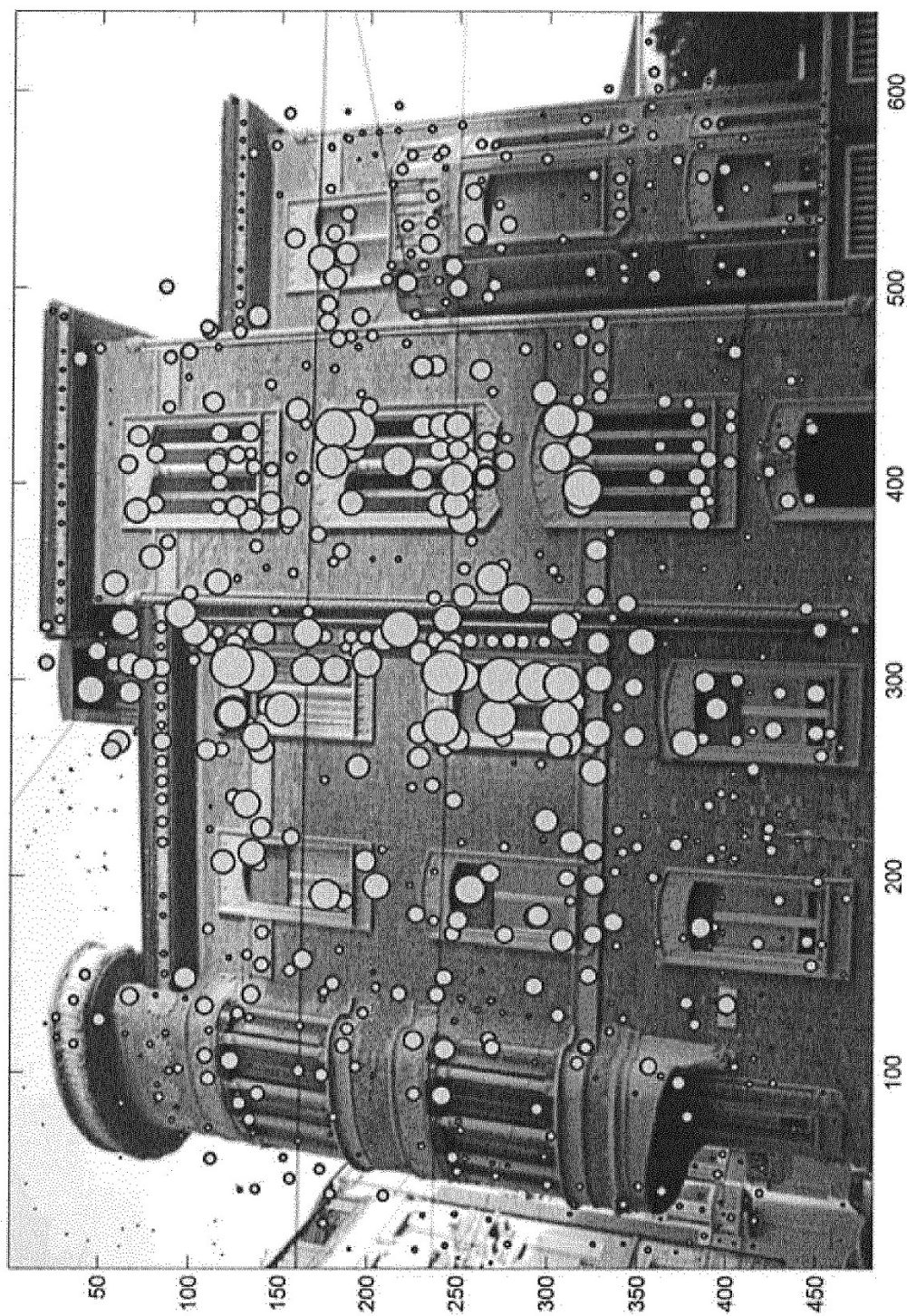
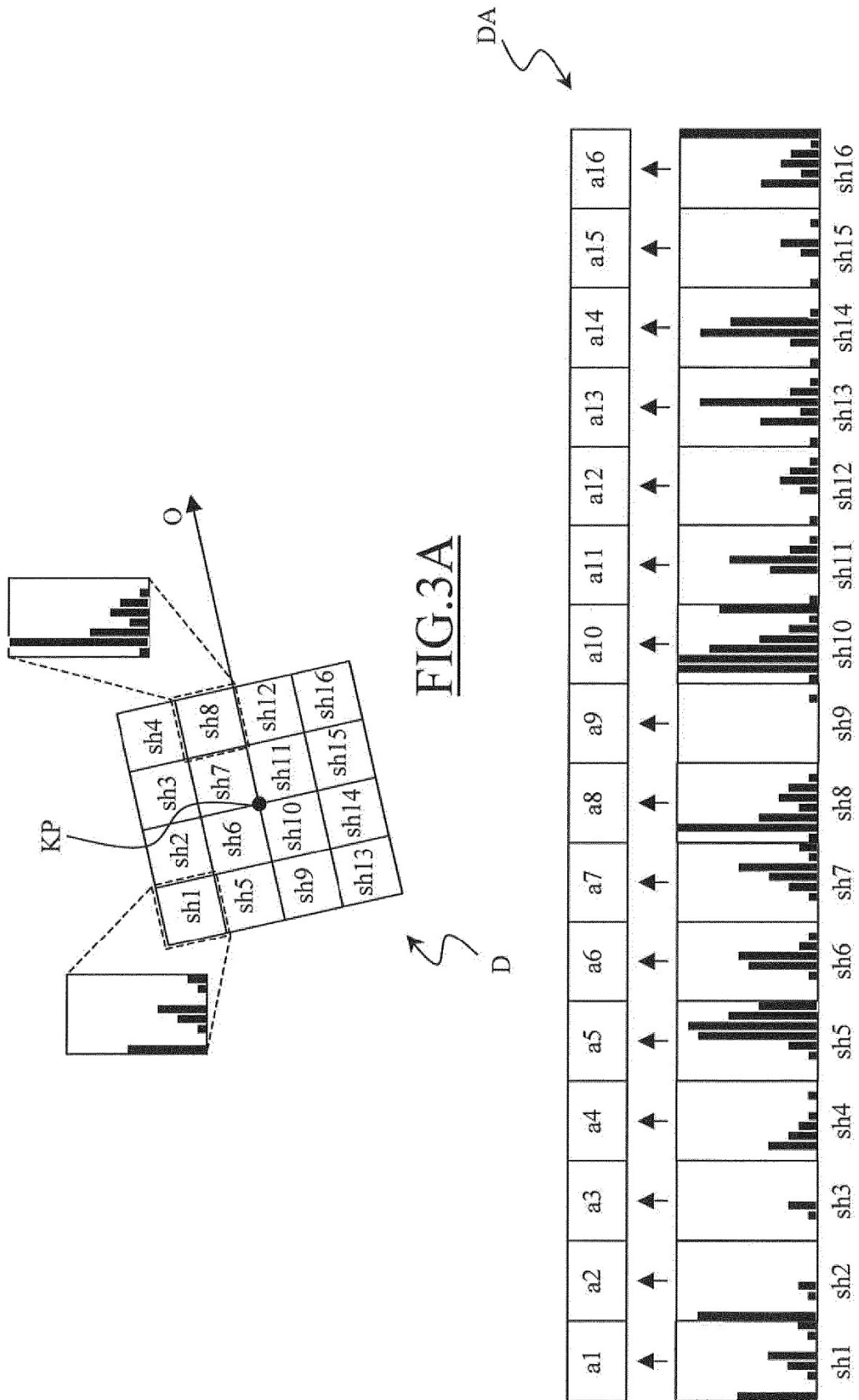


