

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 791 718**

51 Int. Cl.:

G06K 9/62 (2006.01)

G06K 9/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.01.2006 E 12156622 (8)**

97 Fecha y número de publicación de la concesión europea: **19.02.2020 EP 2487624**

54 Título: **Detección y seguimiento de objetos en imágenes**

30 Prioridad:

07.01.2005 US 641734 P

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

05.11.2020

73 Titular/es:

**QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121, US**

72 Inventor/es:

SHAMAIE, ATID

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 791 718 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

Detección y seguimiento de objetos en imágenes

5 **CAMPO TÉCNICO**

[0001] Esta divulgación se refiere al procesamiento de imágenes y, más en particular, a la detección de un objeto en una imagen.

10 **ANTECEDENTES**

15 [0002] La detección y el reconocimiento de objetos en una imagen típicamente implica técnicas de filtrado de imágenes y reconocimiento de patrones. La detección y reconocimiento de objetos en una imagen que contiene cientos o miles de píxeles puede requerir una gran potencia de procesamiento y puede llevar mucho tiempo. Por lo tanto, puede ser útil reducir la dimensionalidad de una imagen antes de procesar la imagen para detectar y reconocer objetos en la imagen. Una técnica conocida para reducir la dimensionalidad de los datos es el análisis de componentes principales (PCA). El PCA se describe en, por ejemplo, Joliffe I.T., Principal Component Analysis, Springer-Verlag, Nueva York (1986).

20 [0003] El PCA usa los valores propios y los vectores propios de la matriz de covarianza de un conjunto de datos como representativos de características valiosas del conjunto de datos, reduciendo de este modo la dimensionalidad del conjunto de datos.

25 [0004] Existen muchos procedimientos y mecanismos que permiten a los humanos interactuar con los ordenadores. Las tecnologías de visión por ordenador pueden permitir que un ordenador detecte un objeto dentro de una imagen capturada por una cámara. Un ordenador que es capaz de detectar y reconocer un objeto dentro de una imagen puede proporcionar al usuario la capacidad de interactuar con el ordenador a través del uso de gestos con las manos.

30 [0005] Se puede mostrar una interfaz de usuario por ordenador en una superficie o pantalla. Una o más cámaras pueden monitorear la actividad en la proximidad de la superficie o pantalla y capturar imágenes de la actividad en la proximidad de la pantalla. A continuación, el ordenador puede procesar estas imágenes, detectar uno o más objetos dentro de las imágenes y percibir que un usuario está usando gestos con las manos para interactuar con la interfaz de usuario por ordenador que se muestra en la superficie o pantalla.

35 [0006] Algunos sistemas pretenden percibir que un usuario está usando gestos con las manos para interactuar con la interfaz de usuario por ordenador que se muestra en la superficie o en la pantalla. Algunos de estos sistemas simplemente perciben el objeto más brillante en una imagen y clasifican ese objeto como una mano o un dedo. En consecuencia, estos sistemas pueden percibir un objeto como una mano o un dedo aunque el objeto no sea ni una mano ni un dedo.

40 [0007] En el documento US 2001/012001 A1, se divulga un panel LED infrarrojo para aplicar siempre luz infrarroja a la superficie posterior de una pantalla semitransparente, una cámara CCD para capturar solo luz infrarroja proveniente de la pantalla semitransparente y un proyector para proyectar una imagen en la pantalla semitransparente con luz que no incluye luz infrarroja. Un dispositivo de control capta una variación en la cantidad de luz de reflexión como información de imagen de detección basada en una señal de formación de imágenes de la cámara CCD. De acuerdo con la información de manipulación obtenida en base a la información de la imagen de detección, el dispositivo de control ejecuta los controles necesarios para cambiar la imagen de una interfaz que se desplazará en la pantalla semitransparente. Un procedimiento de manipulación puede tener una forma que cause una variación en la información de la imagen de detección a través del reflejo de la luz interactiva.

45 [0008] En "Eigentracking: robust matching and tracking of articulated objects using a view-based representation" de Black y col., en COMPUTER VISION ECCV '96, Springer Berlín Heidelberg, se divulga un enfoque para rastrear objetos rígidos y articulados usando una representación basada en la vista, en el desarrollo y extensión del trabajo en representaciones de espacio propio, técnicas de estimación robustas y estimación de flujo óptico parametrizado.

50 [0009] En "An appearance-based approach to gesture-recognition" de Jerome y col., en Field Programmable Logic and Application, Springer Berlín Heidelberg, se divulgan técnicas para el diseño de un sistema capaz de interactuar con el usuario mediante el reconocimiento visual de gestos con las manos, compuesto por tres módulos que incluyen rastreo, clasificación de postura y reconocimiento de gestos. Se presenta un proceso de clasificación para reconocer la postura de la mano usando la distancia en un espacio propio.

55 [0010] En "Applications of Support Vector Machines for Pattern Recognition: A Survey" de Hyeran Byun y col., en SVM 2002 pp. 213-236, se presenta un estudio exhaustivo sobre aplicaciones de máquinas de vectores de soporte para reconocimiento de patrones.

5 [0011] En "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", Athinodoros S. Georghiades y col., IEEE TPAMI vol. 23, núm. 6, 2001, así como en "Support Vector Machines for Object Recognition under Varying Illumination Conditions", Takahiro Okabe y col., 6.^a Conferencia asiática sobre visión por ordenador, 2004, se divulga cómo se puede realizar el reconocimiento de objetos en condiciones de iluminación variables usando un modelo de cono de iluminación.

BREVE EXPLICACIÓN

10 [0012] De acuerdo con un primer aspecto de la invención, se proporciona un procedimiento implementado por ordenador de acuerdo con la reivindicación 1.

15 [0013] De acuerdo con un segundo aspecto de la invención, se proporciona un aparato de acuerdo con la reivindicación 9.

20 [0014] Para identificar con mayor precisión una mano o un dedo en una imagen y mejorar de este modo la eficacia y la eficiencia de la interacción humana con el ordenador, varios sistemas divulgados (1) reconocen un objeto que se toca o se mueve cerca de la superficie o pantalla como una mano o un dedo y (2) excluyen objetos que no se reconocen como una mano o un dedo.

25 [0015] Se presentan procedimientos y aparatos para la detección y reconocimiento de dedos que tocan o se mueven cerca de una superficie o pantalla. Las imágenes se procesan para detectar la presencia de un dedo y rastrear la posición del dedo detectado. La posición de un dedo puede ser usada por un ordenador, por ejemplo, para controlar el puntero del ratón del ordenador. Los procedimientos presentados pueden ser útiles para permitir que los humanos interactúen con los ordenadores de una manera más natural que a través del uso de un ratón y/o teclado convencional.

30 [0016] En una implementación, se crea un modelo de uno o más dedos situados en una o más orientaciones. En particular, una cámara captura un conjunto de imágenes de entrenamiento de uno o más dedos. Aplicando la técnica de PCA, se crea un modelo del conjunto de imágenes de entrenamiento. Se determina la matriz de covarianza del conjunto de imágenes de entrenamiento y se selecciona un número selecto de vectores propios de la matriz de covarianza para definir un espacio propio para el conjunto de imágenes de entrenamiento. Cada una de las imágenes del conjunto de imágenes de entrenamiento se proyecta en el espacio propio definido por los vectores propios seleccionados de la matriz de covarianza. En el espacio propio, cada imagen de entrenamiento está representada por un único punto de entrenamiento. Por lo tanto, la proyección de cada imagen de entrenamiento en el espacio propio crea una nube de puntos de entrenamiento en el espacio propio. La nube de puntos de entrenamiento se modela a continuación mediante un modelo geométrico (por ejemplo, superficie o línea) en el espacio propio.

40 [0017] El modelo se usa a continuación para detectar, reconocer y rastrear los dedos dentro de una imagen. En un sistema, una cámara y un iluminador infrarrojo (IR) se colocan detrás de la superficie o pantalla. Un dedo que toca o se mueve cerca de la superficie o la pantalla refleja parte de la radiación IR proyectada por el iluminador IR. Parte de la radiación IR reflejada es capturada en una imagen por la cámara. La imagen se procesa, como se explica a continuación, para detectar y reconocer la presencia y posición del dedo dentro de la imagen.

45 [0018] El procesamiento incluye extraer de la imagen uno o más objetos que potencialmente pueden ser dedos. Cada objeto extraído se proyecta en el espacio propio que se definió para el conjunto de imágenes de entrenamiento. En el espacio propio, la proyección del objeto extraído está representada por un único punto de imagen. Las coordenadas del punto de imagen único se comparan con el modelo de la nube de puntos de entrenamiento para determinar si el punto de imagen único coincide o casi coincide con el modelo, se determina que el objeto es un dedo. Si se determina que el objeto es un dedo, la posición del dedo con respecto a la pantalla o superficie (por ejemplo, una coordenada (x, y)) se calcula y se ingresa, por ejemplo, en el ordenador y/o la aplicación de software.

55 [0019] De acuerdo con un aspecto general, se determinan las coordenadas para un punto de imagen en un espacio multidimensional, caracterizando el punto de imagen un objeto particular. Se proporciona una ecuación que describe un modelo en el espacio multidimensional, siendo el modelo característico de un conjunto de imágenes de entrenamiento de uno o más de otros objetos. Las coordenadas se aplican a la ecuación para determinar una distancia entre el punto de imagen y el modelo, y se determina si el objeto particular coincide con uno o más de otros objetos en base a la distancia determinada.

60 [0020] Las implementaciones del aspecto general anterior pueden incluir una o más de las características siguientes. Por ejemplo, se puede recibir una imagen del objeto particular. La imagen puede tener un número de elementos de datos, y el espacio multidimensional puede tener una dimensionalidad menor que el número de elementos de datos. La determinación de las coordenadas para el punto de imagen puede incluir proyectar la

imagen en el espacio multidimensional para producir las coordenadas para el punto de imagen en el espacio multidimensional.

5 **[0021]** El modelo puede ser un modelo de un conjunto de puntos de entrenamiento en el espacio multidimensional, correspondiendo cada uno de los puntos de entrenamiento en el conjunto a una o más imágenes en la secuencia de imágenes de entrenamiento.

10 **[0022]** La imagen del objeto particular puede normalizarse antes de ser recibida. La normalización de la imagen del objeto particular puede explicar las variaciones en el brillo. La normalización de la imagen del objeto puede incluir la aplicación de una técnica de equalización de histograma a la imagen del objeto particular.

15 **[0023]** La ecuación puede describir un hiperboloide, un cono, una línea o una combinación de un cono y una línea. El objeto particular puede incluir un objeto del conjunto que consiste en un dedo, un bolígrafo y objetos sustancialmente cilíndricos.

20 **[0024]** Proporcionar la ecuación puede incluir seleccionar la ecuación de entre una primera ecuación y una segunda ecuación, describiendo la primera ecuación un primer modelo característico de un primer conjunto de imágenes de entrenamiento, y describiendo la segunda ecuación un segundo modelo característico de un segundo conjunto de imágenes de entrenamiento.

25 **[0025]** De acuerdo con otro aspecto general, se recibe una imagen de un objeto particular, teniendo la imagen un número de puntos de datos. La imagen se proyecta en un espacio multidimensional que tiene una dimensionalidad que es menor que el número de puntos de datos para producir coordenadas para un punto de imagen en el espacio multidimensional, donde el punto de imagen caracteriza el objeto particular. Se proporciona una ecuación que describe un modelo en el espacio multidimensional, siendo el modelo un modelo para un conjunto de puntos de entrenamiento en el espacio multidimensional, y correspondiendo cada uno de los puntos de entrenamiento en el conjunto a una o más imágenes en un conjunto de entrenamiento de imágenes de uno o más de otros objetos. Las coordenadas se aplican a la ecuación para determinar una distancia entre el punto de imagen y el modelo, y se determina si el objeto particular coincide con los otros objetos en base a la distancia determinada.

30 **[0026]** Las implementaciones del aspecto general anterior pueden incluir una o más de las características siguientes. Por ejemplo, proporcionar la ecuación puede incluir seleccionar la ecuación de entre una primera ecuación y una segunda ecuación, donde la primera ecuación describe un primer modelo característico de un primer conjunto de imágenes de entrenamiento, y la segunda ecuación describe un segundo modelo característico de un segundo conjunto de imágenes de entrenamiento.

35 **[0027]** De acuerdo con otro aspecto general, un sistema incluye una cámara y un dispositivo de procesamiento acoplado a la cámara. El dispositivo de procesamiento está configurado para determinar las coordenadas de un punto de imagen en un espacio multidimensional, caracterizando el punto de imagen un objeto particular. El dispositivo de procesamiento también está configurado para proporcionar una ecuación que describe un modelo en el espacio multidimensional, siendo el modelo característico de un conjunto de imágenes de entrenamiento de uno o más de otros objetos. El dispositivo de procesamiento está configurado además para aplicar las coordenadas a la ecuación para determinar una distancia entre el punto de imagen y el modelo y para determinar si el objeto particular coincide con el uno o más de otros objetos en base a la distancia determinada.

40 **[0028]** Las implementaciones del aspecto general anterior pueden incluir una o más de las siguientes características. Por ejemplo, la cámara puede ser una cámara IR. El sistema puede incluir una fuente IR. El sistema puede incluir una pantalla dispuesta frente a la cámara. La pantalla puede ser al menos translúcida de modo que la luz reflejada por un objeto dispuesto en un lado de la pantalla opuesto a la cámara pueda pasar a través de la pantalla y ser recibido por la cámara.

45 **[0029]** De acuerdo con otro aspecto general, se proporciona una herramienta para determinar coordenadas para un punto de imagen particular en un espacio multidimensional, caracterizando el punto de imagen particular un objeto particular. La herramienta también proporciona una ecuación que describe un modelo en el espacio multidimensional, siendo el modelo característico de un conjunto de imágenes de entrenamiento de uno o más de otros objetos. La herramienta también se proporciona para aplicar las coordenadas a la ecuación para determinar una distancia entre el punto de imagen particular y el modelo, y para determinar si el objeto particular coincide con uno o más de otros objetos en base a la distancia determinada.

50 **[0030]** Las implementaciones del aspecto general anterior pueden incluir una o más de las características siguientes. Por ejemplo, se puede recibir el conjunto de imágenes de entrenamiento, teniendo cada una de las imágenes del conjunto una serie de elementos de datos. El espacio multidimensional puede determinarse en base al conjunto de imágenes de entrenamiento y el espacio multidimensional puede tener una dimensionalidad que es inferior al número de elementos de datos. Se puede generar un conjunto de puntos de entrenamiento proyectando cada imagen en el conjunto de imágenes de entrenamiento en el espacio multidimensional para producir

coordenadas para un punto de entrenamiento correspondiente en el espacio multidimensional. Cada punto de entrenamiento en el conjunto de puntos de entrenamiento puede corresponder al menos a una de las imágenes en el conjunto de imágenes de entrenamiento. Se puede determinar la ecuación que describe el modelo en el espacio multidimensional característico del conjunto de imágenes de entrenamiento.

[0031] Cada imagen en el conjunto de imágenes de entrenamiento puede normalizarse antes de ser recibida. La normalización de cada imagen en el conjunto de imágenes de entrenamiento puede explicar las variaciones en el brillo. La normalización de cada imagen en el conjunto de imágenes de entrenamiento puede incluir la aplicación de una técnica de ecualización de histograma a cada imagen en el conjunto de imágenes de entrenamiento.

[0032] Los diversos aspectos, implementaciones y características pueden implementarse usando, por ejemplo, uno o más de un procedimiento, un aparato, un aparato o herramienta o dispositivo de procesamiento para realizar un procedimiento, un programa u otro conjunto de instrucciones, un aparato que incluye un programa o un conjunto de instrucciones y un medio legible por ordenador. El medio legible por ordenador puede incluir, por ejemplo, instrucciones, software, imágenes y otros datos.

[0033] Los detalles de una o más implementaciones se exponen en los dibujos adjuntos y en la descripción a continuación. Otras características resultarán evidentes a partir de la descripción y de los dibujos y a partir de las reivindicaciones.

DESCRIPCIÓN DE LOS DIBUJOS

[0034] El archivo de patente o solicitud contiene al menos un dibujo ejecutado en color. La Oficina proporcionará copias de esta patente o publicación de solicitud con dibujo(s) en color, previa solicitud y pago de la tasa necesaria.

La FIG. 1 es una ilustración de un sistema para reconocer y rastrear dedos.

La FIG. 2(a) es una imagen de muestra que ilustra dos objetos dentro de una imagen.

La FIG. 2(b) es una matriz de sombra de muestra que proporciona una ilustración de los dos objetos extraídos de la imagen de muestra de la FIG. 2(a).

La FIG. 3 ilustra un diagrama para determinar si un objeto extraído de una imagen es demasiado grande para ser un objeto de interés.

La FIG. 4(a) es un diagrama de flujo de proceso que ilustra un primer proceso para entrenar un sistema de reconocimiento y rastreo de dedos.

