

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 864 834**

51 Int. Cl.:

G06T 7/557 (2007.01)

G06T 7/593 (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.12.2016 PCT/EP2016/081966**

87 Fecha y número de publicación internacional: **26.04.2018 WO18072858**

96 Fecha de presentación y número de la solicitud europea: **20.12.2016 E 16816689 (0)**

97 Fecha y número de publicación de la concesión europea: **25.11.2020 EP 3516626**

54 Título: **Dispositivo y método para obtener información de distancia de las vistas**

30 Prioridad:

18.10.2016 WO PCT/EP2016/074992

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

14.10.2021

73 Titular/es:

**PHOTONIC SENSORS & ALGORITHMS, S.L.
(100.0%)
Calle Guardia Civil 23-2-38
46020 Valencia, ES**

72 Inventor/es:

**BLASCO CLARET, JORGE VICENTE;
MONTOLIU ALVARO, CARLES;
CALATAYUD CALATAYUD, ARNAU;
CARRIÓN, LETICIA y
MARTÍNEZ USÓ, ADOLFO**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 864 834 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

Dispositivo y método para obtener información de distancia de las vistas

Campo técnico

5 La presente invención está comprendida en el campo del procesamiento de imágenes digitales, y más particularmente a métodos y sistemas para estimar distancias y generar mapas de profundidad a partir de imágenes.

Antecedentes de la técnica

10 En la tecnología de campo de luz, los sistemas de visión multivista, tales como una cámara plenóptica o un sistema multicámara (es decir, un arreglo de varias cámaras), se utilizan con frecuencia para estimar la profundidad de las escenas. Las cámaras plenópticas son dispositivos de formación de imágenes que capturan no solo información espacial sino también información angular de una escena, conocida como campo de luz. El campo de luz se puede representar como una función de cuatro dimensiones LF (px, py, lx, ly), donde px y py seleccionan la dirección de llegada de los rayos al sensor y lx, ly son la posición espacial de ese rayo.

15 Una cámara plenóptica está formada típicamente por un arreglo de microlentes colocado delante del sensor de imagen. Este sistema de captura de imagen equivale a capturar la escena desde varios puntos de vista (las llamadas vistas plenópticas, como varias cámaras distribuidas uniformemente sobre la abertura equivalente de la cámara plenóptica). La información sobre las profundidades de los diferentes objetos (la distancia entre el objeto en sí y la cámara) en la escena se captura implícitamente en el campo de luz.

20 Se obtiene una vista plenóptica del campo de luz fijando las variables px, py a un determinado par de valores, lo que equivale a seleccionar únicamente los rayos que atraviesan una determinada parte de la abertura. Otro sistema que puede capturar un campo de luz puede estar formado por un arreglo de varias cámaras. En consecuencia, la información sobre las profundidades de los diferentes objetos (es decir, la distancia entre el objeto en sí y la cámara) de la escena se captura implícitamente en el campo de luz.

25 Un enfoque general para extraer la información de profundidad de un punto del objeto es medir el desplazamiento de la imagen de este punto del objeto sobre las varias vistas plenópticas capturadas de la escena. El desplazamiento o disparidad está directamente relacionado con la profundidad real del objeto. Para obtener la disparidad de un punto, es necesario identificar la posición del mismo punto en varias vistas (o al menos en dos vistas). Para solucionar este problema se suelen utilizar algoritmos de correspondencia entre vistas. Teniendo en cuenta un punto de una determinada vista, estos métodos analizan una región circundante e intentan encontrar la región más similar en el resto de vistas, identificando así la posición del mismo punto en el resto de vistas. Una vez obtenida la disparidad y conociendo los parámetros de la estructura del dispositivo, es posible obtener la profundidad correspondiente mediante métodos de triangulación. También es posible determinar la información de profundidad reenfocando el campo de luz a varios planos de profundidad y detectando las regiones de la imagen que están más enfocadas. El principal inconveniente de estos métodos es que son demasiado intensivos en computación para obtener mapas de profundidad en tiempo real en una plataforma móvil.

35 Otra forma de obtener la información de profundidad de una escena a partir de un campo de luz es analizar las imágenes epipolares, véase por ejemplo, BOLLES R C ET AL: "EPIPOLAR-PLANE IMAGE ANALYSIS: AN APPROACH TO DETERMINING STRUCTUREFROM MOTION", INTERNATIONAL JOURNAL OF COMPUTER VISION, DORDRECHT, NL, 1987-01-01, páginas 7-55. Una imagen epipolar es un corte bidimensional del campo de luz. Se forma una imagen epipolar horizontal fijando las variables py, ly y se forma una imagen epipolar vertical fijando las variables px, lx. Una imagen epipolar horizontal/vertical puede entenderse como una pila de la misma línea ly/lx de las diferentes vistas py/px. Suponiendo que el mismo punto del objeto es capturado por todas las vistas en una cámara plenóptica, se forman líneas correspondientes a diferentes puntos en las imágenes epipolares. El desplazamiento máximo entre vistas adyacentes en una cámara plenóptica es de ± 1 píxeles. Por tanto, los algoritmos de correspondencia se pueden evitar en este tipo de dispositivos ya que cada punto correspondiente a una determinada línea está conectado directamente al mismo punto del resto de vistas en una imagen epipolar. Sin embargo, los algoritmos actuales de cámara plenóptica como las técnicas de dominio de Fourier y las técnicas de profundidad desde el desenfoque son computacionalmente muy ineficientes ya que analizan y procesan todos los puntos de la imagen (no solo los bordes, como en la presente invención). Por otro lado, los métodos simples de gradiente de campo de luz (en las direcciones horizontal y vertical) producen mapas de profundidad muy deficientes, con estimaciones de profundidad poco fiables. Además, estas implementaciones no pueden lidiar con imágenes de video en tiempo real, tomando de cientos de milisegundos a minutos solo para procesar un solo cuadro.

45 Por lo tanto, existe la necesidad de un método extremadamente eficiente que permita a las cámaras plenópticas e imágenes 3D en dispositivos móviles (tales como teléfonos móviles, tabletas u ordenadores portátiles) calcular mapas de profundidad y procesar imágenes de video en tiempo real (por ejemplo, 60 cuadros por segundo).

55 Durante los últimos quince años, la formación de imágenes multivista ha aparecido con mayor frecuencia en la literatura científica, en varios campos de investigación tales como el desvanecimiento de imágenes, la síntesis de vistas virtuales o la reconstrucción de imágenes de alta resolución, solo por nombrar algunos. Una de las principales

limitaciones de utilizar una sola cámara plenóptica es que la resolución espacial se reduce drásticamente para igualar el número de microlentes; por lo tanto, la mayoría de las publicaciones solo consideran mejorar la resolución espacial de dichas cámaras plenópticas mediante técnicas de superresolución, sin considerar mejorar la precisión y rango de las estimaciones de profundidad. Estos enfoques han demostrado ser efectivos para aumentar la resolución espacial de las cámaras plenópticas en un factor de 4X, sin embargo, más allá de 4X su rendimiento cae drásticamente.

Las estimaciones de mapas de profundidad que utilizan cámaras plenópticas son generalmente efectivas cuando la estimación se realiza en un rango de profundidad limitado muy cerca de la cámara. Sin embargo, esta estimación es cada vez más inexacta a medida que aumenta la distancia entre la cámara y el mundo de los objetos.

La visión estéreo es otro enfoque para obtener mapas de profundidad en una escena. Mediante técnicas de triangulación, es posible extraer información 3D de una escena mediante dos miradores, imitando la percepción visual humana. Hay muchos algoritmos estéreo que pueden producir mapas de profundidad utilizando dos cámaras con desplazamiento espacial conocido. Dado que la línea de base de los dispositivos de visión estéreo suele ser más amplia que la línea de base de las cámaras plenópticas, los enfoques de visión estéreo pueden estimar mejor los mapas de profundidad para largas distancias. Sin embargo, estos enfoques estéreo binoculares adolecen de varias desventajas, ya que a menudo dan como resultado mapas de disparidad incompletos (agujeros producidos por oclusiones donde no es posible encontrar el mismo punto de objeto en ambas imágenes) o tienen regiones de discontinuidad de profundidad donde las disparidades entre píxeles vecinos han experimentado brechas de más de un píxel (en visión estéreo, cuando se estima un mapa de profundidad, se acumulan inexactitudes en el cálculo de disparidades entre los puntos correspondientes a nivel de subpíxel; en algún punto, estas inexactitudes pueden ser mayores que un píxel, causando una brecha entre dos puntos consecutivos y dejando un punto sin estimación de profundidad). Además, los enfoques estéreo son altamente costosos desde el punto de vista computacional ya que generalmente requieren algoritmos de correspondencia intensivos en computación.