FIG.2G



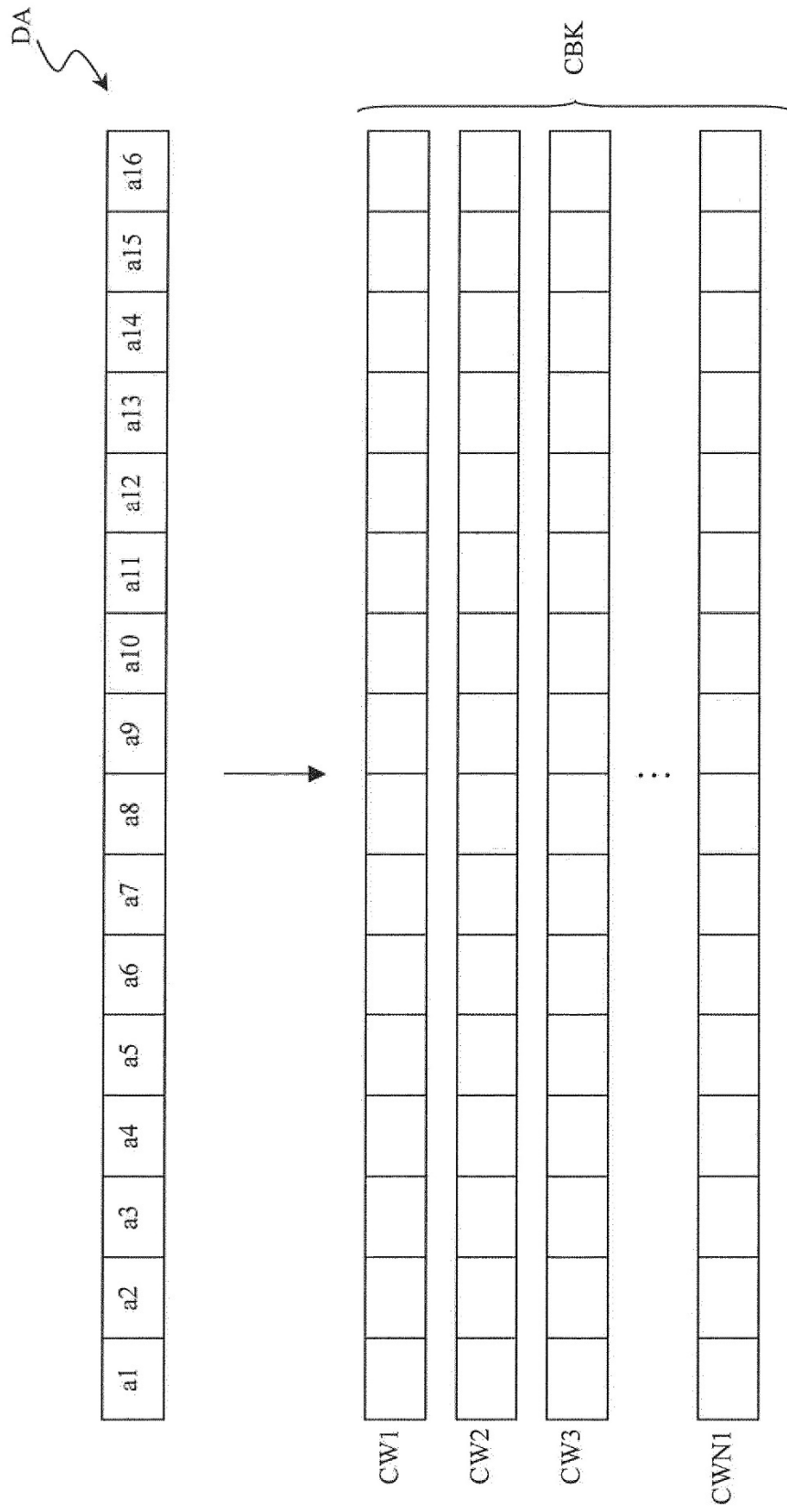


FIG.4A

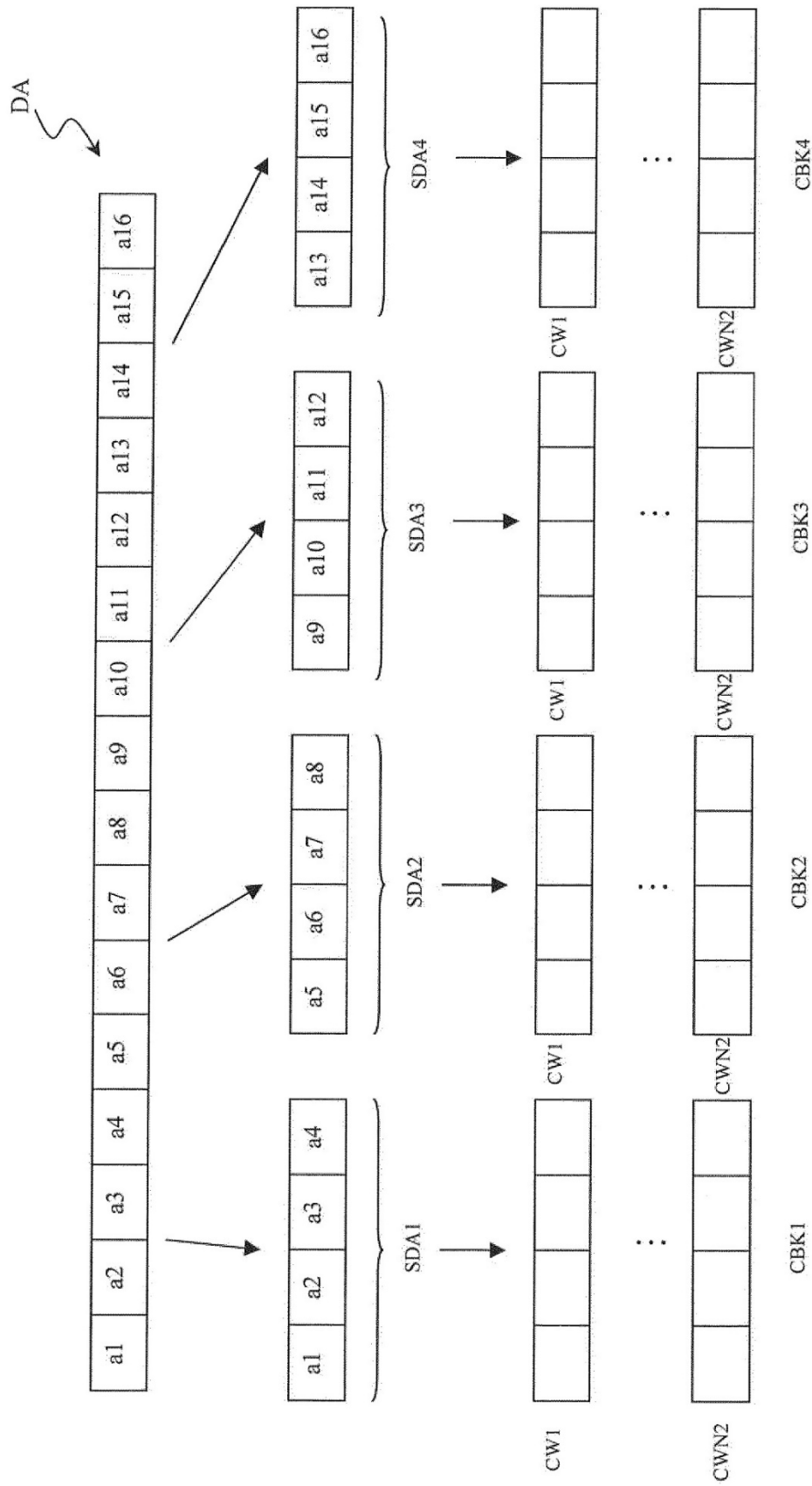


**FIG.4B**