La FIG. 4(b) es un diagrama de flujo de proceso que ilustra un proceso para realizar la operación 410 de la FIG. 4(a).

La FIG. 5 es una imagen de muestra de un dedo.

La FIG. 6 es una ilustración de una nube de puntos de entrenamiento en un espacio propio tridimensional.

La FIG. 7(a) es otra imagen de muestra de un dedo.

La FIG. 7(b) es una imagen de muestra del mismo dedo que en la FIG. 7(a) pero teniendo el dedo una orientación diferente.

La FIG. 8(a) es otra imagen de muestra más de un dedo.

La FIG. 8(b) es una imagen de muestra del mismo dedo que en la FIG. 8(a) pero teniendo el dedo una orientación diferente.

La FIG. 9(a) es una ilustración de una nube de puntos de entrenamiento en un espacio propio tridimensional.

La FIG. 9(b) es una ilustración de las proyecciones de las imágenes de muestra de las FIGS. 8(a) y 8(b) en el espacio propio tridimensional de la FIG. 9(a).

La FIG. 10(a) es una ilustración de la nube de puntos de entrenamiento de la FIG. 9(a) y un modelo tridimensional correspondiente a la nube de puntos de imagen.

La FIG. 10(b) es una versión en color de la FIG. 10(a).

La FIG. 11(a) es una ilustración de un modelo truncado correspondiente a la nube de puntos de entrenamiento de la FIG. 9(a).

5 La FIG. 11(b) es una ilustración del modelo truncado de la FIG. 11(a) y la nube de puntos de entrenamiento correspondientes al modelo.

La FIG. 11(c) es una versión en color de la FIG. 11(b).

10 La FIG. 12 es un diagrama de flujo de proceso que ilustra un proceso para reconocer un dedo en una imagen usando el sistema entrenado con el proceso de la FIG. 4(a).

La FIG. 13 es un diagrama de flujo de proceso que ilustra un segundo proceso para entrenar un sistema de reconocimiento y rastreo de dedos.

15 La FIG. 14 es una ilustración de otra nube de puntos de entrenamiento en un espacio propio tridimensional.

La FIG. 15 es una ilustración de un primer subconjunto de la nube de puntos de entrenamiento de la FIG. 14 que forma una forma cónica.

20 La FIG. 16 es una ilustración de un segundo subconjunto de la nube de puntos de entrenamiento de la FIG. 14 que forma una forma lineal.

La FIG. 17(a) es una ilustración del primer subconjunto de puntos de entrenamiento de la FIG. 15, y un modelo de cono vertical correspondiente al primer subconjunto de puntos de entrenamiento.

25 La FIG. 17(b) es una ilustración del primer subconjunto de puntos de entrenamiento de la FIG. 15, y un modelo de cono vertical truncado correspondiente al primer subconjunto de puntos de entrenamiento.

La FIG. 17(c) es una versión en color de la FIG. 17(a).

30 La FIG. 17(d) es una versión en color de la FIG. 17(b).

La FIG. 18(a) es una ilustración del segundo subconjunto de puntos de entrenamiento de la FIG. 16, y un modelo lineal correspondiente al segundo subconjunto de puntos de entrenamiento.

35 La FIG. 18(b) es una versión en color de la FIG. 18(a).

La FIG. 19 es un diagrama de flujo de proceso que ilustra un proceso para reconocer un dedo en una imagen usando el sistema entrenado con el proceso de la FIG. 13.

40 La FIG. 20 es un diagrama de flujo de proceso que ilustra otro proceso para reconocer un objeto en una imagen.

DESCRIPCIÓN DETALLADA

45 **[0035]** Se describen sistemas y procedimientos para detectar y reconocer dedos que tocan o se mueven cerca de una superficie. Estos sistemas y procedimientos también se pueden usar para detectar y reconocer otros objetos que tocan o se mueven cerca de una superficie. Además, estos sistemas y procedimientos también pueden usarse para detectar y reconocer objetos en un campo de visión sin superficie. Más en general, estos sistemas y procedimientos pueden usarse para detectar y reconocer objetos en una imagen.

50 **[0036]** La FIG. 1 es una ilustración de un sistema ejemplar 100 para reconocer y rastrear uno o más dedos. Un iluminador infrarrojo (IR) 102 situado detrás de una superficie 104 ilumina un área 106 que se extiende tanto detrás como delante de la superficie 104. Una cámara 108 equipada con un filtro IR 110 se sitúa detrás de la superficie 104 y captura imágenes de, por ejemplo, la superficie 104 y sus proximidades. Una imagen capturada por la cámara 108 puede representarse digitalmente por elementos de datos (por ejemplo, píxeles). Un proyector 112 situado detrás de la superficie 104 puede proyectar una interfaz de usuario por ordenador en la superficie 104. La interfaz de usuario por ordenador proyectada en la superficie 104 puede ser una pantalla de ordenador convencional. El proyector 112 puede proyectar IR además de la luz visible. Por lo tanto, el proyector 112 puede estar equipado con un filtro IR opcional 120 para reducir o eliminar el IR proyectado por el proyector 112.

60 **[0037]** Las técnicas y el aparato divulgados pueden permitir que un usuario interactúe con la interfaz de usuario por ordenador usando gestos con las manos.

65 **[0038]** Por ejemplo, se puede rastrear la posición del dedo 114 de un usuario, lo que permite al usuario controlar la localización del puntero del ratón en la interfaz de usuario por ordenador tocando la superficie 104 con su dedo 114 o moviendo su dedo 114 cerca de la superficie 104. La posición del dedo 114 del usuario se puede usar, por

ejemplo, para representar la localización deseada del puntero del ratón. En algunas implementaciones, para discernir con precisión la posición del dedo 114 del usuario, puede ser deseable distinguir el dedo 114 del usuario de la palma 118 del usuario.

5 **[0039]** El sistema 100 ilustrado en la FIG. 1 es simplemente un ejemplo de una implementación y son posibles otras configuraciones. Se puede generar una imagen capturada por la cámara 108 capturando la luz reflejada por los objetos. La luz puede incluir IR, luz visible, ultravioleta o cualquier otra forma de radiación electromagnética. Por tanto, un iluminador IR 102 puede no ser requerido. El iluminador IR puede sustituir otras fuentes de luz. En una implementación alternativa, no se requiere fuente de luz. En cambio, el sistema 100 puede depender únicamente de la luz ambiental reflejada. Además, no se requiere que la cámara 108 se sitúe detrás de la superficie 104. Por ejemplo, la cámara 108 puede situarse en cualquier otro lugar del sistema 100 que se considere ventajoso. Además, el sistema 100 no necesita implementarse para facilitar la interacción humano-ordenador. En cambio, en una implementación, el sistema 100 se puede usar para detectar objetos dentro de imágenes capturadas por una cámara 108. En dicha implementación, puede no requerirse una superficie 104.

15 **[0040]** En general, muchas operaciones de procesamiento pueden estar implicadas en la detección y reconocimiento de un objeto en una imagen como un dedo. Por ejemplo, el sistema 100 primero puede ser entrenado para reconocer un objeto como un dedo. La fase de entrenamiento puede implicar capturar un conjunto de imágenes de entrenamiento de uno o más dedos situados en una o más orientaciones y construir un modelo del conjunto de imágenes de entrenamiento. En la fase de reconocimiento, el sistema 100 puede capturar una imagen, extraer un objeto de la imagen capturada y comparar el objeto extraído con el modelo del conjunto de imágenes de entrenamiento generadas durante la fase de entrenamiento para determinar si el objeto extraído es un dedo. El procesamiento que puede estar implicado en la detección y reconocimiento de un objeto en una imagen como un dedo se describe con más detalle a continuación.

20 **[0041]** Una región de interés (ROI) 116 de una imagen puede definirse como la parte de la imagen que incluye la superficie 104. La ROI de una imagen 116 puede examinarse para detectar y reconocer uno o más dedos 114. Una imagen capturada por la cámara 108 puede incluir reflejos del IR generado por el iluminador IR 102 y reflejado por uno o más dedos dentro de la ROI 116 o el IR puede reflejarse por uno o más de otros objetos dentro de la ROI 116. Por ejemplo, la superficie 104 puede reflejar IR.

25 **[0042]** Para tener en cuenta el IR reflejado consistentemente por la superficie 104 u otros objetos dentro de la ROI 116, se puede crear y restar un modelo de fondo (BGM) del IR reflejado consistentemente por la superficie 104 u otros objetos dentro de la ROI 116, a partir de cada imagen capturada por la cámara 108. Restar el BGM de cada imagen capturada por la cámara 108 puede excluir eficazmente, o al menos reducir, el IR atribuible a los objetos de fondo que se procesan a partir de una imagen capturada. En consecuencia, restar el BGM de cada imagen puede mejorar el contraste entre un objeto de interés dentro de la ROI 116 y la superficie 104 y/u otros objetos consistentemente dentro de la ROI 116.

35 **[0043]** El BGM puede crearse capturando varias imágenes de la ROI 116 sin objetos de interés presentes y a continuación promediando las imágenes capturadas píxel por píxel para crear una nueva imagen promedio o BGM. A continuación, el BGM se puede restar de cada imagen capturada por la cámara 108. Este proceso puede describirse mediante la ecuación:

$$J(x, y) = \begin{cases} 0 & \text{if } |I(x, y) - B(x, y)| < \varepsilon \\ |I(x, y) - B(x, y)| & \text{o} \end{cases} \quad (1)$$

45 donde I es una imagen, B es el BGM, ε es un umbral ajustable (por ejemplo, una constante ajustable) y J es la imagen resultante. El BGM puede actualizarse, por ejemplo, de forma aleatoria, periódica o cuando se produce un evento desencadenante.

50 **[0044]** En referencia a las FIGS. 2(a), 2(b) y 3, se ilustran ejemplos de procesos para detectar la presencia de un objeto dentro de una imagen y hacer una determinación inicial de si el objeto es un objeto de interés.

55 **[0045]** En particular, la FIG. 2(a) ilustra dos objetos 202(a), 204(a) dentro de una imagen de muestra 200(a). Como se ilustra en la FIG. 2(a), los objetos 202(a) y 204(a) se muestran más oscuros que el fondo 206(a). Sin embargo, en una imagen 200(a) formada por la luz reflejada por dos objetos 202(a), 204(a), se esperaría que los dos objetos 202(a), 204(a) fueran más brillantes que el fondo 206(a). Por tanto, se debe apreciar que la imagen 200(a) de la FIG. 2(a) es simplemente un ejemplo de dos objetos 202(a), 204(a) en una imagen 200(a) y que la imagen 200(a) no pretende sugerir que los objetos en una imagen serán más oscuros que el fondo de la imagen.

60 **[0046]** Para extraer los objetos 202(a), 204(a) de la imagen 200(a), se puede emplear un algoritmo de análisis de objetos binarios grandes (blob), tal como el algoritmo Grassfire. El algoritmo Grassfire se describe, por ejemplo, en Pitas I., Digital Image Processing Algorithms, Prentice-Hall, Nueva York (1993). Se pueden emplear otros algoritmos para detectar la presencia de un objeto dentro de una imagen y extraer el objeto de la imagen.

[0047] El algoritmo Grassfire puede buscar objetos en una imagen e identificar cada píxel de cada objeto que detecta con un marcador común. En particular, en una imagen formada por luz reflejada por objetos en los que los objetos son más brillantes que el fondo, el algoritmo Grassfire puede determinar el píxel más brillante de la imagen. Por ejemplo, la imagen puede ser una imagen en escala de grises con 256 valores posibles para cada píxel. Es decir, a cada píxel se le puede asignar un valor de 0-255, donde cero representa el píxel menos brillante posible (por ejemplo, negro absoluto) y 255 representa el píxel más brillante posible (por ejemplo, blanco absoluto). El píxel más brillante de la imagen puede tener un valor de 220. El algoritmo Grassfire puede comparar cada píxel de la imagen con una constante de umbral ajustable en relación con el píxel más brillante de la imagen para determinar si el píxel corresponde a un objeto. Por ejemplo, el valor umbral ajustable puede ser 50. Por lo tanto, el algoritmo Grassfire puede considerar que cualquier píxel de la imagen dentro de los 50 tonos del píxel más brillante representa un objeto. Es decir, cualquier píxel con un valor dentro del intervalo aceptable 170-220 puede considerarse que representa un objeto. El algoritmo Grassfire puede considerar a continuación que cualquier conjunto de píxeles adyacentes que se encuentre dentro del intervalo aceptable constituye un objeto.

[0048] El algoritmo Grassfire puede producir una matriz llamada matriz de sombra que tiene el mismo tamaño que la imagen original. Los elementos de la matriz de sombra que corresponden a píxeles identificados como representantes de un objeto en la imagen original pueden identificarse mediante marcadores comunes. La FIG. 2(b) ilustra un ejemplo de una matriz de sombra 200(b) correspondiente a la imagen de muestra 200(a) ilustrada en la FIG. 2(a). El objeto 202(a) en la imagen de muestra 200(a) está representado por el objeto 202(b) en la matriz de sombra 200(b) y cada elemento del objeto 202(b) se identifica con un marcador común "1". De forma similar, el objeto 204(a) en la imagen de muestra 200(a) está representado por el objeto 204(b) en la matriz de sombra 200(b) y cada elemento del objeto 204(b) se identifica con un marcador común "2". Como se ilustra en la FIG. 2(b), el algoritmo Grassfire puede extraer eficazmente los objetos 202(a) y 204(a) de la imagen de muestra 200(a).

[0049] En referencia nuevamente a la FIG. 1, idealmente, si el dedo 114 toca la superficie 104 o se mueve cerca de la superficie 104, entonces el dedo 114 produciría el objeto más brillante en una imagen. Desafortunadamente, el dedo 114 que toca la superficie 104 o se mueve cerca de la superficie 104 no siempre produce el objeto más brillante dentro de una imagen. Por ejemplo, cuando un usuario toca la superficie 104 con su dedo 114, la palma 118 del usuario puede reflejar suficiente radiación IR para aparecer como un objeto brillante en una imagen. Otros objetos, tal como, por ejemplo, una manga larga, también pueden reflejar suficiente radiación IR para aparecer como objetos brillantes en una imagen. Algunos objetos pueden reflejar más radiación IR que el dedo 114 del usuario y, en consecuencia, pueden aparecer como objetos más brillantes en una imagen que el dedo 114 del usuario.

[0050] Por lo tanto, en una implementación, se pueden extraer diversos objetos de una imagen, no solo el objeto más brillante. Sin embargo, procesar muchos objetos dentro de una imagen para su reconocimiento puede requerir una gran potencia de procesamiento y puede llevar mucho tiempo. Por lo tanto, se pueden aplicar técnicas para reducir el número de objetos extraídos que se procesan para su reconocimiento. Un objeto extraído puede presentar características que hacen improbable que sea un dedo y, por lo tanto, puede descartarse. Un criterio que se puede usar para ordenar los objetos extraídos puede ser el tamaño. Por ejemplo, un objeto extraído puede ser demasiado grande para ser un dedo. De forma similar, un objeto extraído puede ser demasiado pequeño para ser un dedo.

[0051] La FIG. 3 es un diagrama 301 que se usa para ilustrar un ejemplo de un proceso para determinar si un objeto 300 extraído de una imagen es demasiado grande para ser un dedo. Se pueden muestrear uno o más píxeles dentro de cuatro áreas de muestreo 304, 306, 308, 310 que rodean el centro 302 de un objeto extraído 300. La distancia entre el centro 302 de un objeto extraído 300 y las áreas de muestreo 304, 306, 308, 310 puede ser, por ejemplo, una constante ajustable. Como se muestra en la FIG. 3, dos áreas de muestreo 304, 308 que rodean el centro 302 del objeto extraído 300 pueden localizarse en una línea vertical 312 que pasa a través del centro 302 del objeto extraído 300. De forma similar, dos áreas de muestreo 306, 310 que rodean el centro 302 del objeto extraído 300 pueden localizarse en una línea horizontal 314 que pasa a través del centro 302 del objeto extraído.

[0052] Debido al hecho de que una imagen de un dedo 114 a veces puede incluir una estela (véase, por ejemplo, la FIG. 5), las áreas de muestreo 304, 306, 308, 310 pueden considerarse en pares. Por ejemplo, las dos áreas de muestreo 306, 310 localizadas en la línea horizontal 314 que pasa a través del centro 302 del objeto extraído 300 pueden considerarse como un primer par. De forma similar, las dos áreas de muestreo 304, 308 localizadas en la línea vertical 312 que pasa a través del centro 302 del objeto extraído 300 pueden considerarse como un segundo par.

[0053] Si la suma de los valores de píxeles dentro del primer par de áreas de muestreo 306, 310 excede un nivel umbral aceptable, se puede determinar que las áreas de muestreo 306, 310 son parte del objeto extraído 300. Si la suma de los valores de píxel dentro del primer par de áreas de muestreo 306, 310 excede el nivel umbral aceptable, la suma de los valores de píxel dentro del segundo par de áreas de muestreo 304, 308 se puede comparar con el nivel umbral aceptable.