Otro problema que afecta a las cámaras estéreo es la profundidad de campo relativamente pequeña de las cámaras convencionales, ya que este tipo de sistemas pueden estimar las profundidades de manera adecuada solo en el rango en el que ambas cámaras están enfocadas. Con las tecnologías CMOS modernas, los píxeles se han reducido a dimensiones tan pequeñas como una micra y pronto estarán por debajo de una micra. Es bien sabido que a medida que los píxeles de los fotosensores se vuelven más pequeños, la profundidad de campo en el mundo de los objetos (profundidad de enfoque en el mundo de la imagen) se deteriora, por lo tanto, el rango de distancias del mundo real que están enfocadas se vuelve más corto y más corto a medida que los píxeles se vuelven cada vez más pequeños. Sería posible revertir esa tendencia utilizando aberturas más pequeñas, pero a costa de recibir menos luz y, por lo tanto, disminuir el número de cuadros por segundo que se pueden grabar. Por esta razón, las minicámaras utilizadas en telefonía móvil con una gran cantidad de píxeles (10-20 megapíxeles o más) y tamaños de píxeles pequeños (alrededor de una micra) están comenzando a usar soluciones de "enfoque automático" que se implementan principalmente con MEMS (Sistemas micro electromecánicos), elementos móviles que mueven los lentes hacia adelante y hacia atrás a lo largo del eje óptico para enfocar la imagen.

Si un par estéreo usa autoenfoque, ambas cámaras estarán enfocadas, pero la información de las áreas fuera de foco definitivamente se ha hecho borrosa o perdido (mezclando sobre el sensor o la información de la película de diferentes áreas y profundidades del mundo del objeto). De ahí que el proceso estéreo, es decir, la triangulación para conocer la distancia del mismo patrón en ambas cámaras al mundo real, no mejorará la borrosidad en las áreas fuera de foco, contaminando los cálculos de distancia que eventualmente no ofrecerán más datos confiables. Se pueden pensar diferentes soluciones para abordar este problema, por ejemplo, tener una de las dos cámaras enfocada en distancias cortas y la otra enfocada en distancias largas. Sin embargo, esta solución empeora las soluciones de triangulación, al tener que identificar el mismo patrón en áreas borrosas en una de las cámaras y no borrosas en la segunda cámara, lo que aumenta la dificultad e impacta en la confiabilidad de los algoritmos de correspondencia.

Otra posible solución pero mucho más sofisticada es utilizar lentes especiales que dependen del color, de modo que los 3 colores diferentes del patrón de Bayer (o cualquier otro patrón de color fundamental) se enfoquen en tres rangos diferentes para distancias cortas, medias y largas, combinando el resultado posteriormente para obtener lo que se ha denominado EDOF (profundidad de campo extendida). Aunque EDOF se ha aplicado a una sola cámara, potencialmente se puede extender a las dos cámaras de un par estéreo. También se pueden utilizar diferentes permutaciones de colores y posiciones de enfoque en las dos cámaras del par estéreo.

Cualquiera que sea el enfoque que se utilice, finalmente es necesario enfocar ambas cámaras (o colores) alrededor del mismo rango de profundidades en el mundo del objeto (en cuyo caso, información de las áreas fuera de foco en ambas cámaras [o colores] ya no se pueden usar para calcular profundidades) o mezclar imágenes borrosas y no borrosas en el proceso de triangulación, produciendo resultados subóptimos.

Otra posible solución para extender el rango de profundidad, donde se pueden usar enfoques estéreo para estimar profundidades, sería diseñar las cámaras con aberturas extremadamente pequeñas y píxeles relativamente grandes, extendiendo la profundidad de campo desde distancias muy pequeñas (unos pocos centímetros) hasta el infinito, y hacer lo mismo para ambas cámaras en el par estéreo. Sin embargo, esa compensación no es gratuita. En principio, sería posible revertir la tendencia explicada anteriormente con aberturas más pequeñas, pero eso a costa de recibir menos luz y por ende disminuir el número de cuadros por segundo que se pueden grabar (inaceptable en aplicaciones

de video). Finalmente, sería posible agrandar los píxeles, en contra de la tendencia actual de tener un mayor número de megapíxeles con píxeles más pequeños, pero eso daría como resultado sensores extremadamente grandes inapropiados para aplicaciones portátiles y permitidos solo en cámaras profesionales grandes.

5 Como se indicó anteriormente, las cámaras plenópticas se pueden usar para estimar las profundidades de una escena analizando las imágenes epipolares. Las cámaras plenópticas tienen la ventaja de tener una profundidad de campo mucho mayor ya que la abertura se divide efectivamente en varias aberturas pequeñas (generalmente cientos), aumentando drásticamente la profundidad de campo. La profundidad de campo de una cámara plenóptica puede ser prácticamente desde unos pocos centímetros hasta una distancia infinita, lo que hace que estos dispositivos sean mucho más atractivos para grandes profundidades de campo que los enfoques estéreo. En las cámaras plenópticas 10 incluso es posible evitar el requisito de tener MEMS para variar el enfoque de la cámara.

La invención propuesta permite que las cámaras plenópticas calculen mapas de profundidad de una manera extremadamente eficiente, permitiendo el procesamiento de imágenes de video en tiempo real a una alta tasa de cuadros (60 cuadros por segundo o más). Además, la presente invención también aprovecha el sistema multivista para mejorar significativamente la precisión de la estimación de profundidad de las cámaras plenópticas a grandes 15 distancias de la cámara, pudiendo aún disfrutar de técnicas de superresolución y mejoras de resolución lateral, reenfoque y técnicas tradicionales de estimación de profundidad. El procedimiento divulgado en este documento mejora los enfoques más avanzados en términos de eficiencia computacional y requisitos de potencia.

Resumen de la invención

20 La presente invención se refiere a un método implementado por ordenador y un dispositivo que obtiene un mapa de profundidad procesando la imagen de campo de luz capturada por una cámara plenóptica o cualquier otro dispositivo de adquisición de campo de luz, dispositivos de muestreo de función plenóptica o dispositivos de adquisición de imágenes integrales. Se pueden usar otras cámaras en combinación con una cámara plenóptica, tal como una o más cámaras convencionales o cámaras plenópticas adicionales, formando un sistema multivista.

Pueden usarse cámaras plenópticas para estimar las profundidades de una escena analizando las imágenes epipolares. Existe una relación entre la pendiente de las líneas epipolares producidas en las imágenes epipolares en una cámara plenóptica y la profundidad real de un objeto en una escena (en el mundo de los objetos). Por tanto, al 25 detectar la pendiente de las líneas de una imagen epipolar, es posible generar un mapa de profundidad de la escena. El método es muy eficiente computacionalmente, ya que los cálculos se pueden realizar solo para aquellas partes del sensor donde se han encontrado bordes en la escena, evitando así cálculos en regiones del mundo de objetos donde no se detectaron bordes. De esta manera, el método se puede utilizar para obtener mapas de profundidad en tiempo real incluso en dispositivos móviles de bajo coste con procesadores de bajo coste operados por baterías, donde se necesitan cómputos eficientes para evitar agotar las baterías rápidamente. 30

La presente invención utiliza un algoritmo extremadamente eficiente que permite imágenes 3D en cámaras plenópticas, dispositivos móviles (teléfonos móviles, tabletas, ordenadores portátiles, cámaras compactas, etc.), 35 dispositivos de entrada con detección de movimiento y cámaras 3D que procesan imágenes de video en tiempo real (a 60 cuadros por segundo e incluso más) identificando los bordes del objeto y calculando la profundidad solo para los bordes identificados.

Existe una relación entre la pendiente de las líneas producidas en las imágenes epipolares y la profundidad real del objeto en la escena. Por tanto, al detectar la pendiente de las líneas de una imagen epipolar, es posible generar un 40 mapa de profundidad de la escena. Por lo general, se utilizan métodos con base en un gradiente bidimensional de las imágenes epipolares para obtener la pendiente correspondiente. También se pueden emplear métodos similares con base en gradientes de cuatro dimensiones (y, por lo tanto, más costosos desde el punto de vista computacional). A diferencia de todos estos enfoques, el método actual calcula la profundidad de la escena solo para los bordes, lo que reduce drásticamente los requisitos de cómputo.

45 La fotografía de campo de luz captura implícitamente la geometría de la escena 3D y las propiedades de reflectancia en un campo de luz. Un campo de luz es una estructura de cuatro dimensiones donde los rayos de luz incidentes se describen por medio de su posición espacial (2D: lx y ly) y por sus direcciones de llegada (2D: px y py). En la presente invención, un campo de luz 4D (px , py , lx , ly) se considera como la salida de una cámara plenóptica. Estos dispositivos son cada vez más populares debido a su potencial aplicación para estimar el mapa de profundidad de una escena. Si los colores también son capturados por el sensor (por ejemplo, usando los llamados patrones de Bayer o similares), 50 el campo de luz sería una estructura 5D (px , py , lx , ly , c) donde c son los diferentes canales de color capturados. Para mayor claridad y simplicidad, en la presente invención se supone que el campo de luz es una estructura 4D sin información de color. Sin embargo, un experto en el campo comprenderá que la extensión de la información divulgada para los sensores que capturan información de color es trivial y sencilla. Una posible solución sería aplicar los algoritmos presentados en este documento a cada canal de color por separado para aumentar la redundancia de las 55 estimaciones de profundidad.