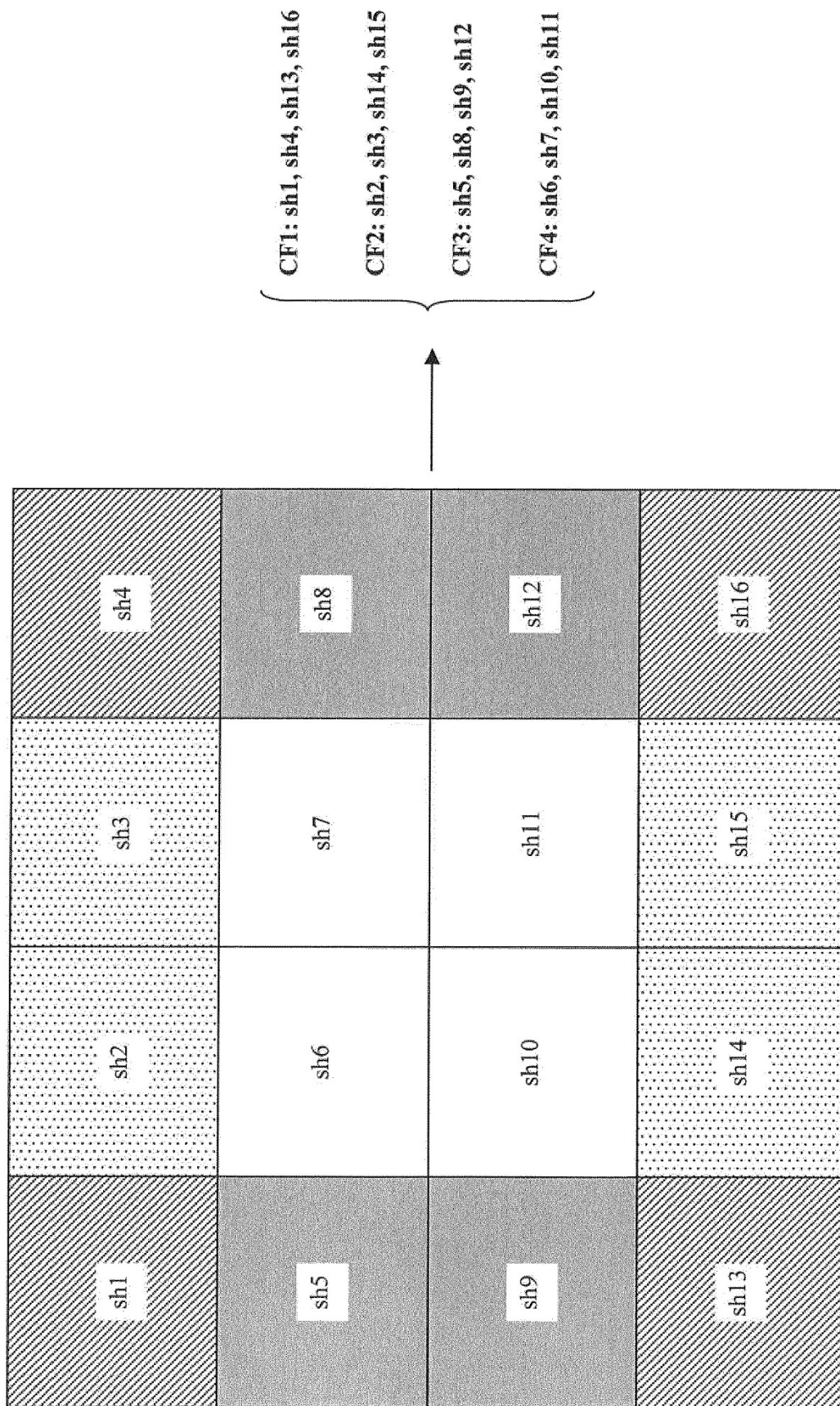


FIG.5

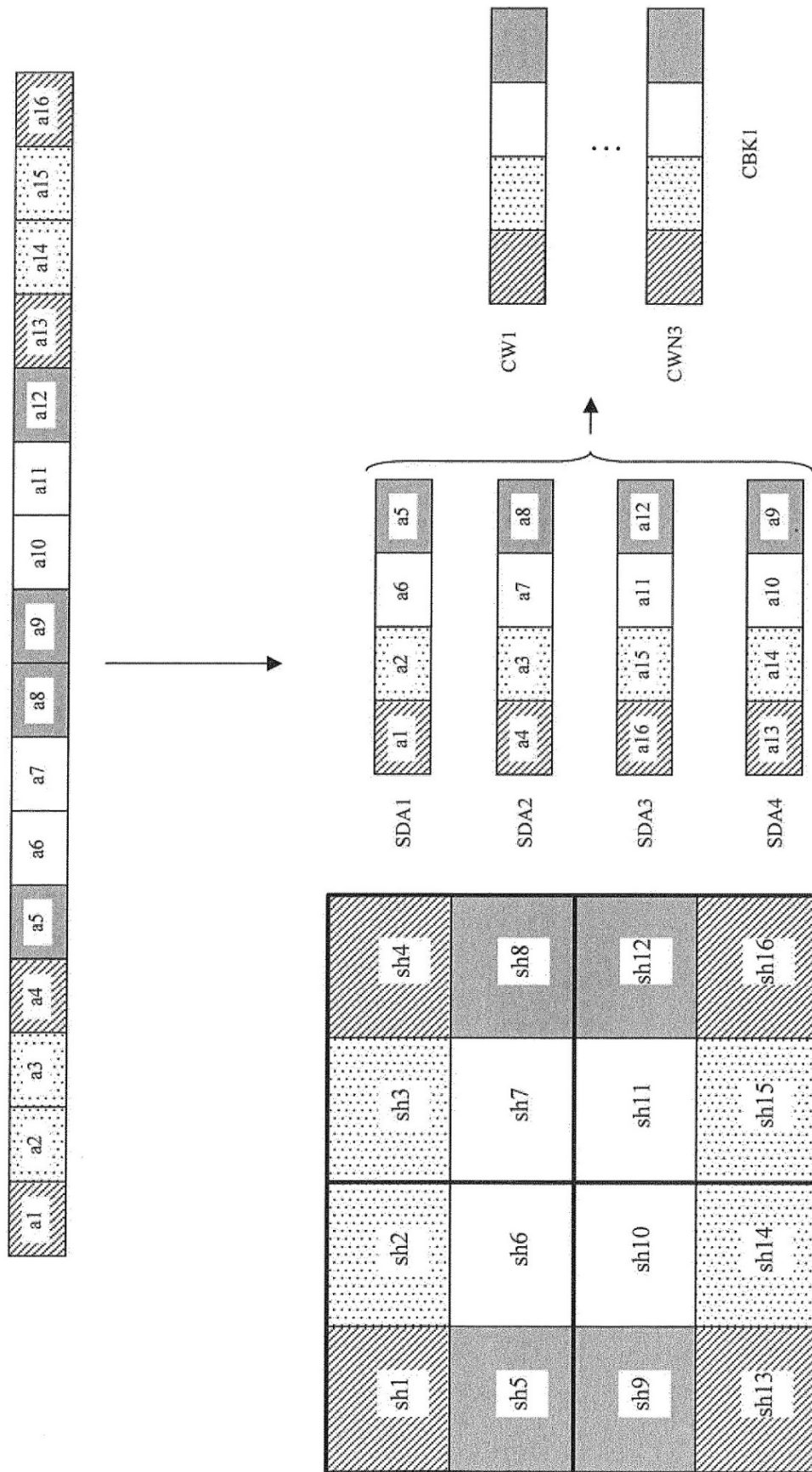


FIG.6A

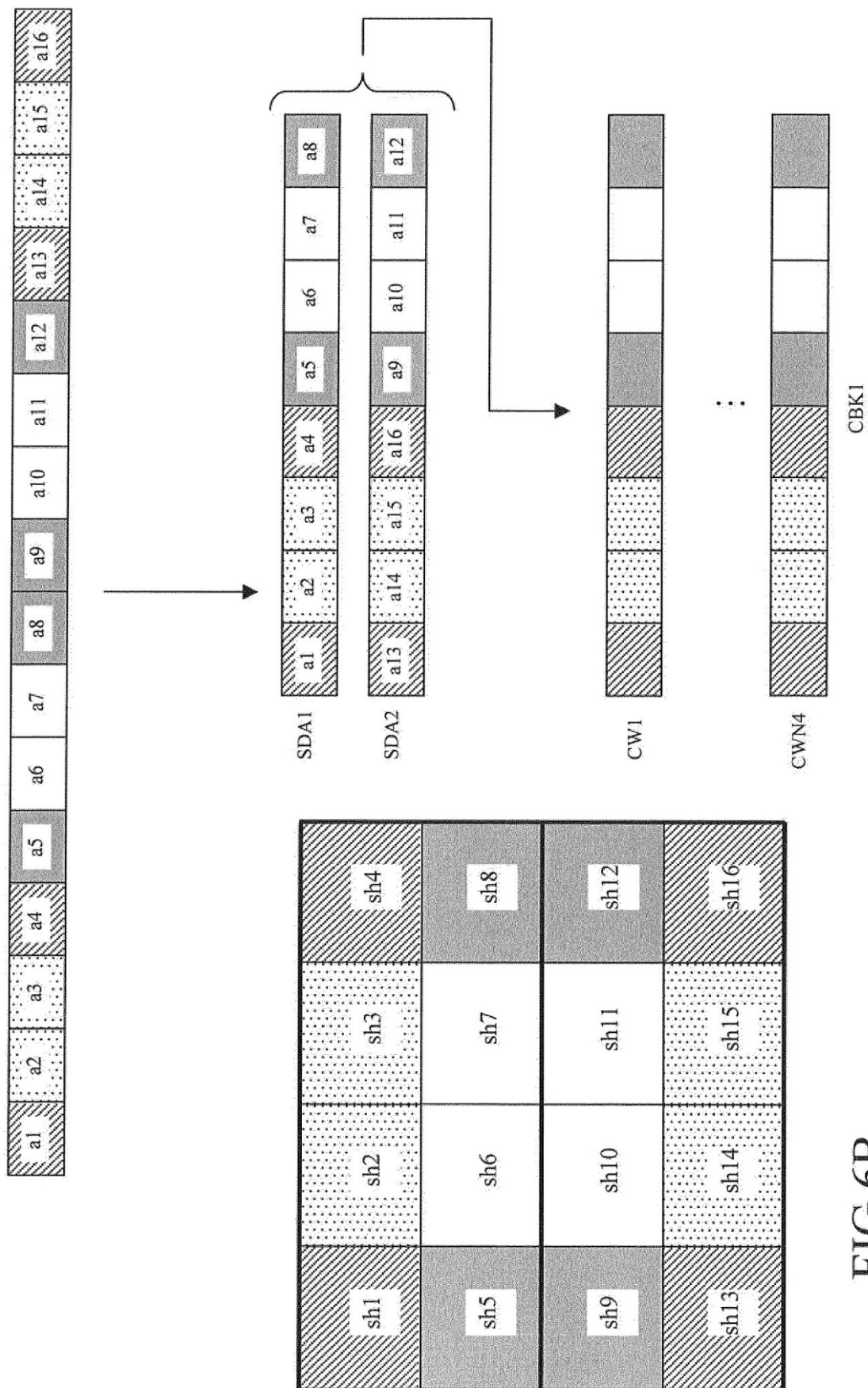