[0054] Si la suma de los valores de píxeles dentro del segundo par de áreas de muestreo 304, 308 excede un nivel umbral aceptable, se puede determinar que las áreas de muestreo 304, 308 son parte del objeto extraído 300. Si la suma de los valores de píxeles dentro del primer par de áreas de muestreo 306, 310 excede el nivel umbral aceptable y la suma de los valores de píxeles dentro del segundo par de áreas de muestreo 304, 308 excede el nivel umbral aceptable, se puede determinar que el objeto es demasiado grande para ser un dedo y, en consecuencia, el objeto puede descartarse.

[0055] De forma similar, los objetos extraídos de una imagen pueden ser demasiado pequeños para ser un dedo. Se puede contar el número de píxeles que representan cada objeto extraído. Si el número de píxeles que representan un objeto es menor que una constante de umbral, se puede determinar que el objeto es ruido o cualquier otro objeto y, en consecuencia, se puede descartar el objeto.

[0056] Reconocer una imagen en un objeto como un dedo puede implicar una fase de entrenamiento y una fase de reconocimiento. En la fase de entrenamiento, el sistema 100 puede entrenarse para reconocer un objeto exponiendo el sistema 100 a un gran conjunto de imágenes de entrenamiento del objeto a reconocer. El conjunto de imágenes de entrenamiento puede contener imágenes del objeto a reconocer situadas en diversas orientaciones diferentes. Durante la fase de reconocimiento, el sistema 100 puede detectar un objeto en una imagen y comparar el objeto con el conjunto de imágenes de entrenamiento, o un modelo del conjunto de imágenes de entrenamiento, para determinar si el objeto es el objeto a reconocer.

[0057] En una implementación, durante la fase de entrenamiento, se captura una gran cantidad de imágenes de uno o más dedos situados en diversas orientaciones diferentes. Se determina la matriz de covarianza del conjunto de imágenes de entrenamiento y se usa un conjunto seleccionado de vectores propios de la matriz de covarianza para definir un espacio propio. Se puede seleccionar cualquier número de vectores propios para definir el espacio propio. La dimensionalidad del espacio propio se define por el número de vectores propios seleccionados para definir el espacio propio. Por ejemplo, un espacio propio tridimensional se define seleccionando tres vectores propios, por ejemplo, los vectores propios correspondientes a los tres valores propios más grandes. Cada imagen de entrenamiento del conjunto de imágenes de entrenamiento se proyecta en el espacio propio tridimensional, creando una colección de puntos tridimensionales en el espacio propio. La colección de puntos tridimensionales en el espacio propio se modela mediante un modelo geométrico tridimensional (por ejemplo, una superficie o línea cuadrática) capaz de ser expresada analíticamente por un polinomio. Por ejemplo, la proyección del conjunto de imágenes de entrenamiento puede formar una superficie similar a un hiperboloide o a un cono. Adicionalmente o de forma alternativa, la proyección del conjunto de imágenes de entrenamiento puede formar una forma geométrica lineal en el espacio propio. En un espacio tridimensional, los hiperboloides, conos y líneas pueden expresarse analíticamente mediante polinomios. Por lo tanto, las formas geométricas (también denominadas modelos geométricos) se pueden usar para modelar dicho conjunto de imágenes de entrenamiento en un espacio propio.

[0058] Durante la fase de reconocimiento de esta implementación, un objeto extraído de una imagen se proyecta en el espacio propio definido durante la fase de entrenamiento. Las tres coordenadas que identifican el punto tridimensional correspondiente a la imagen proyectada en el espacio propio se aplican al polinomio que define el modelo del conjunto de imágenes de entrenamiento en el espacio propio para determinar la distancia del objeto proyectado a partir del modelo en el espacio propio. Si el objeto proyectado se encuentra, por ejemplo, dentro de una distancia definida del modelo, se puede determinar que el objeto proyectado es un dedo.

[0059] En referencia ahora a las FIGS. 5-20, las implementaciones de las fases de entrenamiento y reconocimiento se describen con más detalle.

[0060] La FIG. 4(a) es un diagrama de flujo de proceso que ilustra un proceso ejemplar 400 para entrenar el sistema 100 para reconocer un objeto en una imagen como un dedo. El proceso 400 comienza creando un BGM del IR reflejado por la superficie 104 u otros objetos consistentemente dentro de la ROI 116 (operación 402). A continuación, la cámara 108 (operación 404) captura una gran cantidad de imágenes de entrada de un dedo 114 situado en diversas orientaciones diferentes. El BGM se resta a continuación de cada una de las imágenes de entrada (operación 406). La parte de cada imagen que representa el dedo 114 se extrae de la imagen y se convierte en un tamaño de imagen estándar $n \times n$ (operación 408). La parte de cada imagen que representa el dedo 114 puede extraerse usando un algoritmo de análisis de objetos binarios grandes (blob), tal como el algoritmo Grassfire, o la parte de cada imagen que representa el dedo 114 puede extraerse manualmente en base a la apariencia visual de la imagen. La parte extraída de la imagen que representa el dedo puede incluir una gran cantidad de píxeles. Para reducir la potencia de cálculo implicada en el procesamiento de la parte de la imagen que representa el dedo, puede ser deseable reducir el número de píxeles usados para representar el dedo. Por lo tanto, la resolución de la parte extraída de la imagen que representa el dedo puede reducirse. Por ejemplo, la parte de la imagen que representa el dedo en la imagen original puede ser de 64 píxeles por 64 píxeles. Después de que se extrae la parte de 64x64 de la imagen que representa el dedo, la resolución del dedo extraído se puede reducir de modo que el dedo esté representado por un imagen de 16x16.

[0061] La FIG. 5 ilustra una imagen ejemplar de $n \times n$ 500 de un dedo 502.

[0062] A continuación, se define un espacio propio para el conjunto de imágenes de entrenamiento de $n \times n$ (operación 410). La operación 410 se ilustra con más detalle en la FIG. 4(b). En primer lugar, se determina la matriz de covarianza C del conjunto de imágenes de entrenamiento (operación 410(a)). Si una imagen remodelada, $n \times n$, $I_{(n \times n)}$, está representada por un vector $V_{(m)}$ donde $m = n^2$, la matriz de covarianza C del conjunto de imágenes de entrenamiento $t \{I_1, I_2, \dots, I_t\}$ puede definirse como:

$$C = G^T G, \quad (2)$$

donde G es una matriz $t \times m$ con elementos $G_{ij} = (V_{ij} - \mu_i)$ en los que V_{ij} es el j -ésimo elemento del vector V_i , el vector de la imagen remodelada I_i del conjunto de entrenamiento y μ_i es la media del vector V_i . La matriz de covarianza C , por lo tanto, es una matriz de $m \times m$. A continuación, se pueden determinar los valores propios y los vectores propios de la matriz de covarianza (operación 410(b)). Los valores propios y los vectores propios de la matriz de covarianza se pueden obtener resolviendo la ecuación:

$$\Delta = \Phi^T C \Phi, \quad (3)$$

donde Δ es el conjunto de valores propios de la matriz de covarianza C y Φ es el conjunto de vectores propios de la matriz de covarianza C . Este proceso se describe en la patente de Estados Unidos N.º 5.710.833. Una matriz de covarianza de $m \times m$ tendrá m valores propios y m vectores propios, correspondiendo cada uno de los vectores propios a un único valor propio. Por tanto, una matriz de covarianza de 256×256 para un conjunto de imágenes de entrenamiento de 16×16 tendrá 256 valores propios y 256 vectores propios correspondientes. Además, cada vector propio será un vector columna de longitud 256.

[0063] Todos los vectores propios de la matriz de covarianza C son perpendiculares entre sí. En consecuencia, se puede definir un espacio propio para el conjunto de imágenes de entrenamiento seleccionando un conjunto de vectores propios de la matriz de covarianza C y usando cada uno de los vectores propios seleccionados para definir una dirección (es decir, dimensión) en el espacio (operación 410(c)). Los vectores propios correspondientes a los valores propios más grandes de la matriz de covarianza C indican las direcciones en las que el conjunto de imágenes de entrenamiento presenta la mayor variación. Por lo tanto, una gran parte de los datos contenidos dentro del conjunto de imágenes de entrenamiento puede caracterizarse por la selección de un conjunto de vectores propios correspondientes a varios de los vectores propios más grandes de la matriz de covarianza C . En una implementación, los vectores propios correspondientes a los tres valores propios más grandes de la matriz de covarianza C se seleccionan para definir un espacio propio tridimensional.

[0064] Después de que se haya definido un espacio propio, cada imagen de entrenamiento del conjunto de imágenes de entrenamiento se proyecta en el espacio propio (operación 410(d)). Cada imagen de $n \times n$ $I_{(n \times n)}$ se convierte en un vector $V_{(m)}$ donde $m = n^2$. Por ejemplo, si la imagen $I_{(n \times n)}$ es una imagen de 16×16 , el vector $V_{(m)}$ es un vector fila de longitud 256 ($m = n^2 = 16 \times 16 = 256$). El espacio propio puede definirse mediante una matriz $m \times q$ donde q es el número de vectores propios seleccionados para definir el espacio propio y cada una de las columnas q de la matriz representa uno de los vectores propios seleccionados para definir el espacio propio. A continuación, el vector $V_{(m)}$ puede multiplicarse por la matriz $m \times q$ que define el espacio propio dando como resultado una matriz de $1 \times q$, o vector fila, en el que cada elemento del vector fila identifica una coordenada correspondiente en el espacio propio. Por ejemplo, una imagen de 16×16 $I_{(16 \times 16)}$ puede estar representada por una columna vector de 256×1 $V_{(256 \times 1)}$. Si el espacio propio para el conjunto de imágenes de entrenamiento está definido por tres vectores propios, la transposición del vector de imagen $V_{256 \times 1}$ puede multiplicarse por la matriz de 256×3 que define el espacio propio para obtener un vector fila de 1×3 que define las tres coordenadas de la imagen proyectada en el espacio propio tridimensional.

[0065] Por tanto, la proyección de una imagen en un espacio propio puede reducir eficazmente la dimensionalidad de la imagen. La proyección de una imagen en un espacio propio es un único punto con tantas coordenadas como las dimensiones del espacio propio. Por ejemplo, la proyección de una imagen en un espacio propio tridimensional es un punto tridimensional (es decir, el punto está definido por tres coordenadas). Por tanto, cuando una imagen de $n \times n$ se proyecta en un espacio propio tridimensional, la imagen se transforma de ser definida por n^2 píxeles a ser definida por un único punto identificado por tres coordenadas.

[0066] La FIG. 6 ilustra un ejemplo de un espacio propio tridimensional 600 para un conjunto de imágenes de entrenamiento de uno o más dedos situados en diversas orientaciones. La proyección de cada imagen del conjunto de imágenes de entrenamiento está representada por un punto tridimensional en el espacio propio 600. Por lo tanto, como se ilustra en la FIG. 6, proyectar el conjunto de imágenes de entrenamiento en el espacio propio 600 puede crear una nube de puntos de entrenamiento 602 en el espacio propio 600. La nube de puntos de entrenamiento 602 puede presentar una forma geométrica reconocible. Por ejemplo, la nube de puntos de entrenamiento 602 en la FIG. 6 parece presentar una forma de hiperboloide o de cono.

[0067] La forma de la nube de puntos de entrenamiento 602 puede ser una función de las características especiales de un dedo.

5 **[0068]** Cuando un dedo 114 es perpendicular a la superficie 104, el dedo 114 puede aparecer como un objeto sustancialmente circular en una imagen capturada por la cámara 108. Incluso si el usuario gira su mano, el dedo 114 puede continuar apareciendo como un objeto sustancialmente circular en una imagen capturada por la cámara 108 siempre que el dedo 114 permanezca perpendicular a la superficie 104. En otras palabras, si el dedo 114 permanece perpendicular a la superficie 104, la forma del dedo 114 en una serie de imágenes capturadas por la cámara 108 puede presentar solo ligeras variaciones incluso si se gira la mano. Este proceso de presentar solo pequeñas variaciones puede mantenerse independientemente de dónde, con respecto a la cámara 108, el dedo 114 esté apuntando sobre la superficie 104.

15 **[0069]** Sin embargo, si el dedo 114 del usuario no es perpendicular a la superficie 104, el dedo 114 puede aparecer como un punto brillante con una estela en una imagen capturada por la cámara 108. La estela puede ser IR reflejado por el cuerpo del dedo 114. En consecuencia, si se gira la mano, el ángulo de la estela gira.

[0070] Las FIGS. 7(a) y 7(b) son ilustrativas. La FIG. 7(a) es una imagen de $n \times n$ 700(a) de un dedo 114 con una estela 702(a). En la FIG. 7(a), la estela 702(a) se dirige hacia la esquina superior izquierda de la imagen 700(a). La FIG. 7(b) es una imagen de $n \times n$ 700(b) del mismo dedo 114 con una estela 702(b). En la FIG. 7(b), la estela 702(b) se dirige hacia la esquina superior derecha de la imagen 700(b). La orientación diferente de las estelas 702(a), 702(b) puede explicarse por el hecho de que el dedo 114 está orientado de manera diferente con respecto a la cámara 108 en las dos imágenes 700(a), 700(b). El dedo 114 en la imagen 700(a) y el dedo 114 en la imagen 700(b) forman el mismo ángulo no perpendicular con la superficie 104. Sin embargo, el dedo 114 en la imagen 700(b) se ha girado desde la posición del dedo 114 en la imagen 700(a) de modo que el dedo 114 en la imagen 700(a) y el dedo 114 en la imagen 700(b) forman diferentes ángulos con un plano (no mostrado) normal a la superficie 104.

30 **[0071]** Debido al hecho de que las dos imágenes 700(a), 700(b) son diferentes, estarán representadas por diferentes puntos en la nube de puntos de entrenamiento 602 en el espacio propio 600. Sin embargo, debido a que el proceso de proyectar imágenes en un espacio propio 600 es un proceso lineal, la proyección de un conjunto de imágenes que capturan un dedo 114 con una estela de sustancialmente la misma longitud pero con diferentes ángulos de rotación puede dar como resultado un conjunto de puntos alineados en un patrón sustancialmente circular (no mostrado) en el espacio propio 600. Por lo tanto, mientras las dos imágenes 700(a), 700(b) estén representadas por diferentes puntos en la nube de puntos de entrenamiento 602 en el espacio propio 600, pueden alinearse a lo largo de un patrón sustancialmente circular dentro del espacio propio 600.

40 **[0072]** La longitud de la estela de un dedo en una imagen también puede afectar la localización de la proyección de la imagen en el espacio propio 600. Como se describe anteriormente en relación con las FIGS. 7(a) y 7(b), las proyecciones de imágenes de dedos con sustancialmente la misma longitud de estela pero diferentes ángulos de rotación pueden alinearse en un patrón sustancialmente circular en el espacio propio 600. La proyección de un conjunto de imágenes de un dedo 114 con la misma estela corta pero diferentes ángulos de rotación puede alinearse en un patrón sustancialmente circular con un radio relativo más pequeño que la proyección de un conjunto de imágenes de un dedo 114 con la misma estela larga pero diferentes ángulos de rotación.

45 **[0073]** Las FIGS. 6, 8(a), 8(b), 9(a) y 9(b) son ilustrativos. Los puntos de entrenamiento que forman un vértice 604 de la nube de puntos de entrenamiento 602 en la FIG. 6 pueden asociarse con imágenes de entrenamiento en las que el dedo 114 aparece como una forma sustancialmente circular con poca o ninguna estela. Por el contrario, los puntos de entrenamiento que forman una base 606 de la nube de puntos de entrenamiento 602 en la FIG. 6 pueden estar asociados con imágenes de entrenamiento en las que el dedo 114 es arrastrado por una larga estela.

50 **[0074]** La FIG. 8(a) ilustra una imagen de entrenamiento 800(a) de un dedo 114 con una estela relativamente larga 802(a) causada por un ángulo relativamente pequeño entre el dedo 114 y la superficie 104. La FIG. 8(b) ilustra una imagen de entrenamiento 800(b) de un dedo 114 con una estela relativamente más corta 802(b) causada por un ángulo relativamente grande entre el dedo 114 y la superficie 104.

60 **[0075]** Las FIGS. 9(a) y 9(b) ilustran las proyecciones 800(a)', 800(b)' de las dos imágenes 800(a), 800(b) en el espacio propio tridimensional 600. Como se ilustra en la FIG. 9(a), la imagen 800(a) del dedo 114 con la estela más larga 802(a) se proyecta sobre un punto 800(a)' cerca de la base de la nube de puntos de entrenamiento 602 porque tiene una estela relativamente larga 802(a). Por el contrario, la imagen 800(b) del dedo 114 con la estela más corta 802(b) se proyecta sobre un punto 800(b)' cerca del vértice 604 de la nube de puntos de entrenamiento 602 porque tiene una estela relativamente pequeña 802(b).