La estimación de la profundidad a partir del campo de luz está cada vez más extendida en aplicaciones de campo de luz, especialmente en aplicaciones de formación de imágenes en 3D. Sin embargo, para obtener una reconstrucción

3D o un mapa de profundidad de una escena, los datos contenidos en el campo de luz necesitan un posprocesamiento adicional que transforme la estructura del campo de luz 4D de entrada en una imagen 2D donde para cada píxel capturado por la cámara plenóptica es posible calcular su profundidad en el mundo de los objetos reales. Básicamente, en la formación de imágenes plenópticas, los objetos a diferentes distancias de la cámara producen diferentes patrones de iluminación en el sensor de una cámara plenóptica y, por lo tanto, se puede llevar a cabo un procesamiento apropiado de estos patrones para obtener la distancia correspondiente, es decir, la profundidad a la que estos objetos están en el mundo de los objetos. Los principales inconvenientes de los sistemas de formación de imágenes plenópticas son la pérdida de resolución espacial y el hecho de que su precisión de estimación de profundidad disminuye muy rápidamente a medida que aumenta la distancia a la cámara.

Otra metodología bien conocida para estimar la profundidad del objeto en una escena es mediante visión estereó. Al rastrear el desplazamiento de los puntos de la imagen entre diferentes puntos de vista de la misma escena, es posible estimar la distancia de los objetos de una escena usando triangulación básica. La visión estereó tiene como objetivo identificar los puntos correspondientes del mundo de los objetos grabados (o vistos) desde dos puntos de vista diferentes (dos cámaras diferentes separadas entre sí), calculando su desplazamiento para reconstruir la geometría de la escena como un mapa de profundidad.

De acuerdo con una realización de la presente invención, el sistema y el método de procesamiento descritos en este documento se implementan como un sistema de vista múltiple que incluye (pero no se limita a) al menos una cámara plenóptica de campo de luz y una o más cámaras adicionales (cámaras convencionales y/o cámaras plenópticas). Esta invención crea un mapa de profundidad de alta calidad de una escena con mayor precisión y para distancias más grandes que la técnica anterior. La presente invención permite mejorar las líneas epipolares de las cámaras plenópticas con datos adicionales de una cámara convencional alineada horizontalmente (mejora de la línea epipolar horizontal); sin embargo, esto no limita la generalidad de la invención, que puede incluir múltiples cámaras (siempre que al menos una de ellas sea una cámara plenóptica) y cualquier alineación entre ellas.

Considerando una realización con una sola cámara plenóptica con N' vistas plenópticas (o, de manera equivalente, N' píxeles debajo de cada microlente) y $(M \times N) - 1$ cámaras convencionales dentro de un arreglo de cámaras $M \times N$, la presente invención proporciona las siguientes ventajas principales:

- La invención mejora la eficiencia computacional de las metodologías de última generación, ya que no es necesario elaborar correspondencias de pares estereoscópicos (muy intensivos en computación) entre los puntos registrados por cada cámara, ya que las líneas epipolares formadas en las imágenes epipolares plenópticas se utilizan para encontrar los puntos correspondientes.

- El método de la presente invención no es computacionalmente exigente; además, la invención emplea tareas paralelizables que pueden disfrutar de los beneficios de las modernas plataformas de computación paralela.

- La invención puede ser utilizada en cualquier tipo de dispositivos móviles operados por baterías debido a sus bajos requerimientos de potencia computacional. Esto, acoplado al avance de las minicámaras miniaturizadas, es especialmente útil para la nueva generación de dispositivos móviles. Tener dos cámaras se está volviendo común y la vista múltiple pronto también lo será.

- La invención ofrece mucha más redundancia y tolerancia al ruido ya que la imagen está formada por imágenes $N' + [(M \times N) - 1]$, sumando las imágenes capturadas por las cámaras convencionales $(M \times N) - 1$ como vistas plenópticas adicionales. La presente invención tiene $N' - 1$ imágenes más que un sistema multivista convencional de cámaras convencionales $M \times N$ y $(M \times N) - 1$ vistas plenópticas más que una cámara plenóptica convencional, y al tener estas imágenes (o vistas) adicionales es posible tener líneas de base mucho más amplias que la cámara plenóptica.

- Debido a la pequeña línea de base y al alto número de vistas capturadas por la cámara plenóptica, el efecto de las oclusiones es casi insignificante. De esta forma, se superan los inconvenientes de los sistemas de visión estereó y multivista tradicionales en cuanto a la incompletitud y las discontinuidades producidas en la estimación de profundidad debido a oclusiones en el mundo de los objetos.

- La invención divulgada utiliza la interpolación (o cualquier otro método para establecer una correspondencia entre dos imágenes de diferente resolución del mismo escenario mundial) para mejorar la resolución de la cámara plenóptica, manteniéndose al día con la resolución de las cámaras convencionales. Por lo tanto, la resolución de cada vista plenóptica aumenta significativamente.

- La invención supera los inconvenientes de un sistema basado en una cámara plenóptica con respecto a las inexactitudes producidas en la estimación de profundidad a grandes distancias de la cámara. Esta mejora se logra mediante el uso de varios enfoques diferentes posibles:

- Combinar el mapa de profundidad de la cámara plenóptica y las imágenes 2D de las cámaras convencionales $(M \times N) - 1$.

○ Combinar cada una de las vistas plenópticas N' de la cámara plenóptica y las imágenes 2D de las cámaras convencionales $(M \times N) - 1$ como configuraciones multiestéreo N' veces, considerando como multiestéreo cada una de las $(M \times N)$ cámaras del arreglo. Esto puede incluir el uso de algoritmos de correspondencia tradicionales.

5 ○ Ampliar las imágenes epipolares de la cámara plenóptica con las imágenes 2D de las cámaras convencionales $(M \times N) - 1$.

• La invención divulgada mejora los mapas de profundidad en zonas particularmente difíciles de la imagen (debido, por ejemplo, a oclusiones) mediante la aplicación de algoritmos estéreo clásicos. Suponiendo que se crea un primer mapa de profundidad mediante alguna de las combinaciones descritas anteriormente, en al menos una realización se puede realizar un posible refinamiento de este primer mapa de profundidad aplicando algoritmos estereofónicos clásicos, resolviendo posibles ambigüedades del primer mapa de profundidad obtenido en esas zonas difíciles de la imagen.

• La invención mejora la capacidad de reenfoque que se puede lograr usando solo un par estéreo o multivista.

En la discusión anterior, se supone que solo hay una cámara plenóptica en el arreglo de cámaras $M \times N$, pero la generalización para tener más de una cámara plenóptica es sencilla. Como se explicará más adelante, al tener más de una cámara plenóptica se mejora la redundancia en las medidas para calcular distancias y en el proceso de formación de imágenes, y se mejora la inmunidad al ruido; además, la eficiencia computacional se mejora cuando se usa la información de las cámaras convencionales.

Esta divulgación se refiere a la tecnología de campo de luz y los sistemas de visión multivista para estimar las profundidades de las escenas. También se divulga en este documento un procedimiento de procesamiento de imágenes para producir un mapa de profundidad de una escena estimando la pendiente de líneas epipolares extendidas.

De acuerdo con una realización, la presente invención se refiere a un dispositivo y método para la estimación de la profundidad en tiempo real utilizando un sistema de formación de imágenes de vista múltiple. El sistema comprende al menos una cámara plenóptica de campo de luz y también puede incluir cámaras convencionales adicionales. Un sistema de vista múltiple de este tipo, con los procedimientos de procesamiento de imágenes adecuados, es capaz de crear un mapa de profundidad de la escena con una resolución de muy alta calidad, superando los inconvenientes de las cámaras plenópticas y los sistemas multicámara actuales. La presente invención logra una mayor precisión en las mediciones de profundidad y en la profundidad máxima medible, y al mismo tiempo también proporciona ventajas adicionales tales como transiciones más suaves entre las diferentes profundidades capturadas y representadas en el mapa de profundidad así como una mejor percepción al ojo humano, y también mejora la capacidad de reenfoque digitalmente la imagen después de que se ha capturado la escena y de aplicar efectos artísticos.

El método de la presente invención es extremadamente eficiente en términos de requisitos computacionales, y se puede usar en cualquier tipo de dispositivos móviles operados por baterías debido a sus bajos requerimientos de potencia computacional. El método descrito en este documento también se puede paralelizar eficazmente en varios procesadores y/o GPU, así como en procesadores paralelos específicos para dispositivos móviles que funcionan con baterías.

Para la descripción de la presente invención se considerarán a continuación las siguientes definiciones y acrónimos:

• Arreglo de microlentes: una pluralidad de lentillas (microlentes) dispuestas en un arreglo.

• Arreglo de microlentes regular: arreglo formado por microlentes que han sido diseñados para estar espaciados y contruidos regularmente (paso homogéneo en toda la estructura del arreglo, mismo radio de curvatura para todos los lentes, misma distancia focal, etc.), sin tomar en cuenta la falta de homogeneidad debida a imperfecciones de fabricación.

• Lentilla o microlente: cada lente pequeño forma un arreglo de microlentes.

• Cámara plenóptica: dispositivo que captura no solo la posición espacial sino también la dirección de llegada de los rayos de luz incidentes.

• Cámara convencional: dispositivo que captura únicamente la posición espacial de los rayos de luz incidentes al sensor de imagen, de manera que cada píxel del sensor integra toda la luz que llega en cualquier dirección desde toda la abertura del dispositivo.