FIG.6B

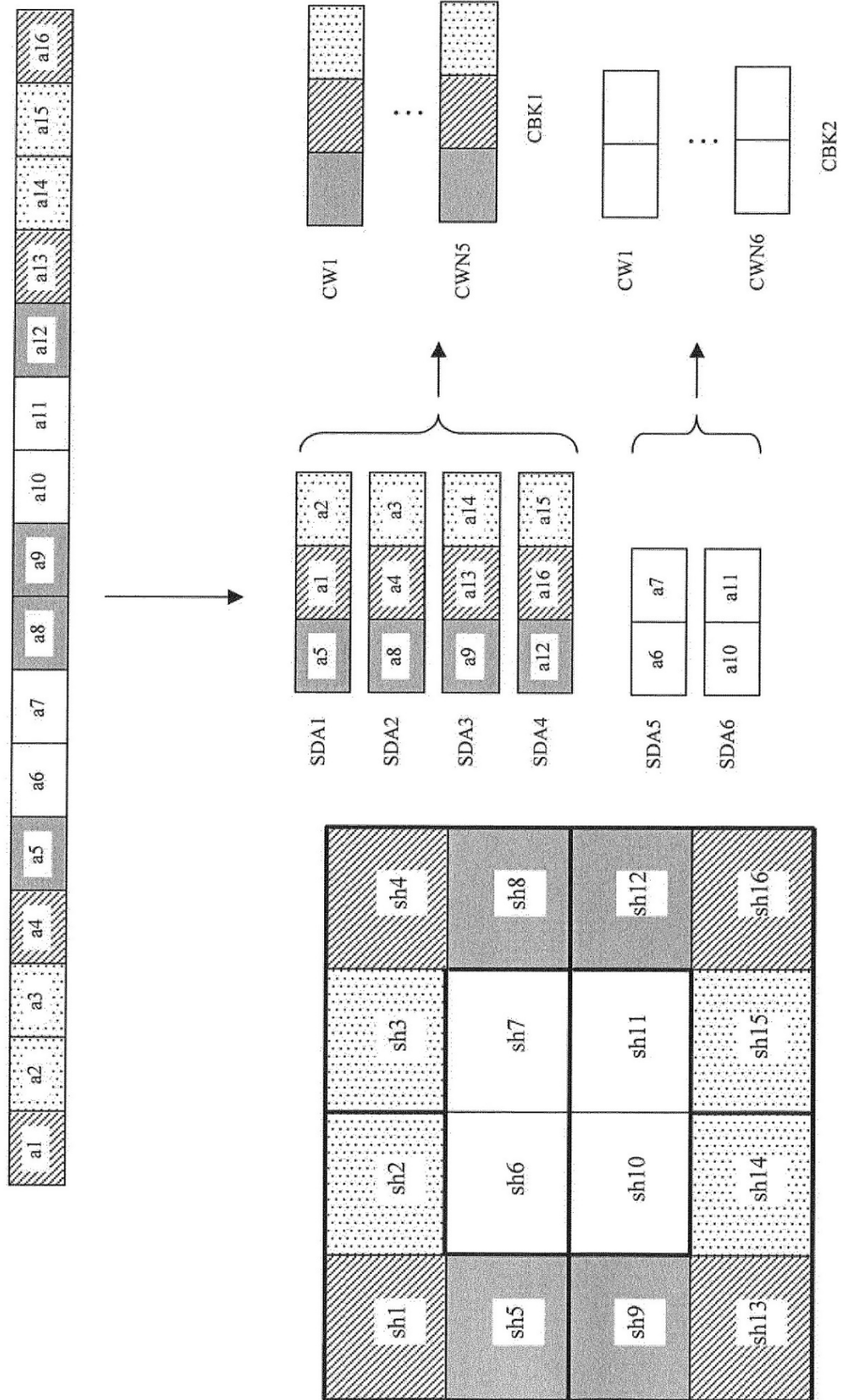


FIG.6C

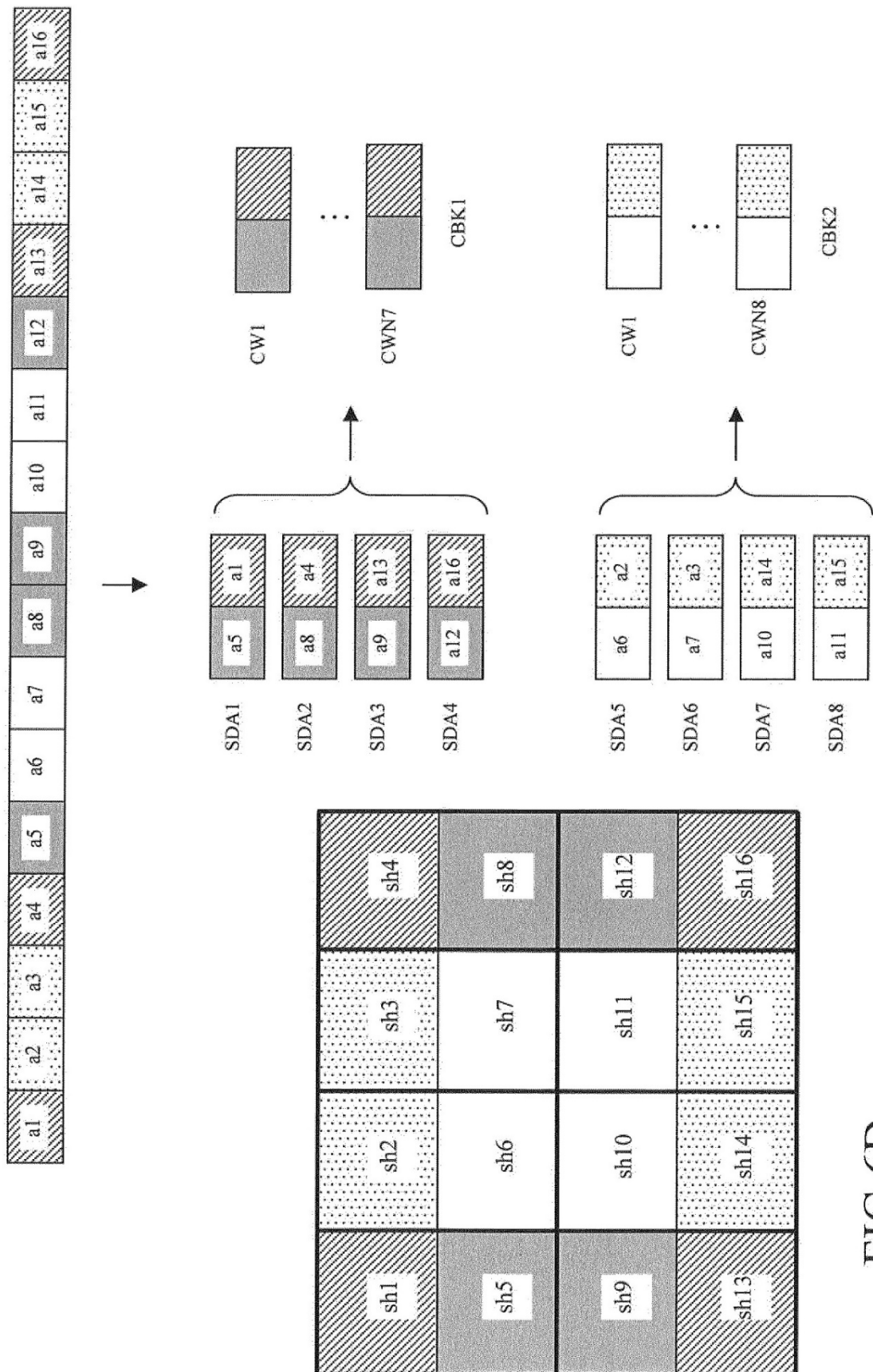


FIG.6D