65 **[0076]** La FIG. 9(b) ilustra las proyecciones 800(a)', 800(b)' de las dos imágenes 800(a), 800(b) en el espacio propio tridimensional 600 aisladas de la nube de puntos de entrenamiento 602. Como se ilustra en la FIG. 9(b), la

proyección 800(a)' está alineada en un patrón sustancialmente circular 902 con proyecciones de imágenes de dedos que presentan la misma estela relativamente larga pero diferentes ángulos de rotación. De forma similar, la proyección 800(b)' está alineada en un patrón sustancialmente circular 904 con proyecciones de imágenes de dedos que presentan la misma estela relativamente corta pero diferentes ángulos de rotación.

[0077] La nube de puntos de entrenamiento 602 puede presentar una forma geométrica reconocible y, por lo tanto, la nube de puntos de entrenamiento 602 puede ser modificada por un modelo geométrico capaz de ser expresado analíticamente por un polinomio. Para facilitar la adaptación de un modelo geométrico a la nube de puntos de entrenamiento 602, se puede aplicar una técnica conocida como la técnica de Transferencia al Centro (TTC) a cada uno de los puntos de entrenamiento (operación 412). La técnica de TTC se describe en, por ejemplo, Shamaie A. y col., "International Journal of Scientia Iranica", 6 (1), (1999). La técnica de TTC transfiere el conjunto de datos al centro del espacio propio. En otras palabras, el centroide del conjunto de datos se mueve al origen del espacio propio.

[0078] Después de que la técnica de TTC se haya aplicado al conjunto de puntos de entrenamiento, se ajusta un modelo a la nube de puntos de entrenamiento 602 (operación 414). La ecuación general para una superficie cuadrática tridimensional viene dada por:

$$F(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + 1 \quad (4)$$

[0079] Si la nube de puntos 602 definiera una superficie cuadrática perfecta, la siguiente ecuación se mantendría para cada punto en la nube 602:

$$F(x, y, z) = 0 \quad (5)$$

donde x, y y z son las coordenadas de cada punto en la nube de puntos 602. Los parámetros desconocidos de la Ecuación 4 son $\Psi = (a, b, c, d, e, f, g, h, i)$. Para ajustar una superficie cuadrática a la nube de puntos de entrenamiento 602, los valores para los parámetros desconocidos en la Ecuación 4 se determinan de modo que se minimice una función de error para el conjunto de puntos en la nube de puntos de entrenamiento 602. La función de error se define mediante la ecuación:

$$E = \sum_{(x,y,z) \in TS} F^2(x, y, z) \quad (6)$$

donde TS es el conjunto de puntos de entrenamiento proyectados. La función de error se puede ampliar a:

$$\begin{aligned} \sum F^2(x, y, z) = & a^2 \sum x^4 + 2ab \sum x^2 y^2 + 2ac \sum x^2 z^2 + 2ad \sum x^3 y + 2ae \sum x^3 z + \\ & 2af \sum x^2 yz + 2ag \sum x^3 + 2ah \sum x^2 y + 2ai \sum x^2 z + 2a \sum x^2 + b^2 \sum y^4 + 2bc \sum y^2 z^2 + \\ & 2bd \sum xy^3 + 2be \sum xy^2 z + 2bf \sum y^3 z + 2bg \sum xy^2 + 2bh \sum y^3 + 2bi \sum y^2 z + \\ & 2b \sum y^2 + c^2 \sum z^4 + 2cd \sum xyz^2 + 2ce \sum xz^3 + 2cf \sum yz^3 + 2cg \sum xz^2 + 2ch \sum yz^2 + \\ & 2ci \sum z^3 + 2c \sum z^2 + d^2 \sum x^2 y^2 + 2de \sum x^2 yz + 2df \sum xy^2 z + 2dg \sum x^2 y + \\ & 2dh \sum xy^2 + 2di \sum xyz + 2d \sum xy + e^2 \sum x^2 z^2 + 2ef \sum xyz^2 + 2eg \sum x^2 z + 2eh \sum xyz + \\ & 2ei \sum xz^2 + 2e \sum xz + f^2 \sum y^2 z^2 + 2fg \sum xyz + 2fh \sum y^2 z + 2fi \sum yz^2 + \\ & 2f \sum yz + g^2 \sum x^2 + 2gh \sum xy + 2gi \sum xz + 2g \sum x + h^2 \sum y^2 + 2hi \sum yz + 2h \sum y + \\ & i^2 \sum z^2 + 2i \sum z + \sum 1 \end{aligned} \quad (7)$$

[0080] La función de error puede minimizarse usando un procedimiento cuasi-newtoniano. Desafortunadamente, los procedimientos de optimización lineal, como el procedimiento cuasi-newtoniano, pueden adherirse a los mínimos locales porque la función de error es cuadrática. Sin embargo, la aplicación de la técnica de TTC al conjunto de puntos de entrenamiento antes de minimizar la función de error reduce el riesgo de que el procedimiento de minimización cuasi-newtoniano se atasque en los mínimos locales. Por lo tanto, la función de error puede converger más rápidamente si la técnica de TTC se aplica al conjunto de puntos de entrenamiento antes de usar el procedimiento cuasi-newtoniano para minimizar la función de error. Minimizar la función de error genera valores para el conjunto de parámetros desconocidos Ψ . Los valores determinados para Ψ minimizando la función de error se conectan a la Ecuación 4 para definir un modelo de la nube de puntos de entrenamiento 602.

[0081] La FIG. 10(a) ilustra la nube de puntos de entrenamiento (puntos oscuros) 602 superponiendo un modelo ejemplar (círculos vacíos) de la nube de puntos de entrenamiento 602. La FIG. 10(b) es una versión en color de la FIG. 10(a). El modelo 1000 puede definirse sustituyendo en la Ecuación 4 los valores para el conjunto de

parámetros Ψ determinados minimizando la función de error. Como se ilustra en las FIGS. 10(a) y 10(b), el modelo tiene una primera superficie 1002 y una segunda superficie 1004 que aparecen ambas como formas hiperboloides con vértices 1002a y 1004a, respectivamente, una frente a la otra, y con ambas superficies apoyadas sobre una eje común (no se muestra). En una implementación, solo la primera superficie 1002 se usa para modelar la nube de puntos de entrenamiento 602. Como se muestra en las FIGS. 10(a)-(b), la nube de puntos de entrenamiento 602 no cubre completamente la primera superficie 1002 y no cubre la segunda superficie 1004 en absoluto.

[0082] La FIG. 11(a) es una ilustración de la primera superficie 1002 del modelo 1000 para modelar la nube de puntos de entrenamiento 602 en el espacio propio 600. La segunda superficie 1004 no está incluida en las FIGS. 11(a)-(c). La FIG. 11(b) ilustra la nube de puntos de entrenamiento (puntos oscuros) 602 superpuestos a la primera superficie 1002 del modelo (círculos vacíos) 1000. La FIG. 11(c) es una versión en color de la FIG. 11(b).

[0083] La FIG. 12 es un diagrama de flujo de proceso que ilustra un proceso ejemplar 1200 para reconocer un objeto en una imagen como un dedo usando un sistema entrenado con el proceso 400. El proceso 1200 comienza creando un BGM del IR reflejado consistentemente por la superficie 104 u otros objetos dentro de la ROI 116 (operación 1202). La cámara 108 captura una imagen de entrada de la superficie 104 y su proximidad (operación 1204), y el BGM se resta de la imagen de entrada (operación 1206).

[0084] Como se describe con más detalle anteriormente en relación con las FIGS. 2(a) y 2(b), un algoritmo de análisis de objetos binarios grandes (blob), como Grassfire, se usa para extraer objetos a partir de la imagen de entrada (operación 1208). Como se analiza con más detalle anteriormente en relación con la FIG. 3, cada uno de los objetos extraídos de la imagen de entrada se procesa para identificar y descartar objetos que son demasiado pequeños o demasiado grandes para ser un dedo (operación 1210). Los objetos restantes se ordenan en una lista en orden ascendente del área general (operación 1212).

[0085] El primer objeto en la lista se convierte en una imagen de $n \times n$ y la imagen redimensionada $n \times n$ se proyecta en el espacio propio 600 para el conjunto de datos de entrenamiento para obtener un punto de imagen (operación 1214). La técnica de TTC se aplica al punto de imagen para imitar la aplicación de la técnica de TTC a los puntos en la nube de puntos de entrenamiento 602 durante la fase de entrenamiento 400 (operación 1216). Los mismos valores de transferencia de TTC usados en la operación 412 pueden usarse en la operación 1216.

[0086] Las coordenadas del punto de imagen se usan para evaluar el polinomio cuadrático que define el modelo 1000 de la nube de puntos de entrenamiento 602 (operación 1218). Si el punto de imagen se encuentra en el modelo 1000, la aplicación de las coordenadas del punto de imagen al polinomio producirá un valor de cero. Por el contrario, si el punto de imagen no se encuentra en el modelo 1000, se obtiene un número real distinto de cero aplicando las coordenadas del punto de imagen al polinomio. El valor obtenido al aplicar las coordenadas del punto de imagen al polinomio representa la distancia entre el punto de imagen y el modelo 1000.

[0087] Una imagen correspondiente a un punto de imagen que se localiza cerca del modelo 1000 en el espacio propio 600 puede presentar características similares a las imágenes que comprenden el conjunto de imágenes de entrenamiento. En consecuencia, puede darse el caso de que cuanto más cerca se encuentre un punto de imagen del modelo 1000, más probable es que la imagen correspondiente al punto de imagen sea un dedo. Por lo tanto, una imagen correspondiente a un punto de imagen que se encuentra en el modelo 1000 o un punto de imagen que se encuentra dentro de una distancia de umbral máxima del modelo 1000 puede determinarse como un dedo.

[0088] Por lo tanto, el valor obtenido al aplicar las coordenadas del punto de imagen al polinomio cuadrático se evalúa para determinar si es menor que una distancia umbral definida (operación 1220). Si el valor obtenido al aplicar las coordenadas del punto de imagen al polinomio cuadrático es menor que la distancia umbral, el objeto correspondiente al punto de imagen se considera un dedo (operación 1222). Si el valor obtenido al aplicar las coordenadas del punto de imagen al polinomio cuadrático es mayor que la distancia umbral máxima, la imagen correspondiente al punto de imagen se descarta y el siguiente objeto de la lista se procesa al pasar a la operación 1214 (operación 1224).

[0089] La reducción de la dimensionalidad de un objeto capturado en una imagen proyectando el objeto en un espacio propio permite comparar el objeto con un modelo de imágenes de objetos de entrenamiento sin tener que comparar cada píxel del objeto capturado con cada píxel del modelo. Como resultado, se ahorra el poder de procesamiento y los recursos y/o se incrementa la velocidad de la comparación.

[0090] En referencia ahora a las FIGS. 13-19, se describen implementaciones de las fases de entrenamiento y reconocimiento.

[0091] La FIG. 13 es un diagrama de flujo de proceso que ilustra un ejemplo de un proceso 1300 para entrenar un sistema de reconocimiento y rastreo de dedos 100 para reconocer un objeto como un dedo. El proceso 1300 incluye la operación de aplicar una técnica de ecualización de histograma a cada imagen de entrenamiento.

[0092] El proceso 1300 comienza creando un BGM del IR reflejado consistentemente por la superficie 104 u otros objetos dentro de la ROI 116 (operación 1302). La cámara 108 captura una gran cantidad de imágenes de entrada de uno o más dedos situados en diversas orientaciones diferentes (operación 1304), y el BGM se resta de cada una de las imágenes de entrada (operación 1306). La parte de cada imagen que representa el dedo se extrae de la imagen y se convierte en un tamaño de imagen estándar de $n \times n$ (operación 1308). Se aplica una técnica de ecualización de histograma a cada imagen de $n \times n$ (operación 1310).

[0093] La técnica de ecualización de histograma se aplica a las imágenes de $n \times n$ para tener en cuenta las variaciones en las condiciones de iluminación. La aplicación de la técnica de ecualización de histograma a una imagen de $n \times n$ implica generar un histograma de las intensidades de los píxeles en la imagen de $n \times n$, normalizando el histograma de la imagen de $n \times n$, y reasignando los valores de los píxeles en la imagen de $n \times n$ en base al histograma de imagen normalizada. En consecuencia, los píxeles individuales conservan su orden de brillo (por ejemplo, permanecen más brillantes o más oscuros que otros píxeles).

[0094] Se crea un espacio propio correspondiente al conjunto de imágenes de entrenamiento de $n \times n$ del dedo y cada imagen de entrenamiento se proyecta en el espacio propio (operación 1312).

[0095] Como se ilustra en la FIG. 14, las imágenes de entrenamiento proyectadas forman una nube de puntos de entrenamiento 1402 en el espacio propio 1400. La técnica de TTC se aplica a la nube de puntos de entrenamiento 1402 para transferir el centroide de la nube de puntos de entrenamiento 1402 al origen del espacio propio 1400 (operación 1314). La forma de la nube de puntos de entrenamiento 1402 en la FIG. 14 parece diferente a la forma de la nube de puntos de entrenamiento 602 en la FIG. 6. En particular, la nube de puntos de entrenamiento 1402 en la FIG. 14 tiene una estela 1406 en un vértice 1408 de una forma cónica 1404. La diferencia de forma entre las dos nubes de puntos de entrenamiento 1402, 602 puede atribuirse a haber aplicado la técnica de ecualización de histograma al conjunto de imágenes de entrenamiento (operación 1310) porque la técnica de ecualización de histograma reduce las variaciones en el conjunto de imágenes de entrenamiento debido a variaciones en condiciones de iluminación. En consecuencia, cuando las imágenes de entrenamiento se proyectan en el espacio propio 1400, se obtiene una forma más uniforme. La forma de la nube de puntos de entrenamiento 602 en la FIG. 6 refleja variación en condiciones de iluminación, así como variación en la forma y orientación del dedo dentro del conjunto de imágenes de entrenamiento. Por el contrario, la forma de la nube de puntos de entrenamiento 1402 en la FIG. 14 refleja principalmente la variación en la forma y orientación de los dedos dentro del conjunto de imágenes de entrenamiento.

[0096] El proceso 1300 incluye ajustar uno o más modelos a la nube transferida de puntos de entrenamiento 1402 (operación 1316). Para modelar la nube de puntos de entrenamiento 1402, el subconjunto de puntos de entrenamiento que forma la forma cónica 1404 y el subconjunto de puntos de entrenamiento que forman la estela 1406 se consideran por separado. La FIG. 15 ilustra el subconjunto de puntos de entrenamiento que forman la forma cónica 1404 sin la estela 1406. La FIG. 16 ilustra el subconjunto de puntos de entrenamiento que forman la estela 1406 sin la forma cónica 1404.

[0097] Se puede crear un modelo del subconjunto de puntos de entrenamiento que forman la forma cónica 1404 y se puede crear un segundo modelo del subconjunto de puntos de entrenamiento que forman la estela 1406. Como se analiza anteriormente, la Ecuación 4 proporciona la fórmula general para una superficie cuadrática tridimensional. Por lo tanto, un modelo del conjunto de puntos de entrenamiento que forman la forma cónica 1404 se puede determinar determinando primero los parámetros desconocidos Ψ de la Ecuación 4 que minimizan la función de error de la Ecuación 7 con respecto al conjunto de puntos en el subconjunto de puntos de entrenamiento que forman la forma cónica 1404. Los valores determinados para Ψ minimizando la función de error se conectan a la Ecuación 4 para definir un modelo de la nube de puntos de entrenamiento 1402.

[0098] De forma alternativa, se puede usar un cono vertical para modelar el subconjunto de imágenes de entrenamiento que forman la forma cónica 1404. Como se ilustra en la FIG. 15, la nube en forma cónica de puntos de entrenamiento 1404 presenta la mayor cantidad de variación a lo largo del eje vertical 1500 que representa el primer componente principal del espacio propio 1400. En otras palabras, la dimensión de altura de la nube en forma cónica 1404 es paralela al primer componente principal del espacio propio 1400. La observación de que la nube cónica de puntos de entrenamiento 1404 presenta la mayor cantidad de variación a lo largo del primer componente principal del espacio propio 1400 es consecuente con el hecho de que el vector propio de la matriz de covarianza correspondiente al primer componente principal representa la dirección en la cual el conjunto de imágenes de entrenamiento presenta la mayor variación. Por lo tanto, puede ser posible usar un cono vertical para modelar el subconjunto de puntos de entrenamiento que forman la forma cónica 1404. La ecuación describe un cono vertical general:

$$H(x, y, z) = \frac{(x-a)^2}{d^2} + \frac{(y-b)^2}{e^2} - \frac{(z-c)^2}{f^2} \quad (8)$$

donde x e y representan los ejes horizontales 1502, 1504 y z representa el eje vertical 1500 del espacio propio 1400. Los parámetros desconocidos de la Ecuación 8 son $\Omega = (a, b, c, d, e, f)$. Para ajustar un cono vertical al subconjunto de imágenes de entrenamiento que forman la forma cónica 1404, se determinan los valores de los parámetros desconocidos Ω que minimizan la Ecuación 8 con respecto al conjunto de puntos en el subconjunto de puntos de entrenamiento que forma la forma cónica 1404. La ecuación puede definir una función de error para la Ecuación 8:

$$E = \sum_{(x,y,z) \in \mathcal{S}} H^2(x, y, z)$$

(9)

Se puede usar un procedimiento cuasi-newtoniano para minimizar la función de error de la Ecuación 9. Minimizar la función de error de la Ecuación 9 genera valores para el conjunto de parámetros desconocidos Ω . Estos valores se conectan a la Ecuación 8 para definir un modelo de cono vertical del subconjunto de puntos de entrenamiento que forman la forma cónica 1404.