• Campo de luz: estructura de cuatro dimensiones $LF(px, py, lx, ly)$ que contiene información de la luz captada por los píxeles (px, py) debajo de los microlentes (lx, ly) en una cámara plenóptica.

• Vista plenóptica: imagen bidimensional formada al tomar un subconjunto de la estructura del campo de luz eligiendo un determinado valor (px, py) , el mismo (px, py) para cada una de los microlentes (lx, ly) .

- Profundidad: distancia entre el plano de un objeto puntual de una escena y el plano principal de la cámara, ambos planos son perpendiculares al eje óptico.
 - Mapa de profundidad: imagen bidimensional en la que se añaden los valores de profundidad calculados (dz) del mundo del objeto como valor de dimensión adicional a cada píxel de la imagen bidimensional, componiendo (dx, dy, dz).
 - Mapa de disparidad: diferencia en la posición de la imagen del mismo conjunto de puntos 3D en el mundo de los objetos cuando se captura con dos cámaras desde dos perspectivas diferentes. La disparidad se puede utilizar para determinar la profundidad mediante triangulación.
 - Imagen epipolar: corte bidimensional de la estructura del campo de luz compuesto al elegir un valor determinado de (px, lx) (imagen epipolar vertical) o (py, ly) (imagen epipolar horizontal).
 - Línea epipolar: conjunto de píxeles conectados dentro de una imagen epipolar detectada como borde de un objeto.
 - Línea epipolar válida: línea epipolar cuya forma cumple con una forma que se espera que sea creada por un borde en el mundo de los objetos en una cámara ideal libre de aberraciones, desalineaciones y tolerancias de fabricación.
 - Línea epipolar extendida: conjunto de píxeles de la línea epipolar de una cámara plenóptica extendido por uno o más píxeles (correspondientes al mismo punto del mundo de objetos) de una o más cámaras convencionales.
 - Línea de base: en un sistema multivista, distancia entre el centro de las aberturas de dos cámaras consecutivas (cámaras plenópticas o convencionales o cualquier cámara).
 - Minicámaras inteligentes: módulos de cámara en miniatura de pequeñas dimensiones para dispositivos móviles que pueden tener características adicionales como la capacidad de ajustar su tasa de cuadros automáticamente con cambio de iluminación, enfoque a diferentes distancias, acercar y alejar, etc., transformando la imagen capturada de acuerdo con criterios predefinidos.
 - Correspondencia estéreo (o simplemente correspondencia): técnica que hace coincidir los puntos de una imagen con los puntos de otra imagen, identificando el mismo punto en el mundo de los objetos visto desde diferentes puntos de vista. Este proceso determina qué partes de una imagen corresponden a qué partes de otra imagen, donde las diferencias se deben a diferentes perspectivas.
 - Microimagen: imagen de la abertura principal producida por un determinado microlente de una cámara plenóptica sobre el sensor de imagen.
 - FOV: campo de visión.
- De acuerdo con un aspecto de la presente invención, se proporciona un método de acuerdo con la reivindicación 1 para obtener información de profundidad a partir de un campo de luz. El método comprende los siguientes pasos: generar una pluralidad de imágenes (por ejemplo, al menos una imagen epipolar horizontal, al menos una imagen epipolar vertical, o una combinación de las mismas) a partir de un campo de luz capturado por un dispositivo de adquisición de campo de luz (tal como una cámara plenóptica); un paso de detección de bordes para detectar, en las imágenes epipolares, bordes de objetos en la escena capturados por el dispositivo de adquisición de campo de luz; en cada imagen epipolar, detectando líneas epipolares válidas formadas por un conjunto de bordes; determinar las pendientes de las líneas epipolares válidas.
- En una realización, el paso de detección de bordes comprende calcular una segunda derivada espacial para cada píxel de las imágenes epipolares y detectar los cruces por cero de las segundas derivadas espaciales. El paso de determinar las pendientes de las líneas epipolares válidas puede comprender aplicar un ajuste de línea a los bordes detectados.
- La detección de líneas epipolares válidas en una imagen epipolar puede comprender determinar las líneas epipolares como un conjunto de bordes conectados y analizar las líneas epipolares para determinar si las líneas epipolares son válidas o no. Las líneas epipolares se determinan preferiblemente como un conjunto de píxeles de borde conectados. El análisis de las líneas epipolares para determinar si son válidas o no comprende comprobar el cumplimiento de al menos un criterio. Un criterio se refiere al número de píxeles que forman la línea epipolar que superan un umbral determinado (por ejemplo, el número de píxeles que forman la línea epipolar debe ser al menos igual al número de píxeles de la altura de la imagen epipolar correspondiente). Otro criterio puede referirse a la consistencia de la dirección de los píxeles de los bordes dentro de la imagen epipolar. En una realización, se emplea una combinación de los criterios anteriores.
- Alternativamente, en lugar de comprobar el cumplimiento de al menos un criterio, el análisis de las líneas epipolares para determinar si las líneas epipolares son válidas o no puede comprender un análisis morfológico, un método

heurístico o un algoritmo de aprendizaje automático. El análisis de las líneas epipolares puede incluir ignorar una o varias filas de píxeles en la parte superior y/o en la parte inferior de la imagen epipolar.

5 El método también puede comprender generar un mapa de pendientes asignando pendientes a posiciones en el mundo de los objetos. En una realización, el paso de generar un mapa de pendientes comprende asignar valores de pendiente solo a los bordes detectados. El paso de generar un mapa de pendientes también puede comprender aplicar un algoritmo de relleno para asignar valores de pendiente a las posiciones del mapa de pendientes teniendo en cuenta las pendientes obtenidas previamente para los bordes detectados. El método puede comprender generar un único mapa de pendientes a partir de una combinación de pendientes redundantes obtenidas de diferentes líneas epipolares válidas para la misma posición. Se descartan las pendientes asignadas a una determinada posición con alta dispersión con respecto al resto de los valores de dicha posición.

10 El método puede comprender además generar un mapa de profundidad asignando valores de profundidad a posiciones en el mundo de objetos, en donde el mapa de profundidad se obtiene aplicando una conversión de pendiente a profundidad al mapa de pendiente. El método comprende obtener valores de profundidad correspondientes a las pendientes de las líneas epipolares válidas y generar un mapa de profundidad asignando valores de profundidad a posiciones en el mundo de objetos.

15 El paso de generar un mapa de profundidad puede comprender asignar valores de profundidad solo a los bordes detectados. El paso de generar un mapa de profundidad puede comprender aplicar un algoritmo de relleno para asignar valores de profundidad a las posiciones del mapa de profundidad teniendo en cuenta los valores de profundidad obtenidos previamente para los bordes detectados. El método puede comprender generar un único mapa de profundidad a partir de una combinación de valores de profundidad redundantes obtenidos de diferentes imágenes epipolares para la misma posición para generar un único mapa de profundidad. Se descartan los valores de profundidad asignados a una determinada posición con alta dispersión con respecto al resto de los valores de dicha posición.

20 El método comprende la generación de un mapa de pendiente y/o un mapa de profundidad, en donde el número de posiciones del mapa de pendiente y/o profundidad es mayor que el número de microlentes utilizando la precisión de subpíxeles obtenida en los cruces por cero.

25 Se obtiene sólo un valor de pendiente por línea epipolar válida. El método también puede comprender un paso de aplicar un filtro a las imágenes epipolares para obtener imágenes epipolares filtradas antes de la etapa de detección de bordes. En una realización, el dispositivo de adquisición de campo de luz es una cámara plenóptica.

30 La detección de líneas epipolares válidas comprende extender las líneas epipolares de las imágenes epipolares desde el dispositivo de adquisición de campo de luz con información adicional de imágenes capturadas por al menos un dispositivo de adquisición de imágenes adicional para obtener una línea epipolar extendida.

35 Las imágenes epipolares pueden extenderse añadiendo, por encima y/o por debajo, la información adicional dependiendo de las posiciones relativas del al menos un dispositivo de adquisición de imágenes adicional al dispositivo de adquisición de campo de luz. La información adicional se añade preferiblemente a una cierta distancia por encima y/o por debajo de las imágenes epipolares de acuerdo con los desplazamientos horizontales y verticales computados previamente en un proceso de calibración. En una realización, las imágenes epipolares horizontales se extienden añadiendo la información adicional del al menos un dispositivo de adquisición de imágenes adicional que está alineado horizontalmente con el dispositivo de adquisición de campo de luz. Las imágenes epipolares verticales pueden extenderse añadiendo la información adicional del al menos un dispositivo de adquisición de imágenes adicional que está alineado verticalmente con el dispositivo de adquisición de campo de luz.

40 La información adicional puede comprender píxeles de borde contenidos en imágenes capturadas por al menos una cámara convencional, en donde dichos píxeles de borde corresponden al borde del objeto representado por la línea epipolar. Alternativamente, o además de, la información adicional puede comprender líneas epipolares contenidas en imágenes capturadas por al menos un dispositivo de adquisición de campo de luz adicional, en donde dichas líneas epipolares corresponden al borde del objeto representado por la línea epipolar.