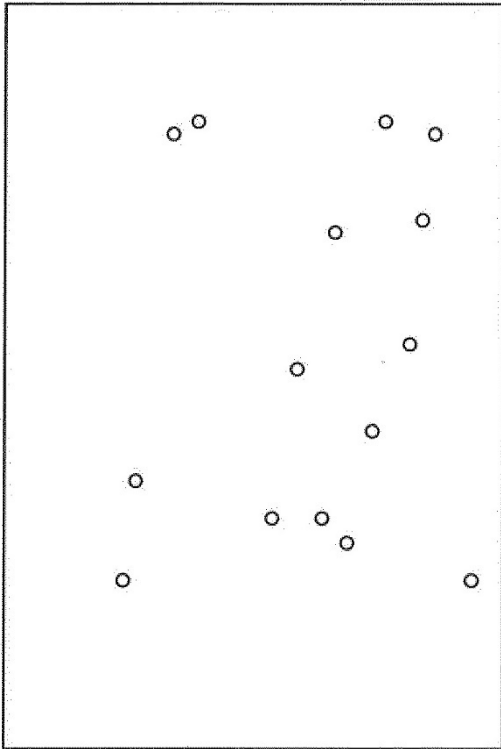


FIG. 7A

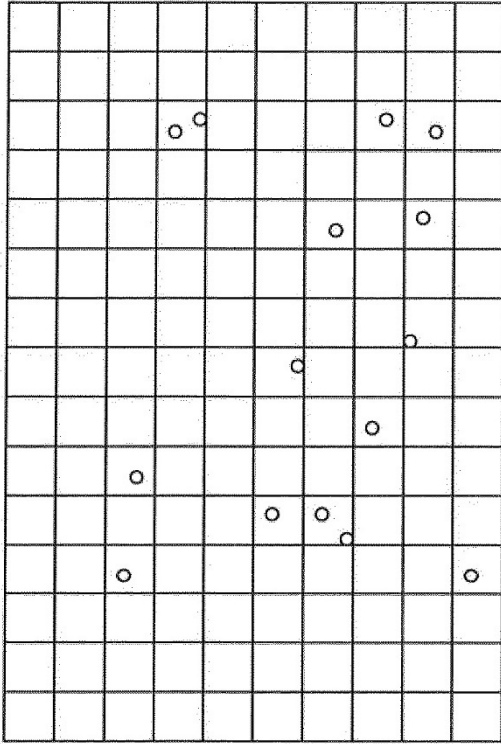


FIG. 7B

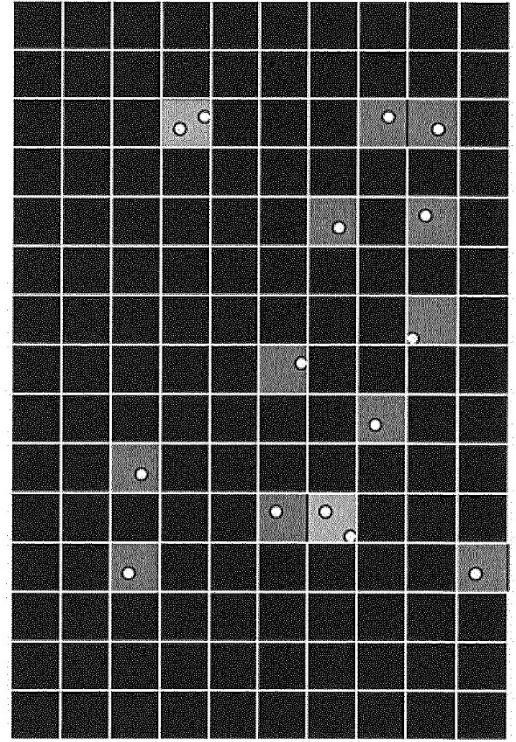