[0099] La FIG. 17(a) ilustra la nube de puntos de entrenamiento de la forma cónica (diamantes oscuros) 1404 superponiendo un modelo de cono vertical ejemplar (círculos vacíos) 1700 de la nube de puntos de entrenamiento de la forma cónica 1404. La FIG. 17(c) es una versión en color de la FIG. 17(a). Como se ilustra en las FIGS. 17(a) y 17(c), el modelo de cono vertical 1700 tiene tanto una superficie inferior 1702 como una superficie superior 1704. Cada superficie 1702 y 1704 forma una forma cónica con las dos formas cónicas reunidas en un vértice común 1706. Las dos formas cónicas se encuentran en un eje común (no mostrados). La nube de puntos de entrenamiento que forman la forma cónica 1404 no cubren completamente la superficie 1702 y no cubren la superficie 1704 en absoluto. En una implementación, solo la superficie inferior 1702 del cono 1700 se usa para modelar la nube de puntos de entrenamiento de la forma cónica 1404. La FIG. 17(b) ilustra la nube de puntos de entrenamiento de la forma cónica (diamantes oscuros) 1404 superponiendo la superficie inferior 1702 del modelo de cono vertical ejemplar (círculos vacíos) 1700 de la nube de puntos de entrenamiento de la forma cónica 1404. La FIG. 17(d) es una versión en color de la FIG. 17(b).

[0100] Un modelo para la estela 1406 es una línea vertical. El modelo de línea vertical de la estela 1406 se determina calculando la media de los puntos de entrenamiento de la estela 1406 a lo largo de los ejes horizontales 1602, 1604. La FIG. 18(a) ilustra la nube de puntos de entrenamiento de la estela (puntos oscuros) 1406 superponiendo un modelo de línea vertical ejemplar (línea oscura) 1800 de la nube de puntos de entrenamiento que forman la estela 1406. La nube de puntos de entrenamiento que forman la estela 1406 se agrupan densamente alrededor de una porción inferior 1810 de la línea 1800 y adoptan gradualmente una forma sustancialmente lineal más cerca de una porción superior 1820 de la línea 1800. La FIG. 18(b) es una versión en color de la FIG. 18(a). En una implementación alternativa, la línea 1800 puede no ser vertical. En cambio, la línea 1800 puede estar orientada para que coincida mejor con el conjunto de puntos de entrenamiento que forman la estela 1406.

[0101] La FIG. 19 es un diagrama de flujo de proceso que ilustra un ejemplo de un proceso 1900 para reconocer un dedo en una imagen. El proceso 1900 usa un sistema entrenado con el proceso 1300.

[0102] El proceso 1900 comienza creando un BGM del IR reflejado consistentemente por la superficie 104 u otros objetos dentro de la ROI 116 (operación 1902). La cámara 108 captura la imagen de entrada de la superficie 104 y sus proximidades (operación 1904), y el BGM se resta de la imagen de entrada (operación 1906). Como se describe con más detalle anteriormente en relación con las FIGS. 2(a) y 2(b), un algoritmo de análisis de objetos binarios grandes (blob), como Grassfire, se usa para extraer objetos a partir de la imagen de entrada (operación 1908). Como se analiza anteriormente en relación con la FIG. 3, cada uno de los objetos extraídos de la imagen de entrada se procesa para identificar y descartar objetos que son demasiado pequeños o demasiado grandes para ser un dedo (operación 1910). Los objetos restantes se ordenan en una lista en orden ascendente del área general (operación 1912). El primer objeto en la lista se convierte en una imagen de $n \times n$ (operación 1914) y la misma técnica de ecualización de histograma aplicada en el proceso 1300 se aplica a la imagen redimensionada (operación 1916). Después de aplicar la técnica de ecualización del histograma a la imagen, la imagen se proyecta en el espacio propio para el conjunto de imágenes de entrenamiento y la técnica de TTC se aplica al punto de imagen para imitar la aplicación de la técnica de TTC a los puntos en la nube de puntos de entrenamiento 1402 durante la fase de entrenamiento 1300 (operación 1918).

[0103] Como se describe anteriormente, dos modelos, la superficie inferior 1702 de un cono vertical 1700 y una línea vertical 1800, se usan para modelar la nube de puntos de entrenamiento 1402. Por lo tanto, se determina si la coordenada vertical del punto de imagen está por encima o por debajo del vértice 1706 de la superficie inferior 1702 del modelo de cono vertical 1700 (operación 1920).

[0104] Si la coordenada vertical del punto de imagen está por debajo del vértice 1706 del modelo de cono vertical 1700, la rama "N" se lleva a cabo fuera de la operación 1920. Las coordenadas del punto de imagen se usan para evaluar la Ecuación 8 que define el modelo de cono vertical 1700 de la nube cónica de los puntos de entrenamiento

1404 (operación 1922). Si el punto proyectado se encuentra en la superficie del modelo de cono vertical 1700, al aplicar las coordenadas del punto proyectado a la Ecuación 8 se obtiene un valor de cero. Por el contrario, si el punto de imagen no se encuentra en la superficie del modelo de cono vertical 1700, se obtiene un número real distinto de cero aplicando las coordenadas del punto de imagen a la Ecuación 8. El valor obtenido al aplicar las coordenadas del punto de imagen a la Ecuación 8 representa la distancia entre el punto de imagen y el modelo de cono vertical 1700.

[0105] Una imagen correspondiente a un punto de imagen que se sitúa cerca de la superficie inferior 1702 del modelo de cono vertical 1700 en el espacio propio 1400 puede presentar características similares a algunas de las imágenes que comprenden el conjunto de imágenes de entrenamiento. En consecuencia, puede darse el caso de que cuanto más cerca se sitúe una imagen de la superficie inferior 1702 del modelo de cono vertical 1700, más probable es que la imagen correspondiente al punto de la imagen sea un dedo. Se puede determinar que una imagen correspondiente a un punto de imagen que se encuentra en la superficie inferior 1702 del modelo de cono vertical 1700 o un punto de imagen que se encuentra dentro de una distancia umbral máxima de la superficie inferior 1702 del modelo de cono vertical 1700 es un dedo.

[0106] Por lo tanto, el valor obtenido al aplicar las coordenadas del punto de imagen a la Ecuación 8 se evalúa para determinar si es menor que una distancia umbral definida (operación 1924). Si el valor obtenido al aplicar las coordenadas del punto de imagen a la Ecuación 8 es menor que la distancia umbral (rama "Y" fuera de la operación 1924), la imagen correspondiente al punto de imagen se considera un dedo (operación 1926). Si el valor obtenido al aplicar las coordenadas del punto de imagen a la Ecuación 8 es mayor que la distancia umbral (rama "N" fuera de la operación 1924), la imagen correspondiente al punto de imagen se descarta y se procesa el siguiente objeto en la lista procediendo a la operación 1914 (operación 1928).

[0107] Si la coordenada vertical del punto de imagen está por encima del vértice 1706 del modelo de cono vertical 1700, la rama "Y" se lleva a cabo fuera de la operación 1920. Las coordenadas del punto de imagen se usan para evaluar la ecuación que define el modelo de línea vertical 1800 (operación 1930). Si el punto de imagen se encuentra en la línea 1800, la aplicación de las coordenadas del punto de imagen a la ecuación que define la línea generará un valor de cero. Por el contrario, si el punto de imagen no se encuentra en la línea 1800, se obtendrá un número real distinto de cero aplicando las coordenadas del punto de imagen a la ecuación que define la línea 1800. El valor obtenido al aplicar las coordenadas del punto de imagen a la ecuación que define la línea 1800 representa la distancia entre el punto de imagen y la línea 1800.

[0108] Una imagen correspondiente a un punto de imagen que se localiza cerca del modelo de línea vertical 1800 en el espacio propio 1400 puede presentar características similares a algunas de las imágenes que comprenden el conjunto de imágenes de entrenamiento. En consecuencia, puede darse el caso de que cuanto más cerca se localice una imagen del modelo de línea vertical 1800, más probable es que la imagen correspondiente al punto de imagen sea un dedo. Una imagen correspondiente a un punto de imagen que se encuentra en el modelo de línea vertical 1800 o un punto de imagen que se encuentra dentro de una distancia umbral máxima del modelo de línea vertical 1800 se puede determinar como un dedo.

[0109] Por lo tanto, el valor obtenido al aplicar las coordenadas del punto de imagen a la ecuación que define la línea se evalúa para determinar si es menor que la distancia umbral (operación 1932). Si el valor obtenido al aplicar las coordenadas del punto de imagen a la ecuación que define la línea es menor que la distancia umbral (rama "Y" fuera de la operación 1932), la imagen correspondiente al punto de imagen se considera un dedo (operación 1926). Si el valor obtenido al aplicar las coordenadas del punto de imagen a la ecuación que define la línea es mayor que la distancia umbral (rama "N" fuera de la operación 1932), la imagen correspondiente al punto de imagen se descarta y el siguiente objeto en la lista puede procesarse procediendo a la operación 1914 (operación 1928).

[0110] Existen diversas técnicas para ajustar la distancia umbral. Por ejemplo, una técnica para ajustar la distancia umbral implica aplicar las coordenadas de los puntos de entrenamiento al modelo de los puntos de entrenamiento. Una distancia umbral que incluye el 90 % de los puntos de entrenamiento se considera una opción razonable para la distancia umbral en una implementación. Por supuesto, se pueden seleccionar otras distancias o porcentajes umbral.

[0111] La FIG. 20 es un diagrama de flujo de proceso que ilustra un ejemplo de un proceso 2000 para reconocer un objeto en una imagen. El proceso comienza determinando las coordenadas para un punto de imagen que caracteriza un objeto en un espacio multidimensional (operación 2002). En una implementación, una imagen de un objeto particular es capturada por una cámara y recibida como entrada por un ordenador o aplicación de software. El punto de imagen caracteriza el objeto particular capturado en la imagen.

[0112] Se proporciona una ecuación que describe un modelo geométrico en el espacio multidimensional que es característico de las imágenes de entrenamiento (operación 2004). Las coordenadas para el punto de imagen se aplican a la ecuación que describe el modelo geométrico para determinar una distancia entre el punto de imagen y el modelo geométrico (operación 2006).

[0113] En base a la distancia determinada entre el punto de imagen y el modelo geométrico, se determina si el objeto caracterizado por el punto de imagen coincide con las imágenes de entrenamiento (operación 2008). Más específicamente, se determinará que el objeto coincide o no con el tipo de objeto capturado por las imágenes de entrenamiento.

5

[0114] En una implementación, determinar las coordenadas para el punto de imagen implica proyectar la imagen en un espacio multidimensional diferente. El espacio multidimensional puede ser, por ejemplo, un espacio propio. En esta implementación, el modelo geométrico que es característico de las imágenes de entrenamiento también se encuentra en el espacio multidimensional.

10

[0115] Se han descrito varias implementaciones de la invención. Sin embargo, se entenderá que pueden hacerse diversas modificaciones. Por ejemplo, un disco compacto (CD), un dispositivo de procesamiento u otro medio legible por ordenador puede contener un programa, instrucciones o segmentos de código para implementar cualquiera de los procedimientos divulgados. Además, se puede proporcionar una herramienta para implementar cualquiera de los procedimientos divulgados. La herramienta puede incluir, por ejemplo, un medio legible por ordenador, un dispositivo de procesamiento, una cámara, un proyector o una combinación de estos y posiblemente otros componentes. Un dispositivo de procesamiento puede incluir, por ejemplo, un procesador, un ordenador, un dispositivo lógico programable o un circuito integrado.

15

20

[0116] Las implementaciones y características pueden implementarse, al menos en parte, en una variedad de dispositivos. Los ejemplos incluyen un ordenador como se describe anteriormente, que incluye un ordenador portátil u otro dispositivo de procesamiento. Los ejemplos también incluyen un teléfono portátil; un asistente digital personal; un dispositivo de mensajería tal como, por ejemplo, un buscapersonas o un dispositivo portátil de correo electrónico (tal como, por ejemplo, un Blackberry®); un reproductor de música portátil tal como, por ejemplo, un iPod®; u otro dispositivo portátil electrónico de mensajería, entretenimiento, organización o juego.

25

[0117] Además, aunque los sistemas y procedimientos divulgados en general se han descrito en el contexto del reconocimiento de un objeto en una imagen como un dedo, se contempla la capacidad de reconocer otros objetos en una imagen. Los sistemas y procedimientos descritos se pueden usar para reconocer cualquier objeto que pueda ser modelado por un modelo geométrico en un espacio, por ejemplo, un espacio propio. Por ejemplo, los sistemas y procedimientos descritos se pueden usar para reconocer un objeto en una imagen como un bolígrafo o un recipiente. Además, los elementos de diferentes implementaciones se pueden combinar, complementar, modificar o eliminar para producir otras implementaciones.

30

35

[0118] Si bien los procedimientos descritos se describieron como que incluyen múltiples operaciones, se pueden agregar operaciones adicionales a los procedimientos divulgados. Además, puede no ser necesario realizar cada operación y algunas operaciones, por lo tanto, pueden omitirse. Además, las operaciones divulgadas no necesariamente tienen que realizarse en el orden en que fueron descritas.

40

[0119] Finalmente, se pueden usar, combinar y modificar diversas tecnologías para producir una implementación, incluyendo dichas tecnologías, por ejemplo, una variedad de hardware, software, firmware, componentes integrados, componentes discretos, dispositivos de procesamiento, dispositivos de memoria o almacenamiento, dispositivos de comunicación, lentes, filtros, dispositivos de visualización y dispositivos de proyección. En consecuencia, otras implementaciones están dentro del alcance de las siguientes reivindicaciones.

45

REIVINDICACIONES

1. Un procedimiento implementado por ordenador que comprende:

- 5 detectar, dentro de una imagen, un objeto cerca de una superficie (104) usando radiación electromagnética reflejada por el objeto;
- determinar si el objeto detectado es de un tipo de objeto predeterminado; y
- 10 si se determina que el objeto detectado es del tipo de objeto predeterminado, ingresar una posición del objeto detectado en una aplicación informática,

caracterizado por que

- 15 para determinar si el objeto detectado es de un tipo de objeto predeterminado se usa un modelo que incluye una forma geométrica que representa un conjunto de puntos de entrenamiento en un espacio multidimensional, y
- 20 la determinación se basa en una comparación de una proyección del objeto detectado en el espacio multidimensional y el modelo.

2. Un procedimiento implementado por ordenador de la reivindicación 1, en el que el modelo está entrenado para determinar si el objeto detectado es del tipo de objeto predeterminado basado en una apariencia del objeto detectado.

25 3. Un procedimiento implementado por ordenador de la reivindicación 1, en el que la radiación electromagnética se proyecta sobre un primer lado de la superficie (104) y el objeto está cerca de un segundo lado de la superficie, siendo el primer lado anverso al segundo lado.

30 4. Un procedimiento implementado por ordenador de la reivindicación 1, en el que detectar el objeto cerca de la superficie comprende detectar el objeto que toca la superficie.

35 5. Un procedimiento implementado por ordenador de la reivindicación 1, en el que el modelo es característico de un conjunto de imágenes de entrenamiento de uno o más objetos del tipo de objeto predeterminado.

 6. Un procedimiento implementado por ordenador de la reivindicación 5, en el que cada uno de los puntos de entrenamiento en el conjunto corresponde a imágenes de entrenamiento del conjunto de imágenes de entrenamiento, o en el que el modelo describe un cono en el espacio multidimensional.

40 7. Un procedimiento implementado por ordenador de la reivindicación 1, en el que el objeto detectado incluye un objeto del conjunto que consiste en un dedo, un bolígrafo y un recipiente.

45 8. Un medio de almacenamiento no transitorio legible por ordenador codificado con instrucciones ejecutables que, cuando se ejecutan por un procesador, causan que el procesador realice el procedimiento de una cualquiera de las reivindicaciones 1 a 7.