45 En una realización, el método comprende determinar una región de búsqueda en las imágenes capturadas por las cámaras convencionales donde se buscan los píxeles de borde correspondientes a la línea epipolar. El método puede comprender determinar una región de búsqueda en las imágenes capturadas por los dispositivos de adquisición de campo de luz adicionales donde se busca el píxel del borde central de las líneas epipolares de los dispositivos de adquisición de campo de luz adicionales correspondientes al borde del objeto representado por la línea epipolar. En ambos casos, la región de búsqueda puede ser una ventana unidimensional o una ventana bidimensional. El tamaño de la región de búsqueda se selecciona preferiblemente con base en la incertidumbre de las mediciones de profundidad del dispositivo de adquisición de campo de luz esperadas de la curva de dispersión en una primera distancia de profundidad estimada considerando solo el dispositivo de adquisición de campo de luz.

50 De acuerdo con una realización, el método comprende:

- Calcular una regresión lineal de la línea epipolar del dispositivo de adquisición de campo de luz.

- Obtener una línea de extensión a partir de la imagen captada por una cámara convencional.
- Ampliar la imagen epipolar del dispositivo de adquisición de campo de luz con la línea de extensión de la cámara convencional.
- Calcular el punto de intersección de la línea epipolar y la línea de extensión.

- 5 • Definir una región de búsqueda alrededor del punto de intersección.

La imagen epipolar se extiende preferiblemente con la línea de extensión utilizando desplazamientos horizontales y verticales computados previamente durante un proceso de calibración. El proceso de calibración para obtener los desplazamientos horizontal y vertical puede comprender colocar un punto luminiscente en el plano conjugado del dispositivo de adquisición de campo de luz alineado con el eje óptico y calcular los desplazamientos necesarios para obtener una línea epipolar vertical alineada verticalmente con el píxel de la cámara convencional que contiene la luz producida por el punto luminiscente.

10

El paso de obtener una línea de extensión puede comprender determinar una equivalencia entre las coordenadas verticales y/u horizontales de la imagen capturada por el dispositivo de adquisición de campo de luz y las coordenadas verticales y/u horizontales de la imagen de la cámara convencional. En una realización, la equivalencia entre las coordenadas verticales y/u horizontales de los píxeles de los sensores de imagen de los dispositivos de adquisición se obtiene colocando un patrón luminiscente en el plano conjugado del dispositivo de adquisición de campo de luz alineado con el eje óptico y calcular la relación entre los tamaños vertical y/u horizontal de los patrones de luz capturados por cada uno de los sensores de imagen de los dispositivos de adquisición. El método puede comprender además aplicar un proceso de correspondencia para encontrar el píxel del borde en la imagen de la cámara convencional que coincida con el borde del objeto representado por la línea epipolar.

15

20

El paso de obtener la línea epipolar extendida puede comprender asignar pesos a la línea epipolar y la información adicional. El método puede comprender además obtener imágenes totalmente enfocadas desde un sistema de vista múltiple que comprende el dispositivo de adquisición de campo de luz y al menos una cámara convencional. El paso para obtener imágenes totalmente enfocadas comprende:

- 25 • Para objetos ubicados a una distancia más allá de un umbral del sistema de vista múltiple, obtener imágenes enfocadas de la al menos una cámara convencional.

- Para objetos ubicados a una distancia por debajo de un umbral del sistema de vista múltiple, obtener imágenes reenfoadas desde el dispositivo de adquisición de campo de luz.

- 30 • Componer una imagen totalmente enfocada final tomando, para distancias por debajo del umbral, los objetos más nítidos de las imágenes reenfoadas del dispositivo de adquisición de campo de luz y, para distancias más allá del umbral, tomando las imágenes enfocadas de al menos una cámara.

En otra realización, el método comprende además un paso de reenfoque de imágenes de un sistema de vista múltiple que comprende el dispositivo de adquisición de campo de luz y al menos una cámara convencional. El paso de reenfoque de imágenes comprende:

- 35 • Calcular un mapa de profundidad.
- Para objetos ubicados a una distancia por debajo de un umbral del sistema de vista múltiple, utilizando imágenes reenfoadas del dispositivo de adquisición de campo de luz.

- 40 • Para objetos ubicados a una distancia más allá de un umbral del sistema de vista múltiple, seleccionar un rango enfocado de distancias de la al menos una cámara convencional y desenfoque objetos en la imagen colocados a una distancia más allá del rango enfocado seleccionado. La borrosidad se realiza preferiblemente usando un filtro gaussiano.

El método puede comprender generar un mapa de pendiente para el dispositivo de adquisición de campo de luz y para al menos un dispositivo de adquisición de campo de luz adicional. En una realización, los diferentes mapas de pendientes generados se combinan en un único mapa de pendientes considerando los desplazamientos verticales y horizontales entre los dispositivos de adquisición de campo de luz. El método puede comprender generar un mapa de profundidad utilizando algoritmos de correspondencia estéreo entre todas las vistas capturadas por los dispositivos de adquisición de campo de luz y las imágenes capturadas por las cámaras convencionales. En una realización formada por una cámara plenóptica y una o más cámaras convencionales, el método incluye usar información proporcionada por las cámaras convencionales para mejorar la precisión de las pendientes estimadas en las imágenes epipolares de la cámara de campo de luz.

50

Según otro aspecto de la presente invención, se proporciona un dispositivo de acuerdo con la reivindicación 14 para generar un mapa de profundidad a partir de un campo de luz. El dispositivo comprende unos medios de procesamiento

- 5 configurados para realizar los pasos del método explicado anteriormente. En una realización, el dispositivo puede comprender un dispositivo de adquisición de campo de luz, tal como una cámara plenóptica. En otra realización, el dispositivo puede comprender un sistema multivista que comprende una pluralidad de cámaras en las que al menos una de ellas es un dispositivo de adquisición de campo de luz. El sistema multivista puede comprender al menos una cámara convencional. En una realización, el dispositivo puede comprender una o más cámaras plenópticas y una o más cámaras convencionales. Las una o más cámaras convencionales pueden alinearse vertical y/u horizontalmente con el dispositivo de adquisición de campo de luz. El sistema de vista múltiple puede comprender un arreglo bidimensional de dispositivos de adquisición de campo de luz y cámaras convencionales.
- 10 Preferiblemente, el dispositivo es un dispositivo móvil electrónico, tal como un teléfono inteligente, una tableta, un ordenador portátil o una cámara compacta. Los medios de procesamiento pueden comprender una primera CPU configurada para obtener y analizar imágenes epipolares horizontales y una segunda CPU configurada para obtener y analizar imágenes epipolares verticales. En otra realización, los medios de procesamiento comprenden un procesador de múltiples núcleos. Alternativamente, o además de, los medios de procesamiento pueden comprender una unidad de procesamiento de gráficos.
- 15 Según otro aspecto más de la presente invención, se proporciona un producto de programa informático de acuerdo con la reivindicación 15 para generar un mapa de profundidad a partir de una imagen capturada por una cámara plenóptica o generar un mapa de profundidad a partir de un conjunto de imágenes capturadas por una o más cámaras plenópticas y una o más cámaras convencionales, que comprenden instrucciones de código informático que, cuando son ejecutadas por un procesador, hacen que el procesador realice el método anteriormente explicado. En una
20 realización, el producto de programa informático comprende al menos un medio de almacenamiento legible por ordenador que tiene grabadas en él las instrucciones del código informático.
- Breve descripción de los dibujos
- A continuación se describen muy brevemente una serie de dibujos que ayudan a comprender mejor la invención y que se relacionan expresamente con un modo de realización de dicha invención, presentados como ejemplo no limitativo
25 de la misma.
- La figura 1A representa una cámara plenóptica que captura la luz de un objeto colocado en el plano conjugado del arreglo de microlentes. La figura 1B ilustra la luz capturada por el sensor de imagen de la cámara plenóptica.
- Las figuras 2A y 2B muestran una cámara plenóptica que captura la luz de un objeto colocado más cerca que el plano conjugado del arreglo de microlentes.
- 30 Las figuras 3A y 3B representan una cámara plenóptica que captura la luz de un objeto colocado más allá del plano conjugado del arreglo de microlentes.
- Las figuras 4A-4D representan el proceso de formación de imágenes epipolares centrales horizontales y verticales para los ejemplos de las figuras 1, 2 y 3.
- 35 La figura 5 representa, de acuerdo con una realización, un diagrama de un flujo de proceso para determinar la profundidad de un punto en el mundo de los objetos analizando las líneas detectadas en las imágenes epipolares.
- Las figuras 6A-6C muestran diversos ejemplos de líneas epipolares válidas y no válidas en una imagen epipolar. Las figuras 6D-6G muestran el proceso de cálculo de la pendiente de la línea epipolar en el ejemplo de la figura 6A. Las figuras 6H-6J ilustran el proceso de cálculo de la pendiente de la línea epipolar de la figura 6B.
- 40 Las figuras 7A-7E muestran una imagen epipolar que incluye varias líneas epipolares y el proceso de cálculo de las pendientes correspondientes.
- La figura 8 muestra un ejemplo de un mapa de escasa profundidad que muestra tres objetos a diferentes profundidades.
- La figura 9 muestra un ejemplo de un mapa de profundidad denso que muestra tres objetos a diferentes profundidades.
- 45 Las figuras 10A-10C muestran diferentes realizaciones de dispositivos móviles electrónicos que ejecutan el método de la presente invención cuando el dispositivo de captura es una única cámara plenóptica.
- La figura 11A-11D muestra la incertidumbre introducida por el tamaño de píxel no infinitesimal del sensor al medir la pendiente de las líneas epipolares producidas por un objeto ubicado a cierta distancia.
- 50 La figura 12 muestra la distribución de probabilidad de la profundidad estimada por una cámara plenóptica para las profundidades de dos objetos ubicados a diferentes profundidades (a la izquierda cerca de la cámara y a la derecha más lejos de la cámara).
- Las figuras 13A-13E muestran un dispositivo portátil que contiene cinco posibles configuraciones del sistema multivista formado por una cámara plenóptica y varias cámaras convencionales.