FIG. 7C

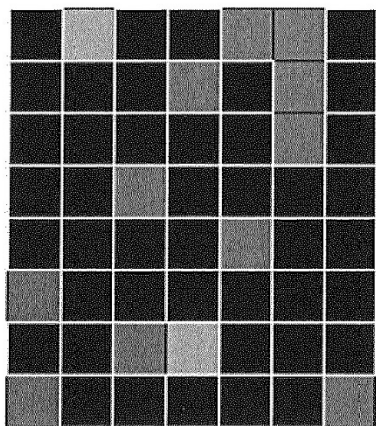


FIG. 7E

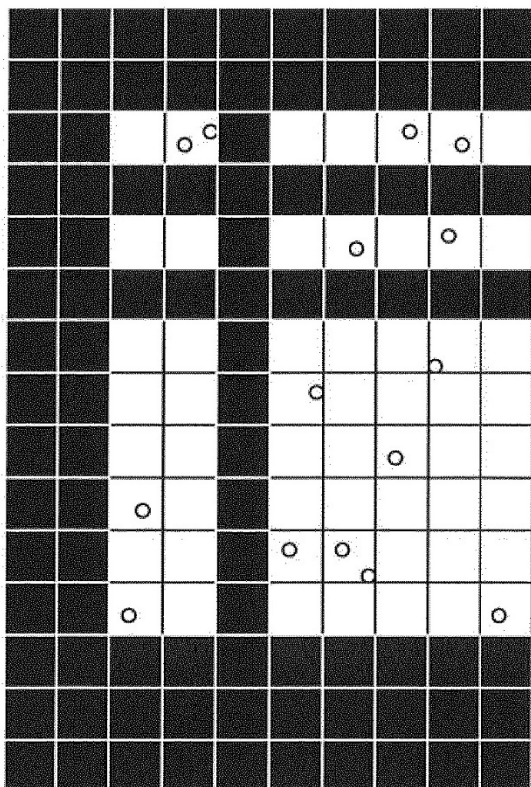


FIG. 7D