9. Un aparato, que comprende:

- 50 un medio para detectar, dentro de una imagen, un objeto cerca de una superficie (104) usando radiación electromagnética reflejada por el objeto;
- un medio para determinar si el objeto detectado es de un tipo de objeto predeterminado; y
- 55 un medio para ingresar una posición del objeto detectado a una aplicación informática si se determina que el objeto detectado es del tipo de objeto predeterminado,

caracterizado por que

- 60 los medios para determinar si el objeto detectado es de un tipo de objeto predeterminado usan un modelo que incluye una forma geométrica que representa un conjunto de puntos de entrenamiento en un espacio multidimensional, y
- 65 la determinación se basa en una comparación de una proyección del objeto detectado en el espacio multidimensional y el modelo.

10. El aparato de la reivindicación 9, en el que los medios para proyectar radiación electromagnética se localizan en un primer lado de la superficie, anverso a un segundo lado de la superficie, estando localizado el objeto cerca del segundo lado de la superficie.

5 **11.** El aparato de la reivindicación 9, en el que los medios para detectar el objeto comprenden medios para detectar el objeto que toca la superficie.

12. El aparato de la reivindicación 9, en el que el modelo es característico de un conjunto de imágenes de entrenamiento de uno o más objetos del tipo de objeto predeterminado.

10

13. El aparato de la reivindicación 12, en el que el modelo describe un cono en el espacio multidimensional.

14. Un procedimiento implementado por ordenador de la reivindicación 1, en el que el espacio multidimensional tiene una dimensionalidad que es menor que el número de elementos de datos del objeto detectado.

15

15. El aparato de la reivindicación 9, en el que el espacio multidimensional tiene una dimensionalidad que es menor que el número de elementos de datos del objeto detectado.

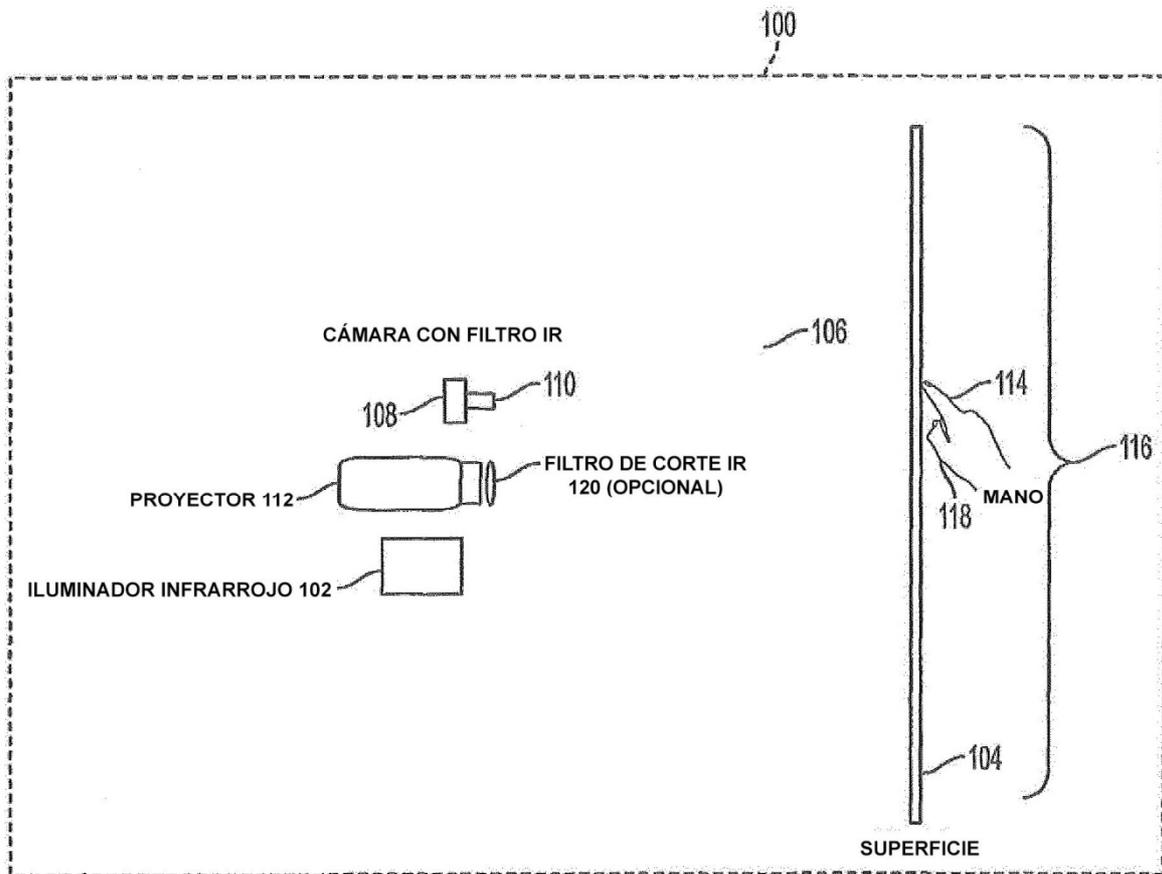


FIG. 1

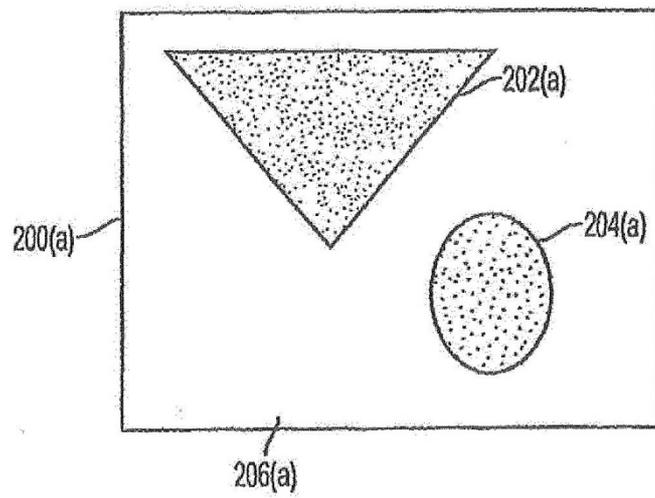


FIG. 2(a)

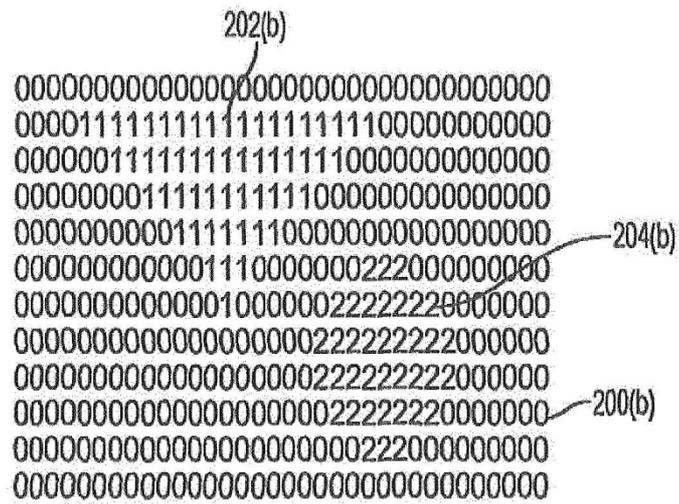


FIG. 2(b)

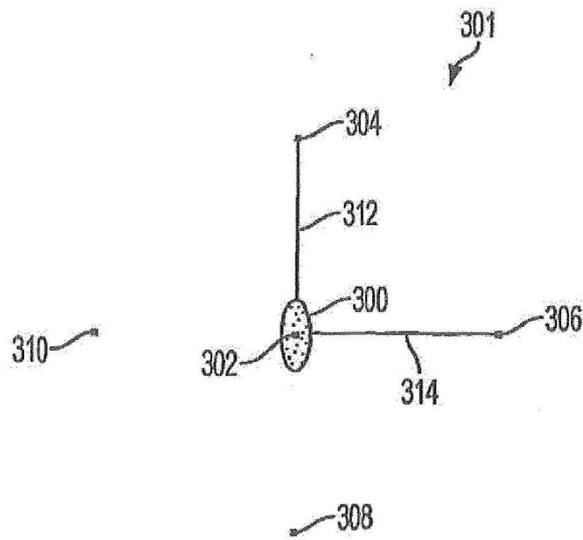


FIG. 3

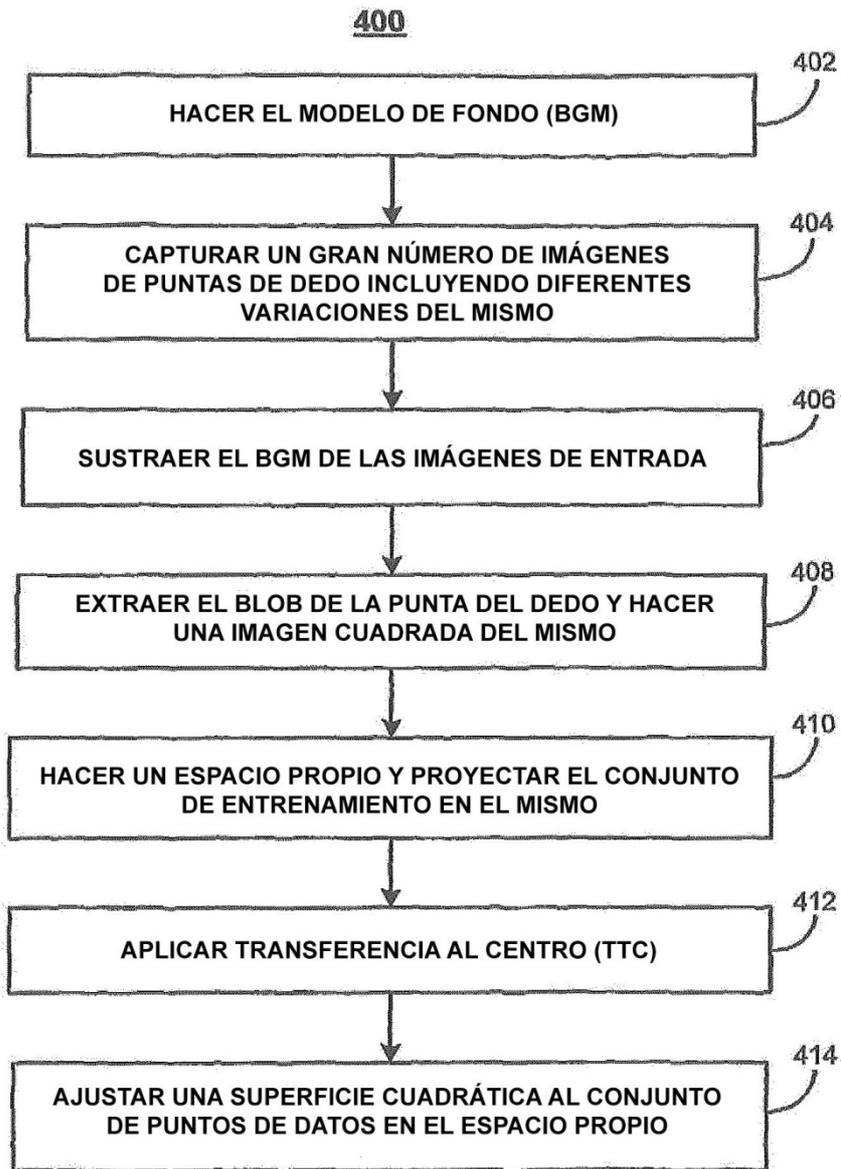
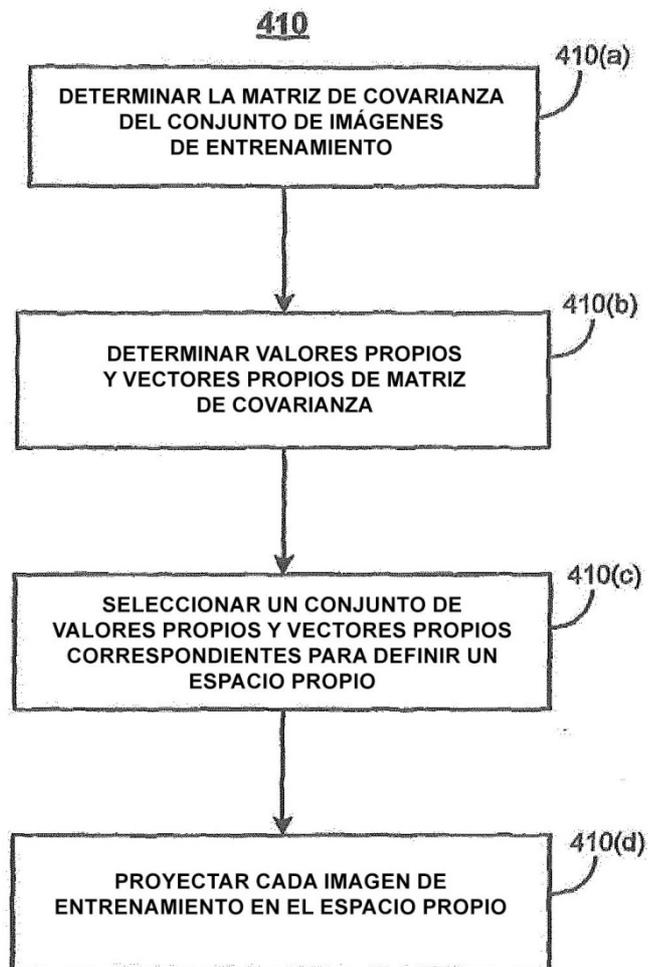


FIG. 4(a)



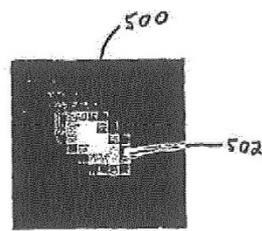


FIG. 5

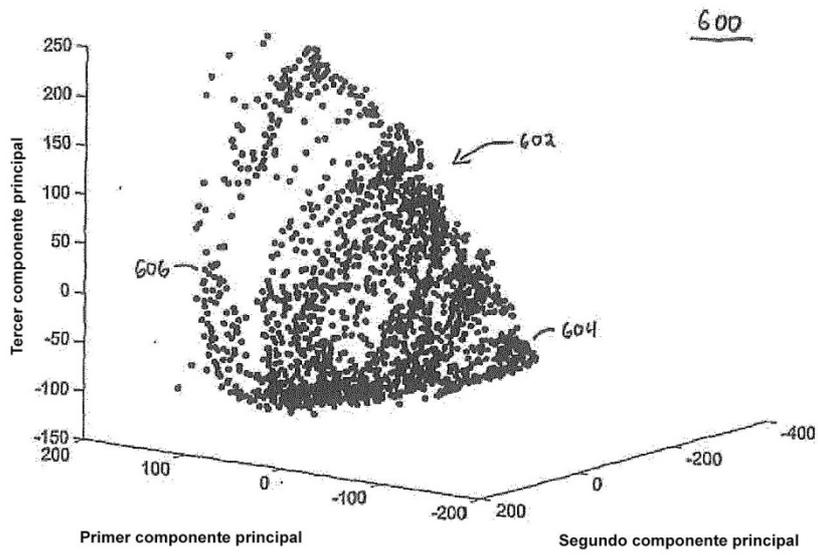


FIG. 6

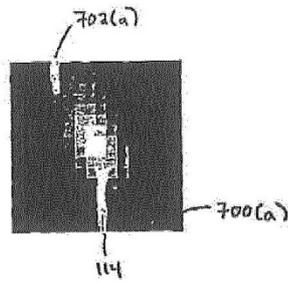


FIG. 7(a)

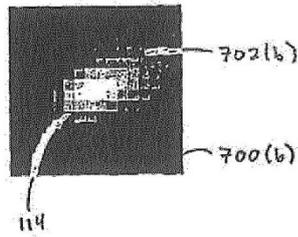


FIG. 7(b)

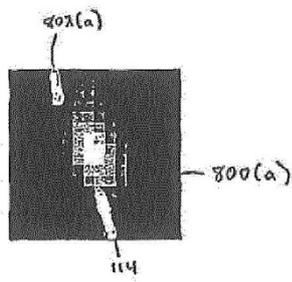


FIG. 8(a)

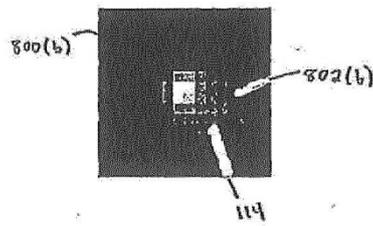


FIG. 8(b)