La figura 14 ilustra el proceso de extensión de una imagen epipolar capturada con una cámara plenóptica con una imagen 2D de la misma escena capturada por una cámara convencional.

La figura 15 muestra la primera estimación de pendiente obtenida considerando solo una línea epipolar plenóptica y cómo la estimación de pendiente se mejora al considerar la imagen 2D de una cámara convencional.

5 Las figuras 16A-16B representan, de acuerdo con una realización, los varios pasos del procedimiento de la presente invención para mejorar las estimaciones de pendiente de una sola cámara plenóptica.

Las figuras 17A-17B representan las ventanas o regiones (1D y 2D) donde se busca el píxel del borde en la imagen de la cámara convencional.

10 La figura 18 muestra una posible realización del sistema multivista de esta invención: un arreglo bidimensional de cámaras plenópticas y/o cámaras convencionales.

Las figuras 19A y 19B comparan la distribución de probabilidad al medir la profundidad de un objeto ubicado a cierta profundidad, con una sola cámara plenóptica (figura 19A) y con un sistema multivista compuesto por una cámara plenóptica y una cámara convencional (figura 19B).

15 La figura 19A-19B ilustra cómo una posible configuración formada por una cámara plenóptica y convencional captura la luz emitida por un objeto en el mundo colocado en el plano conjugado del arreglo de microlentes de la cámara plenóptica.

La figura 20A-20B muestra el proceso de cálculo del desplazamiento horizontal H.

La figura 21 ilustra la línea epipolar extendida obtenida para un sistema multivista formado por dos cámaras plenópticas y una cámara convencional.

20 La figura 22A representa un diagrama de flujo para la detección de líneas epipolares válidas en una realización que usa una única cámara plenóptica. La figura 22B ilustra un diagrama de flujo para la detección de líneas epipolares válidas en una realización utilizando una cámara plenóptica y la información adicional capturada por al menos una cámara adicional.

25 Las figuras 23A-23C muestran diferentes realizaciones de dispositivos móviles electrónicos que ejecutan el método para el sistema de vista múltiple.

Descripción detallada

30 La presente invención se refiere a un dispositivo y método para generar un mapa de profundidad a partir de un campo de luz. Un campo de luz puede ser capturado por varios tipos de dispositivos. Por simplicidad, primero se considerarán solo cámaras plenópticas. Posteriormente, se describe el método cuando se aplica a un sistema multivista que consta de una o más cámaras plenópticas y una o más cámaras convencionales. No obstante, el método descrito en este documento se puede aplicar a campos de luz capturados por cualquier otro dispositivo, incluidos otros dispositivos de formación de imágenes integrales.

35 Una cámara convencional solo captura información espacial bidimensional de los rayos de luz capturados por el sensor. Además, la información de color también se puede capturar mediante el uso de los denominados sensores con patrón de Bayer u otros sensores de color. Una cámara plenóptica captura no solo esta información sino también la dirección de llegada de los rayos. Por lo general, una cámara plenóptica se fabrica colocando un arreglo de microlentes entre el lente principal y el sensor. Cada uno de los microlentes (lx, ly) está formando una pequeña imagen de la apertura principal en el sensor. Estas pequeñas imágenes se conocen como microimágenes, de modo que, cada píxel (px, py) de cualquier microimagen captura los rayos de luz que provienen de una parte diferente de la apertura principal, cada una de las micro imágenes debajo de cualquier microlente es una imagen de la apertura del lente principal, y cada píxel en la posición px1, pyl o pxn, pyn en cada microlente integra luz proveniente de una parte dada de la apertura (axn, ayn) sin importar la posición del microlente. La luz que atraviesa la apertura en posición (axn, ayn) proveniente de diferentes ubicaciones del mundo de objetos incidirá en diferentes microlentes, pero siempre estará integrada por el píxel (pxn, pyn). En consecuencia, las coordenadas (px, py) de un píxel dentro de una microimagen determinan la dirección de llegada de los rayos capturados a un microlente dado y (lx, ly) determinan la posición espacial bidimensional. Toda esta información se conoce como campo de luz y se puede representar mediante una matriz de cuatro dimensiones LF (px, py, lx, ly) o una matriz de cinco dimensiones LF (px, py, lx, ly, c) si la información de color (c) se considera. En lo sucesivo, solo se considerarán los sensores monocromáticos. Estos sensores capturan la intensidad de la luz detectada para todo el espectro para el que han sido diseñados. Sin embargo, la invención descrita en este documento puede extenderse directamente a sensores que también capturan información de color, como resultará obvio para un experto en el campo. Una posible adaptación de la presente invención para este tipo de sensores es aplicar el método descrito en este documento a cada canal de color por separado para aumentar aún más la redundancia de las estimaciones de profundidad.

Los objetos en el mundo a diferentes profundidades o distancias de la cámara producen diferentes patrones de iluminación en la imagen capturada por el sensor de imagen de una cámara plenóptica. La figura 1A representa una vista bidimensional esquemática de una cámara 100 plenóptica que comprende un lente 102 principal, un arreglo 104 de microlentes (formado por una pluralidad de microlentes 105 reunidos en filas y columnas) y un sensor 106 de imagen colocado detrás del arreglo 104 de microlentes para detectar intensidad, color e información direccional. En el ejemplo que se muestra en la figura 1A, la cámara 100 plenóptica está capturando los rayos 108 de luz entrantes desde un punto 110 de objeto colocado en el plano conjugado del arreglo 104 de microlentes. La figura 1B representa la luz capturada por el sensor 106 de imagen de la cámara 100 plenóptica. Cada celda de la cuadrícula representa la microimagen 112 producida por cada microlente 105 sobre el sensor 106 de imagen.

Cuando la imagen de un punto 110 de objeto se enfoca en el arreglo 104 de microlentes, el punto 110 de objeto se coloca en el plano conjugado del MLA a través del lente 102 principal de la cámara 100 plenóptica y solo un punto infinitesimal sobre un microlente 105 está iluminado (en realidad, no es un punto infinitesimal sino un patrón de difracción). Además, dado que la separación entre los microlentes 105 y el sensor 106 de imagen es aproximadamente la distancia focal de los microlentes 105, todos los píxeles de la correspondiente microimagen 112 recogen exactamente la misma intensidad de luz, como se muestra en la figura 1B. En todas las imágenes sobre el plano del sensor de imagen que se muestran en este documento, el color negro se usa para representar la falta de luz y cuanto más blancos son los píxeles, más iluminados están con niveles de gris que significan iluminaciones parciales.

Por otro lado, los puntos 110 de objeto de la escena que están más cerca que el plano conjugado del arreglo 104 de microlentes en el mundo de objetos iluminarán más microlentes 105 ya que el punto de enfoque en el mundo de la imagen habría estado más lejos que el arreglo 104 de microlentes (más hacia el lado derecho), y el patrón capturado por los píxeles del sensor será diferente. El diagrama de este escenario se ilustra en la figura 2A, mientras que la figura 2B muestra el patrón correspondiente producido sobre el sensor 106 de imagen.

A la inversa, un punto 110 de objeto que está más lejos que el plano conjugado del arreglo 104 de microlentes ilumina también más microlentes 105 pero ahora el punto de enfoque está más cerca del lente 102 principal que la posición del arreglo 104 de microlentes y, por lo tanto, el patrón capturado por el sensor 106 de imagen difiere de las dos situaciones anteriores, como se muestra en las figuras 3A y 3B. Los niveles de gris en algunas de las microimágenes 112 corresponden a píxeles parcialmente iluminados, mientras que en los píxeles blancos toda el área del píxel ha sido alcanzada por la luz procedente del punto 110 de objeto en el mundo del objeto.

Estos diversos patrones del campo de luz capturados por el sensor 106 de imagen se pueden representar en imágenes epipolares tomando cortes bidimensionales del campo de luz. Las figuras 4A-4C representan, respectivamente para cada uno de los escenarios de las figuras 3A-3C, el proceso de generación de imágenes 400 epipolares horizontales (fila de arriba) e imágenes 402 epipolares verticales (fila de abajo), reorganizando los píxeles capturados por el sensor 106 de imagen. La figura 4A muestra el patrón creado sobre el sensor para un punto en el mundo de objetos ubicado en el plano conjugado del arreglo de microlentes. La figura 4B representa el patrón creado para un punto en el mundo de objetos ubicado más cerca de la cámara que el plano conjugado del arreglo de microlentes. La figura 4C representa el patrón creado para un punto en el mundo de objetos ubicado más lejos de la cámara que el plano conjugado del arreglo de microlentes.