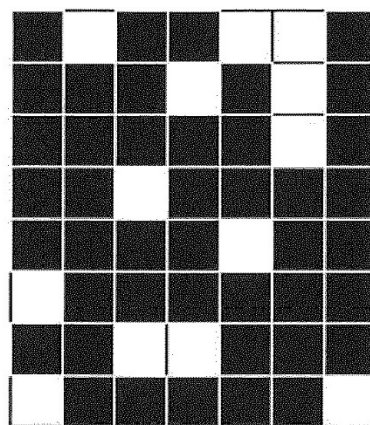


FIG. 7F

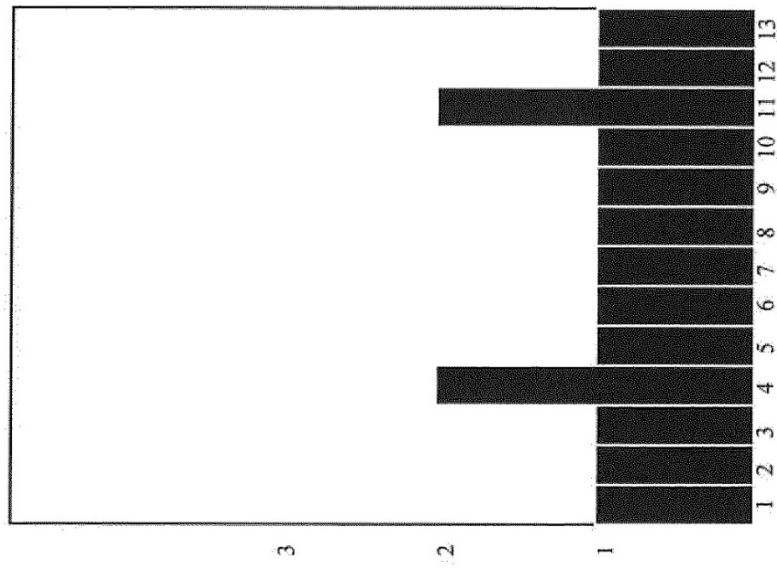


FIG.8B

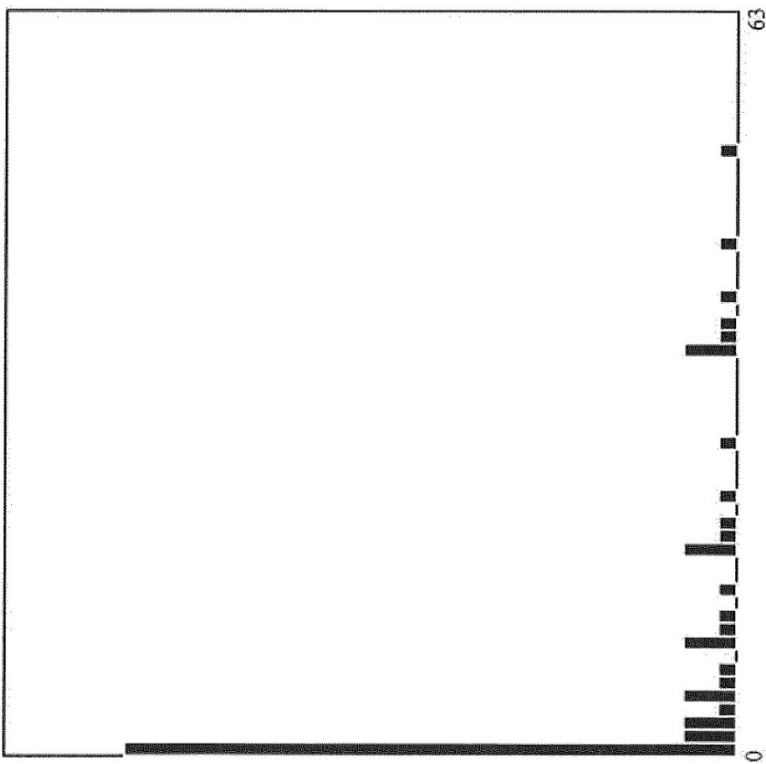