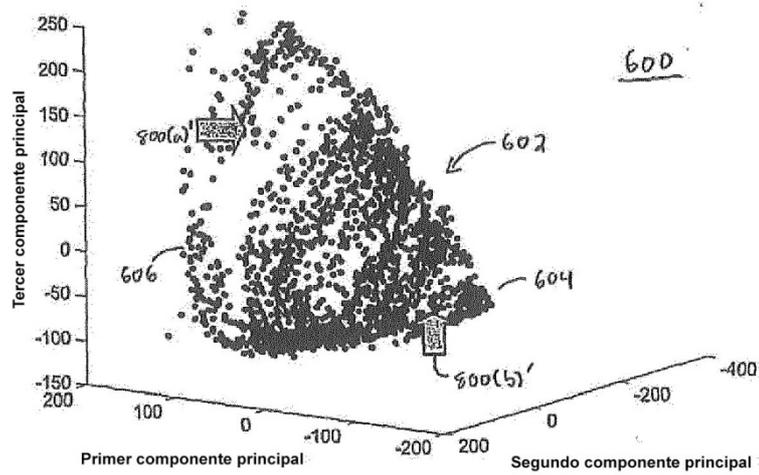


FIG. 9(a)

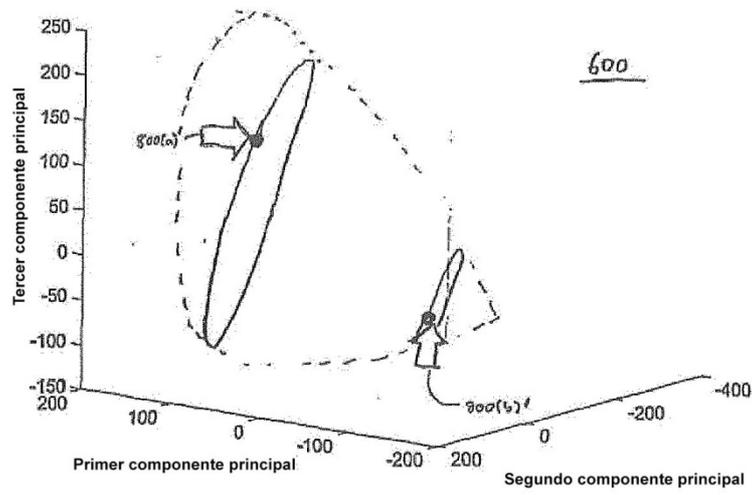


FIG. 9(b)

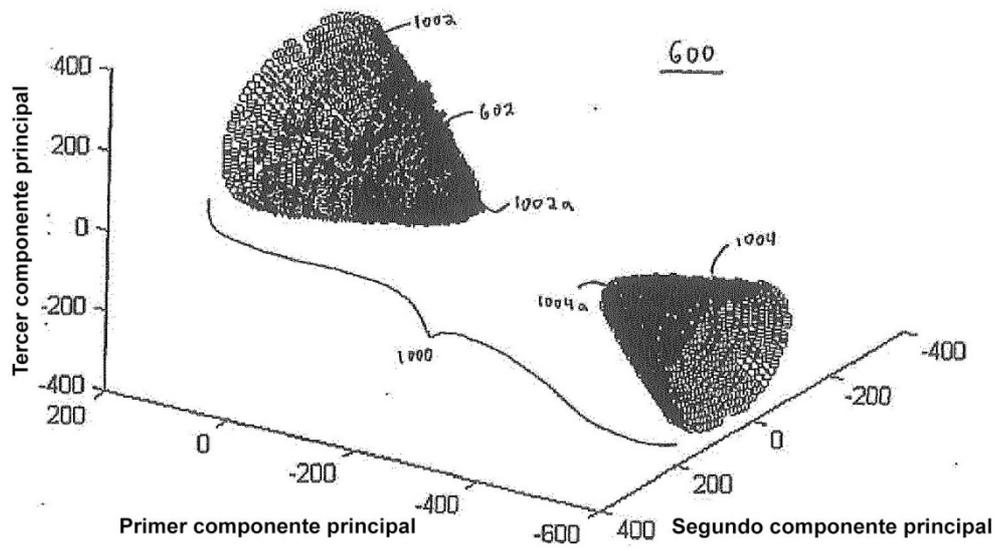


FIG. 10 (a)

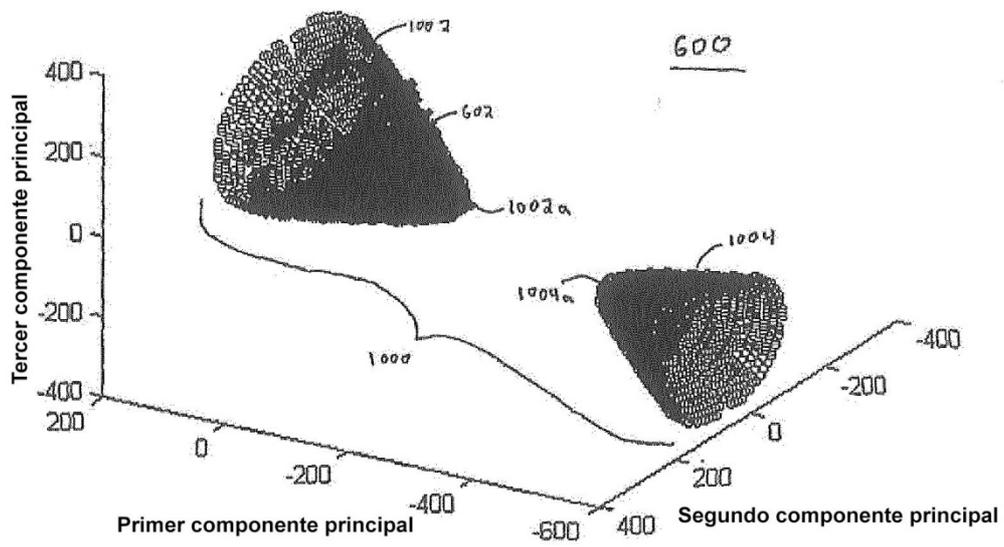


FIG. 10(b)

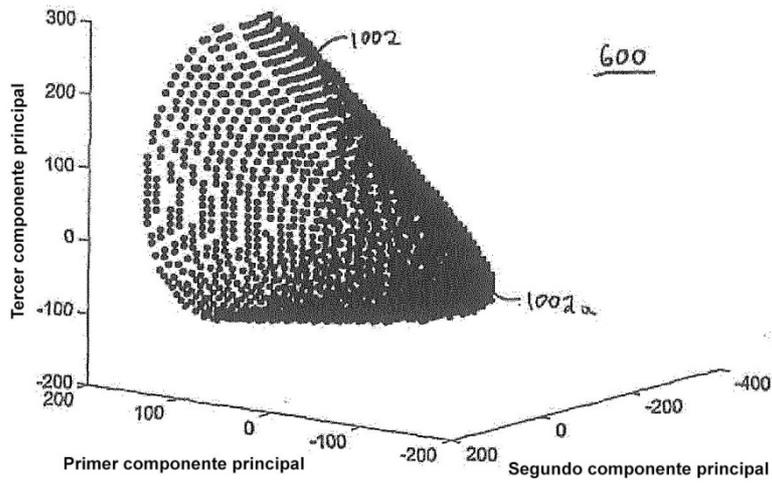


FIG. 11(a)

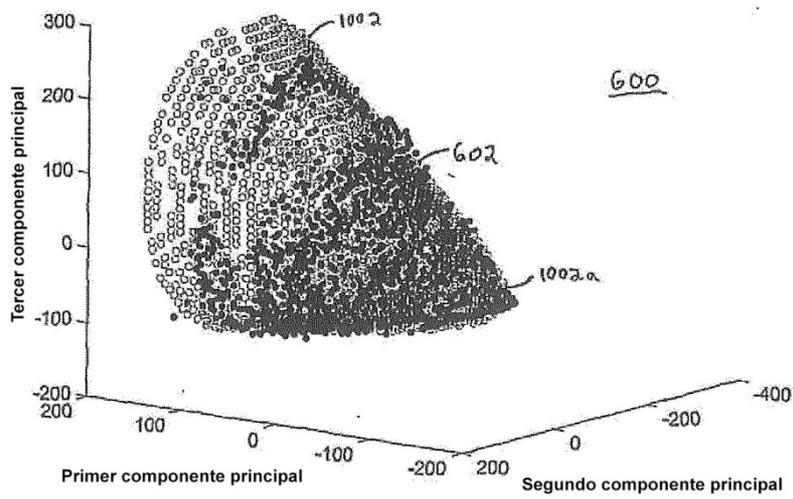


FIG. 11(b)

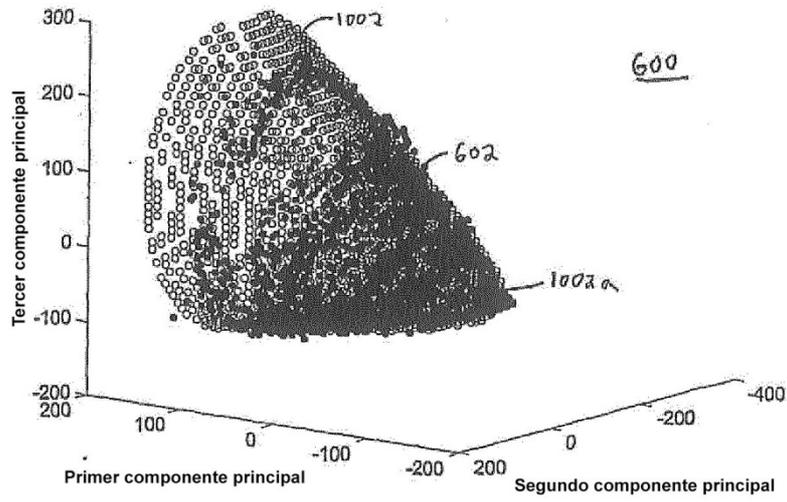


FIG. 11(c)