Las imágenes 400 epipolares horizontales se forman fijando las coordenadas (p_y , l_y) del campo de luz, mientras que las imágenes 402 epipolares verticales se forman fijando las coordenadas (p_x , l_x).

En las figuras 4A-4C las imágenes 400 epipolares horizontales y las imágenes 402 epipolares verticales son, respectivamente, imágenes epipolares horizontales-centrales e imágenes epipolares verticales-centrales ya que los píxeles p_y y p_x que se han fijado para las imágenes epipolares son los píxeles central-horizontal y central-vertical de sus respectivos microlentes. Las figuras 4A-4C muestran cómo las imágenes 402 epipolares verticales (fila de abajo) y las imágenes 400 epipolares horizontales (fila de arriba) se forman directamente a partir del campo de luz capturado.

La figura 4D muestra con más detalle el proceso de generación de una imagen 400 epipolar horizontal (una vista de zoom de la fila de arriba de la figura 4A), formada apilando las líneas (410, 412, 414, 416, 418) de píxeles ubicadas en altura $p_y=Y$ de las microimágenes 112 correspondientes a los microlentes 105 ubicados en la misma fila l_y (en $l_y = 3$ en el ejemplo de la figura 4D, los microlentes 105 horizontales-centrales del arreglo 104 de microlentes). Dado que la altura seleccionada $p_y=Y$ de las líneas (410, 412, 414, 416, 418) de píxeles en las microimágenes 112 es la altura central, la imagen 400 epipolar horizontal se considera una imagen epipolar horizontal-central. Los píxeles individuales ($p_x = 1, p_x = 2, \dots$) que forman cada línea (410, 412, 414, 416, 418) de píxeles en la figura 4D no están representados. Por el contrario, cada imagen (402) epipolar vertical se forma apilando las líneas de píxeles colocadas a un ancho determinado $p_x = X$ de las microimágenes 112 correspondientes a los microlentes 105 ubicados en la misma columna l_x .

Como puede verse en las figuras 4A-4D, en las imágenes 400 epipolares horizontales y las imágenes 402 epipolares verticales se forma una línea 430 epipolar (coloreada en blanco). Todos los píxeles iluminados (píxeles blancos) de esta línea 430 epipolar corresponden al mismo punto 110 de objeto en el mundo del objeto, como se ilustra en los ejemplos de las figuras 1B-3B. Una línea 430 epipolar es un conjunto de píxeles iluminados conectados (no píxeles negros) dentro de una imagen epipolar que se detectan como bordes. Además, la pendiente de la línea 430 epipolar

está directamente relacionada con el tipo de patrón iluminado sobre los microlentes 104 y sobre el sensor 106 de imagen y también con la profundidad correspondiente del punto 110 de objeto en el mundo del objeto. En el ejemplo de la figura 4D, la pendiente de la línea 430 epipolar es ∞ (ángulo= 90° con respecto al eje horizontal), que corresponde a una distancia tal que el punto 110 de objeto se coloca en el plano conjugado del arreglo 104 de microlentes (figura 1A). Si la pendiente es positiva (el ángulo es menor que 90°), el punto 110 de objeto está más cerca del lente 102 principal (figura 2A), mientras que si la pendiente es negativa (el ángulo es mayor que 90°), el punto 110 de objeto está más lejos del lente 102 principal (figura 3A).

Por tanto, conociendo este patrón, es posible rastrear los patrones muestreados por los píxeles a través de la cámara 100 plenóptica y obtener la profundidad exacta (dz) del punto 110 de objeto que produce dicho patrón. La relación entre profundidad y pendiente depende de las dimensiones físicas y el diseño (que se conocen) de la cámara 100 plenóptica utilizada para capturar el campo de luz.

En consecuencia, una cierta pendiente de una línea 430 epipolar está inequívocamente relacionada con una cierta profundidad de un punto 110 de objeto de la escena del mundo tridimensional real.

La pendiente estimada de una línea epipolar contiene información de profundidad de un determinado objeto. Pendiente y profundidad son dos caras de la misma moneda (es posible obtener profundidades de pendientes de forma determinista y viceversa, con solo errores de cuantificación en las conversiones debido a que los píxeles del sensor no son infinitesimales). La pendiente en sí es suficiente para obtener información sobre la profundidad relativa de los diferentes objetos de una escena. Esta información relativa (es decir, la pendiente) puede ser útil para algunas aplicaciones en las que no es necesario proporcionar información de profundidad absoluta, tal como identificar los diferentes objetos de una escena que se encuentran a la misma profundidad (misma pendiente). Por lo tanto, en tales escenarios, el cálculo de pendientes es suficiente y se puede omitir la conversión de pendiente a profundidad.

El método de la presente invención se basa en el cálculo de profundidades solo para las áreas donde hay bordes en la proyección del mundo sobre el arreglo 104 de microlentes (o lo que es lo mismo, bordes en el mundo objeto). En una realización preferida, se aplica una regresión lineal a los píxeles iluminados que forman una línea 430 epipolar para obtener una cierta pendiente. Al analizar una línea 430 epipolar en una imagen epipolar horizontal 400 o vertical 402, todas las vistas plenópticas distribuidas a lo largo de la dimensión horizontal (px) o vertical (py) se consideran ya que el mismo punto 110 de objeto ha sido capturado por todas estas vistas. Por lo tanto, la técnica de regresión lineal reduce el ruido estadístico al aprovechar la información redundante en una dimensión.

Además, el método incluye una etapa adicional para reducir aún más el ruido estadístico analizando el mismo punto 110 de objeto en las imágenes epipolares horizontales 400 y verticales 402 y considerando los valores de profundidad obtenidos con las diversas imágenes (400, 402) epipolares que contienen información del mismo punto 110 de objeto (por ejemplo, está claro que un punto 110 de objeto único en el mundo de objetos, como se muestra en las figuras 1 a 4, produce varias impresiones en varios puntos del sensor 106 de imagen y esas impresiones aparecen en varias imágenes epipolares verticales y varias horizontales).

En una realización, se identifican todas las líneas 430 epipolares formadas en las imágenes epipolares horizontales 400 y verticales 402 y se calcula la pendiente correspondiente. Luego, se calcula la profundidad correspondiente del punto 110 de objeto considerando las dimensiones físicas del dispositivo.

Sólo se calcula un valor de pendiente y profundidad por línea 430 epipolar ya que una línea epipolar está formada por el mismo punto 110 de objeto capturado desde varios puntos de vista. Por lo tanto, la cantidad de datos se reduce drásticamente debido a los siguientes dos factores:

(i) En comparación con otros enfoques que procesan todos los puntos capturados por el sensor 106 de imagen, el presente método solo procesa los puntos de interés, es decir, las áreas del mundo de objetos que se detectan como bordes porque crean líneas epipolares (como áreas del mundo de los objetos completamente uniforme, sin bordes, no producen ninguna línea epipolar sino colores uniformes).

(ii) Es posible almacenar solo un valor de pendiente por línea 430 epipolar en lugar de almacenar un valor por cada píxel que forma la línea 430 epipolar.

Por lo tanto, la salida de este proceso de cálculo puede ser solo los valores de profundidad correspondientes de estas pendientes detectadas.

Las pendientes obtenidas analizando las imágenes epipolares horizontales 400 y verticales 402 y las líneas 430 epipolares se combinan en una matriz de cuatro dimensiones para reducir el ruido estadístico, debido al hecho de que la confiabilidad de la salida se mejora mediante la redundancia de mediciones adicionales, ya que se considera el mismo píxel del sensor cuando se analizan las imágenes epipolares verticales 402 y horizontales 400 y, por lo tanto, varios valores de pendiente pueden haber sido producidos por el mismo punto del mundo del objeto.

Las pendientes calculadas se transforman a las correspondientes profundidades del objeto considerando los parámetros físicos de la cámara 100 plenóptica. Esta etapa de transformación se realiza después de combinar todas las pendientes redundantes, reduciendo drásticamente el número de transformaciones de pendiente a profundidad.

La matriz de cuatro dimensiones de profundidades/pendientes generada previamente se combina en un mapa de profundidad/pendiente escasa bidimensional (escasa porque ofrece lecturas solo donde hay bordes en el mundo de los objetos), reduciendo aún más el ruido estadístico y, así, aumentar la calidad del mapa de profundidad.

5 Las profundidades/pendientes calculadas para las líneas 430 epipolares en las imágenes epipolares horizontales 400 y verticales 402 se combinan directamente en un mapa disperso bidimensional de profundidad/pendiente, realizando por tanto una única etapa de combinación, lo que aumenta la eficiencia computacional.

El mapa de profundidad/pendiente disperso se rellena aplicando técnicas de relleno de imagen para obtener valores de profundidad/pendiente para cada píxel (dx, dy).

10 Solo las imágenes epipolares horizontales-centrales (formadas estableciendo la coordenada py para que sea igual al píxel central en la dimensión py dentro de una microimagen 112), y/o solo las imágenes epipolares verticales-centrales (formadas tomando la coordenada px igual al píxel central en la dimensión px dentro de una microimagen), como se muestra en las figuras 4A-4D se consideran con el objetivo de reducir el número de imágenes epipolares a analizar y, así, aumentar el rendimiento a costa de reducir la redundancia estadística.