FIG.8A

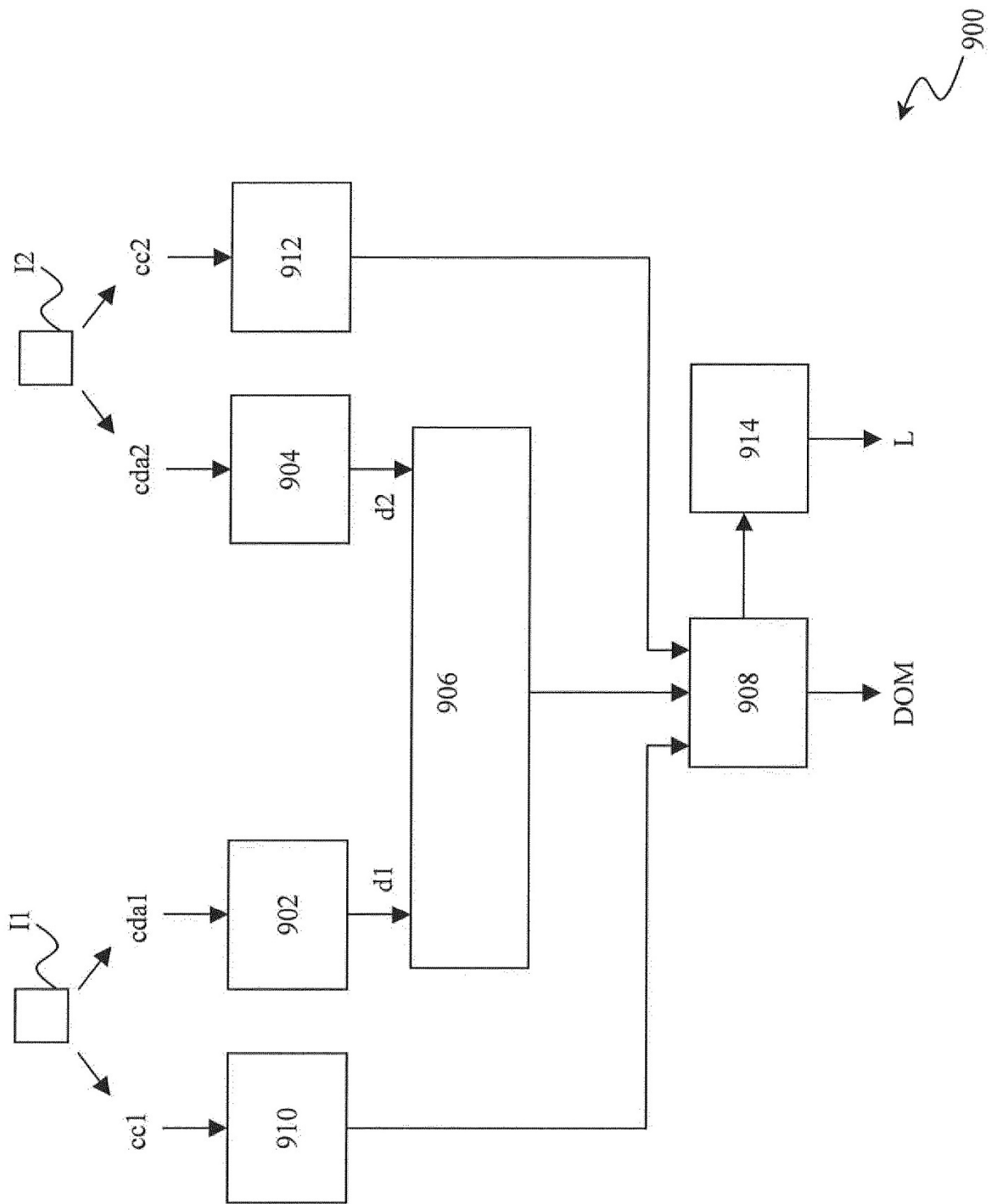


FIG.9

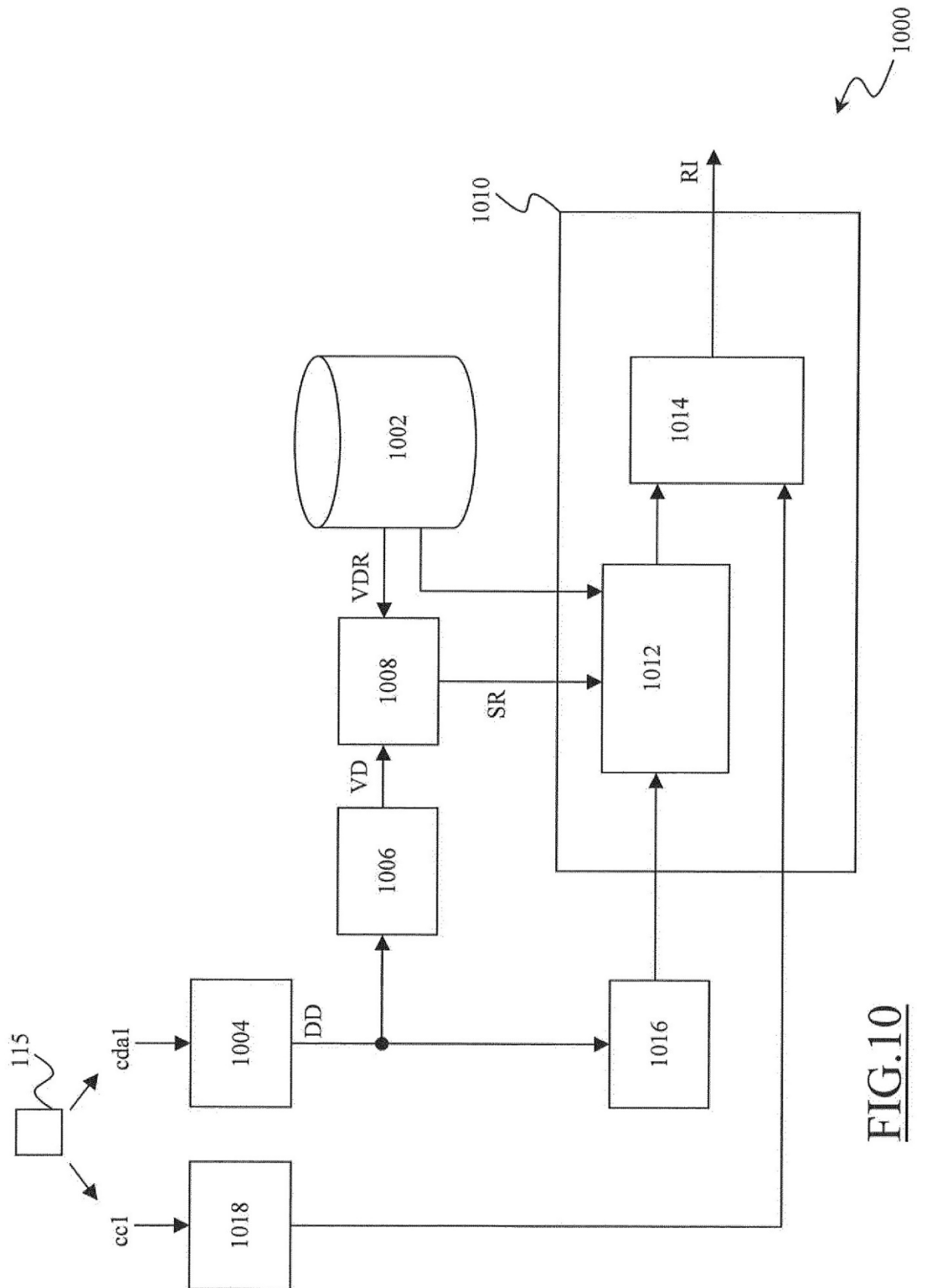


**FIG.10**