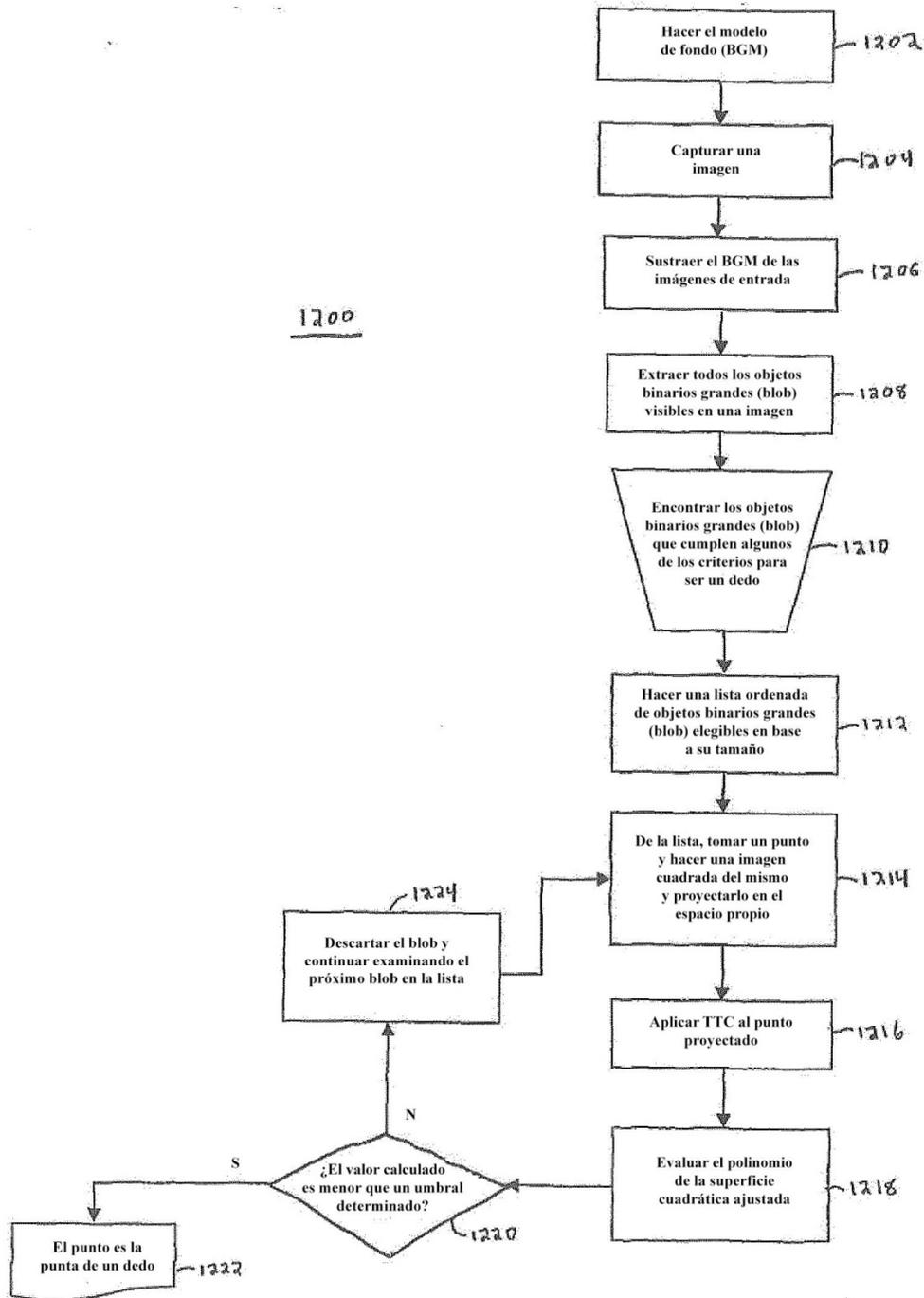


FIG. 12

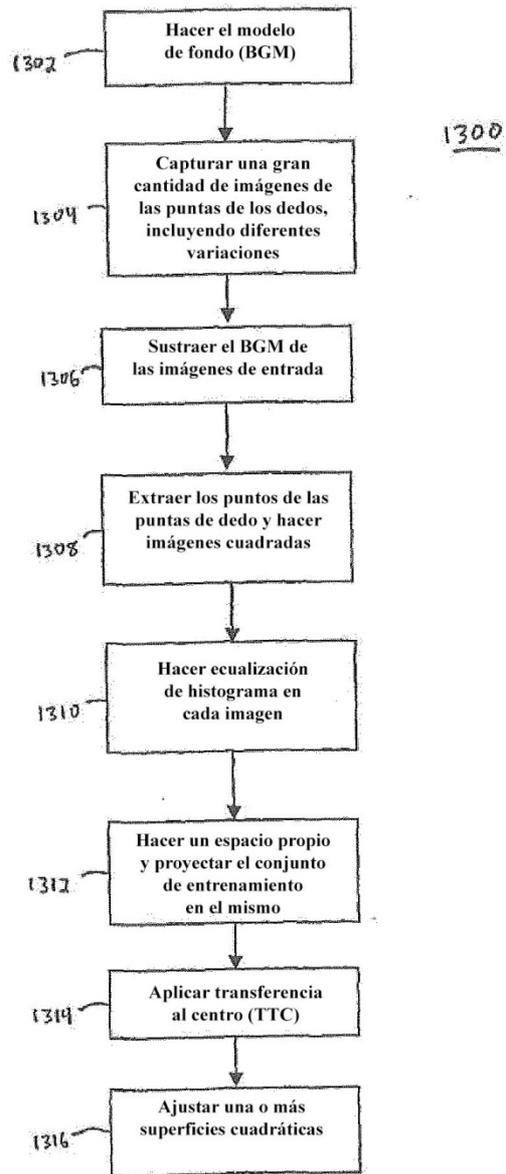


FIG. 13

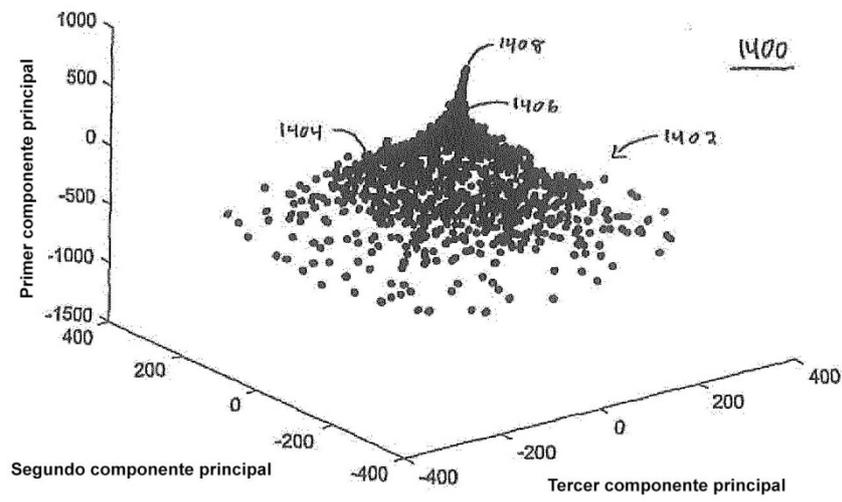


FIG. 14

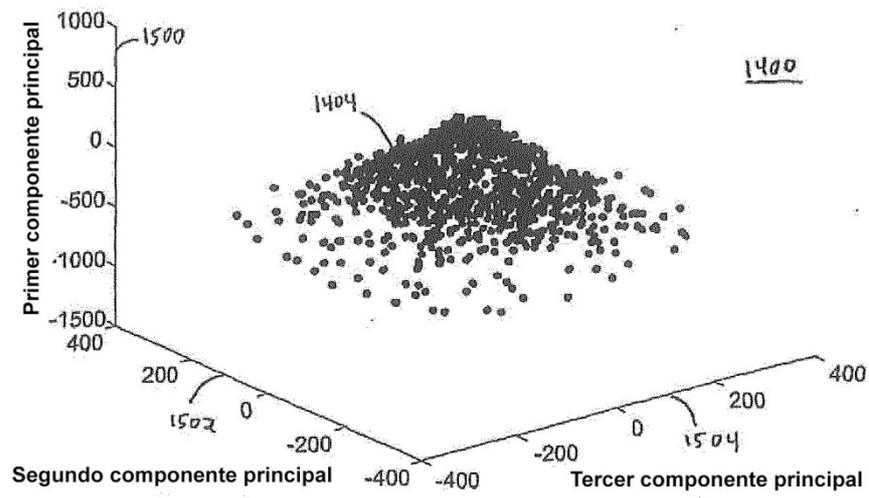


FIG. 15

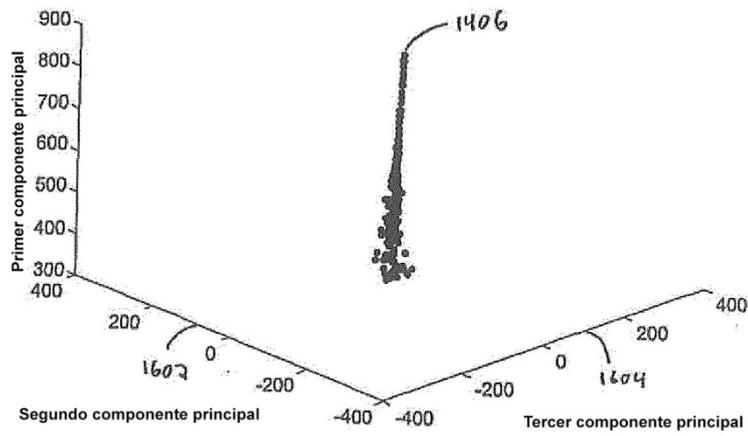


FIG. 16

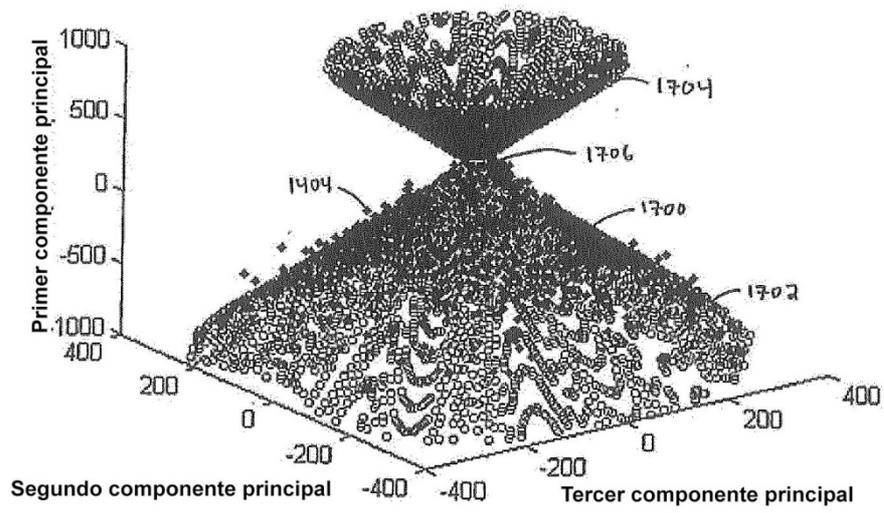


FIG. 17(a)

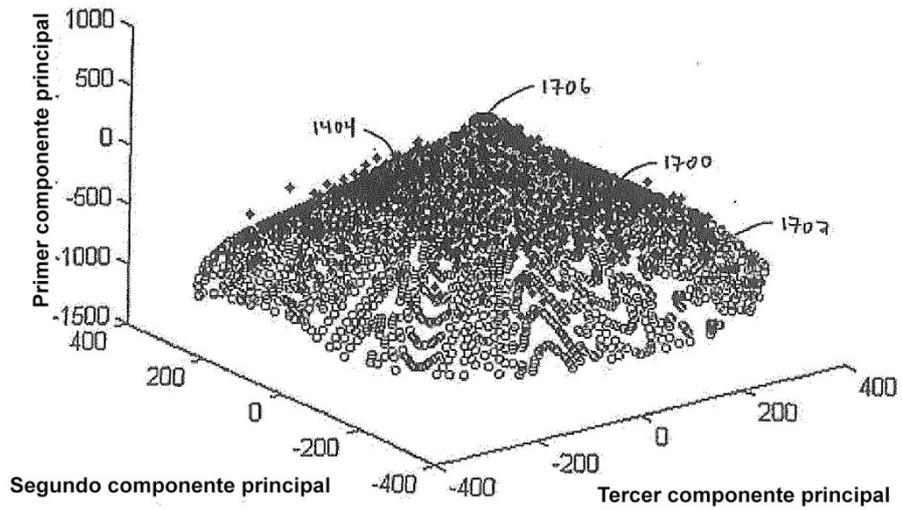


FIG. 17(b)

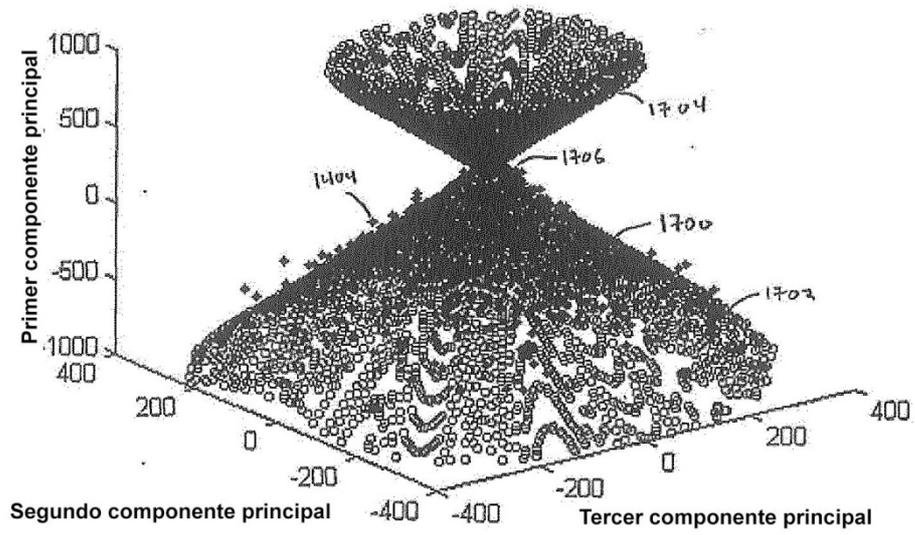


FIG. 17(c)

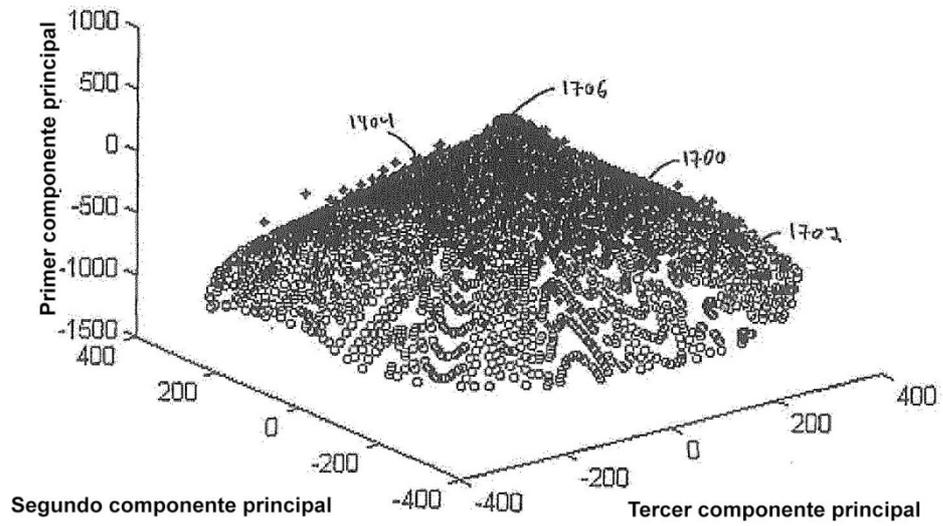


FIG. 17(d)

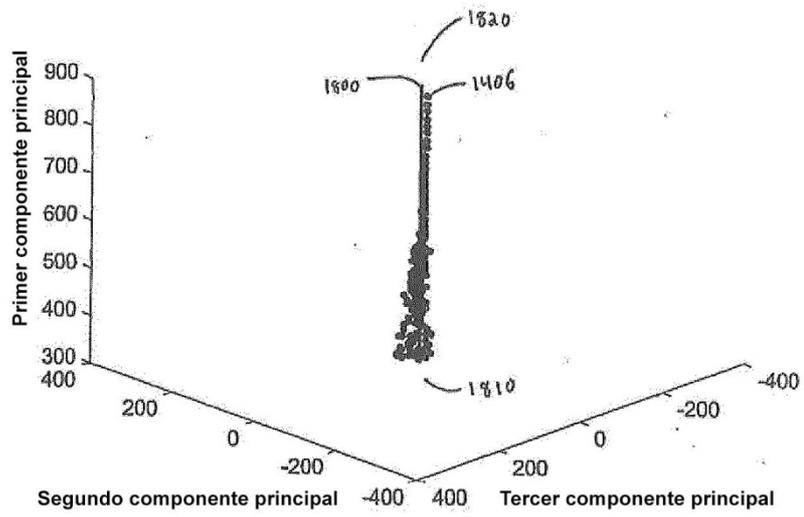


FIG. 18 (a)

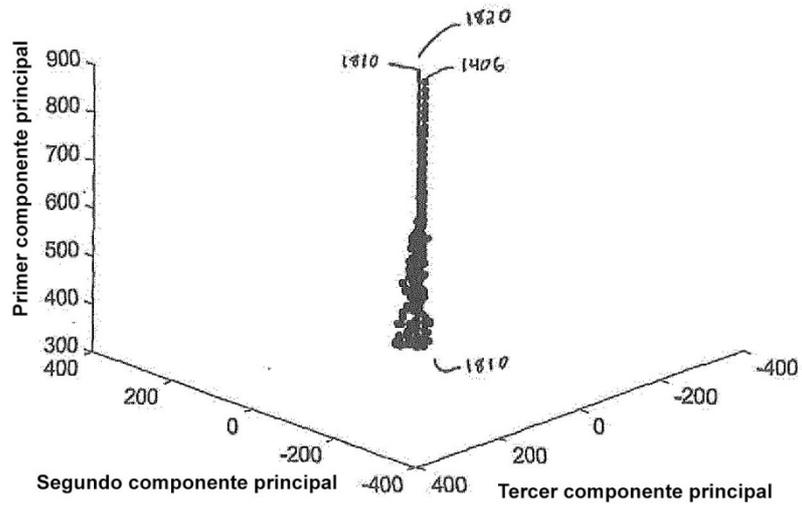


FIG. 18(b)

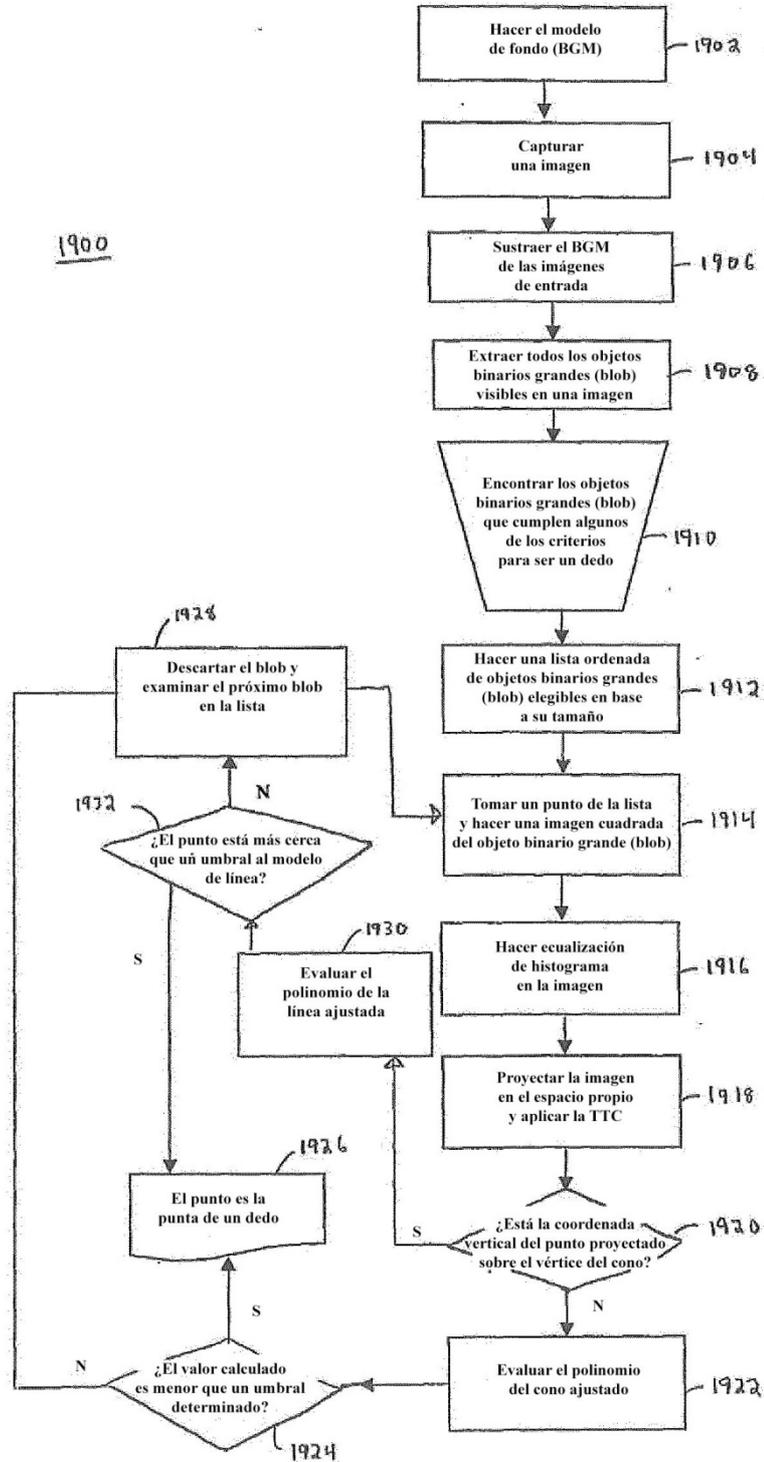


FIG. 19

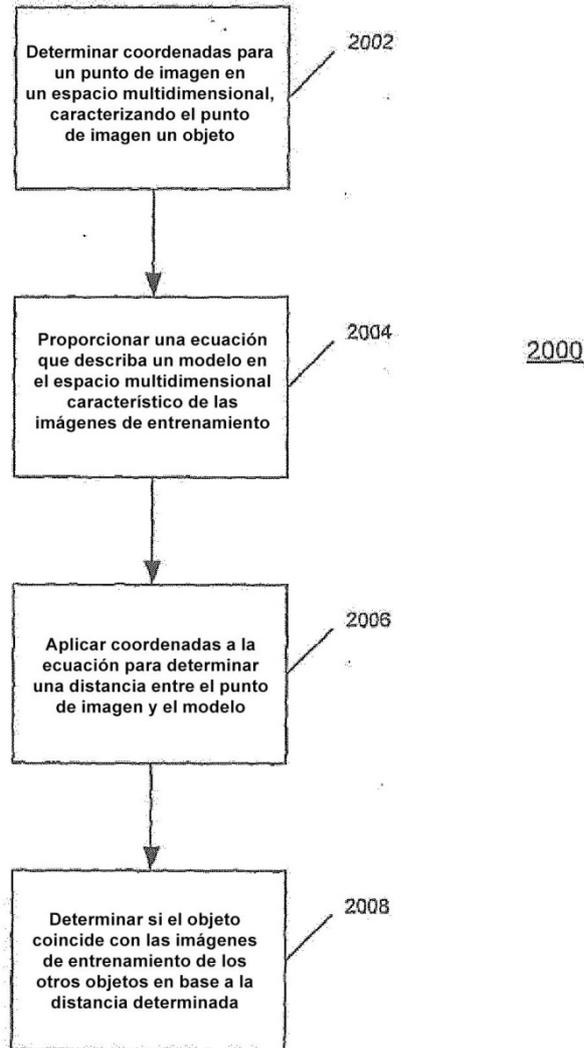


FIG. 20