15 El método de la presente invención se puede implementar en dispositivos móviles (por ejemplo, teléfonos inteligentes, tabletas u ordenadores portátiles) equipados con una cámara plenóptica.

20 La figura 5 muestra un diagrama de flujo de un método para generar mapas de profundidad de acuerdo con una realización. Para generar un mapa de profundidad, el método genera imágenes 502 epipolares horizontales y verticales 503 a partir de un campo 501 de luz capturado por una cámara 100 plenóptica. Para cada imagen epipolar horizontal 502 y vertical 503 generada, las líneas (510, 511) epipolares válidas dentro de las imágenes epipolares se identifican. A continuación, se calculan las pendientes (512, 513) de estas líneas (510, 511) epipolares válidas y finalmente se obtienen los correspondientes valores (514, 515) de profundidad.

La figura 5 describe el proceso de identificación y procesamiento de las líneas (510, 511) epipolares válidas tomando como entrada un campo 501 de luz capturado y procesando todas las imágenes epipolares ("EPI" en la figura 5) horizontales 502 y verticales 503 realizando los siguientes pasos:

- 25
- Para cada imagen 502 epipolar horizontal, obtenida para un par fijo de valores (py, ly):
 - Aplicar un filtro unidimensional (o mayor) a lo largo de la dimensión lx para reducir el ruido, obteniendo una imagen 504 epipolar horizontal filtrada.
 - Para cada píxel (px, lx), calcular la segunda derivada 506 espacial en el píxel (px, lx) sobre la intensidad de la luz o el contraste de los píxeles a lo largo de la dimensión lx.
- 30
- Determinar los bordes 508 del mundo de objetos analizando las líneas epipolares con precisión de subpíxeles, más específicamente detectando el cruce por cero de las segundas derivadas espaciales.
 - Buscar cada uno de los cruces por cero que estén correctamente ordenados formando una línea 510 epipolar válida, descartando líneas epipolares inválidas.
- 35
- Para cada imagen 503 epipolar vertical, obtenida para un par fijo de valores (px, lx):
 - Aplicar un filtro unidimensional a lo largo de la dimensión ly para reducir el ruido, obteniendo una imagen 505 epipolar vertical filtrada.
 - Para cada píxel (py, ly), calcular la segunda derivada 507 espacial a lo largo de la dimensión ly.
 - Determinar los bordes 509 del mundo de objetos analizando las líneas epipolares con precisión de subpíxeles, más específicamente detectando el cruce por cero de las segundas derivadas espaciales.
- 40
- Buscar cada uno de los cruces por cero que estén correctamente ordenados formando una línea 511 epipolar válida, descartando líneas epipolares inválidas.
 - Para cada línea (510, 511) epipolar válida encontrada en las imágenes epipolares horizontales y verticales, los bordes de precisión de subpíxeles se utilizan para determinar la pendiente (512, 513) de la línea (510, 511) epipolar válida realizando una técnica de regresión lineal (pero también se puede utilizar cualquier otra técnica de ajuste).
- 45
- Para cada pendiente calculada, se aplica una conversión de pendiente a profundidad (514, 515).
 - Finalmente, se generan dos matrices de profundidades, una matriz 516 de profundidad horizontal para las imágenes 502 epipolares horizontales y una matriz 517 de profundidad vertical para las imágenes 503 epipolares verticales.

Los pasos del filtro de reducción de ruido para obtener imágenes epipolares horizontales 504 o verticales 505 filtradas pueden descartarse opcionalmente para aumentar la velocidad de procesamiento.

Las dos matrices de pendiente (obtenidas de las imágenes 502 epipolares horizontales y verticales 503) se combinan en una única matriz de pendiente y finalmente obtienen una única matriz de profundidad.

5 Los cruces por cero de las segundas derivadas espaciales se identifican mediante valores consecutivos positivo-negativo o negativo-positivo de la segunda derivada. Además, para obtener una precisión de subpíxeles, se considera la magnitud de la segunda derivada de estos puntos para determinar dónde está teniendo lugar el cruce por cero real. Un experto en la técnica reconocería que muchos otros métodos de detección de bordes (tales como el operador del detector de bordes Canny, métodos de ajuste de curvas o métodos basados en momentos) también se pueden aplicar para este propósito y las técnicas descritas en este documento no se limitan al método de cruce por cero. Sin embargo, es sumamente importante obtener la máxima precisión posible a la hora de determinar la pendiente de las líneas formadas por los bordes detectados, por eso la precisión de subpíxeles para determinar los bordes es muy importante. Uno de los objetivos del método propuesto es ser computacionalmente eficiente (este requisito debe tenerse en cuenta al elegir el algoritmo de detección de bordes que se utilizará).

15 Las áreas del mundo de objetos completamente uniformes (sin ninguna textura o contraste de color) no producirán ninguna línea epipolar ya que todos los píxeles registrarán la misma intensidad de luz, independientemente de la distancia de las fuentes de luz a la cámara. Todos los ejemplos mostrados en las figuras 1 a 4 corresponden a un mundo de objetos "oscuro" con sólo una fuente de luz de punto radiante (punto 110 de objeto) creando líneas 430 epipolares dentro de imágenes (400, 402) epipolares.

20 En una situación real, las líneas 430 epipolares se crean mediante un cambio de contraste o un cambio de color, y es por eso que las líneas 430 epipolares corresponden a bordes (cambios de color o contraste) en el mundo de los objetos.

25 Por tanto, las líneas 430 epipolares son producidas por los bordes de los objetos. La primera derivada de las imágenes epipolares (es decir, sobre la intensidad de los píxeles) proporciona el gradiente (es decir, la rapidez con la que cambia la intensidad de la luz o el contraste). La segunda derivada indica dónde cambia más rápidamente el contraste (que corresponde a los bordes de los objetos en el mundo de los objetos). Dado que la segunda derivada no necesariamente tendrá el cruce por cero en un píxel dado (ya que depende de los valores de intensidad de la luz en píxeles, por ejemplo, la imagen epipolar en la figura 6A tiene algún nivel de gris), los bordes del objeto se determinan con subpíxeles precisión.

30 Debido a la propia naturaleza y las limitaciones de diseño de una cámara 100 plenóptica, los píxeles que forman una línea (510, 511) epipolar válida dentro de una imagen epipolar, deben estar necesariamente en posiciones vecinas (es decir, los puntos que forman una línea válida línea epipolar debe estar conectada) y debe componer una línea con todos sus puntos yendo en la misma dirección a medida que avanzamos hacia arriba-abajo o hacia abajo-arriba en la línea epipolar.

35 Las figuras 6A-6C representan un ejemplo (figura 6A) de una línea 610 epipolar válida en una imagen 600 epipolar y varios ejemplos (figuras 6B y 6C) de líneas (612, 614) epipolares no válidas en las respectivas imágenes (602, 604) epipolares. En una realización preferida, solo se consideran las posiciones vecinas cuando se buscan bordes en una imagen epipolar para formar una línea epipolar válida (comenzando desde el píxel central detectado como borde, las flechas en las figuras 6A-6C representan las posiciones vecinas que se consideran para determinar los píxeles de borde conectados que forman la línea epipolar). En consecuencia, las líneas 610 epipolares como la que se muestra en la figura 6A se consideran válidas, mientras que las líneas 612 epipolares como la que se muestra en la figura 6B se detectan como no válidas como el píxel en la parte superior 620 y el píxel en la parte inferior 622 de la imagen 602 epipolar no está conectada con el resto de la línea 612 epipolar.

45 A primera vista, las líneas 614 epipolares como la que se muestra en la figura 6C pueden considerarse como una línea epipolar válida. Sin embargo, debido a la naturaleza de las cámaras 100 plenópticas, tales líneas no ocurrirían en un dispositivo impecable (los píxeles en la parte superior 630 y en la parte inferior 632 no siguen la misma dirección que el resto de la línea epipolar). Estos píxeles (630, 632) extremos de este tipo de líneas se pueden omitir al calcular la pendiente de las líneas epipolares, y aún así ser considerados como líneas epipolares válidas, ya que los píxeles externos posiblemente provengan de aberraciones del lente principal. De esta manera, intercambiamos la potencia de luz recibida y las capacidades de discriminación de pendiente para reducir las aberraciones de píxeles extremos, formados por rayos que atravesaban la parte periférica más aberrada de la abertura. También es posible que toda la línea pueda etiquetarse como no válida para evitar realizar cálculos con líneas epipolares no válidas.

50 Heurísticamente, es fácil para un ser humano discriminar entre líneas epipolares válidas y no válidas inspeccionando visualmente la morfología de las líneas. Sin embargo, los algoritmos para tomar una decisión en un ordenador no son sencillos. Para un experto en el tema no es difícil concebir varios algoritmos diferentes para realizar esa tarea y las implementaciones particulares de cualquier algoritmo que analice la morfología son irrelevantes para el contenido de la invención. Se ha definido heurísticamente cómo identificar líneas epipolares válidas y se pueden desarrollar muchas soluciones informáticas para realizar esa tarea.

55

Sólo las líneas epipolares que tienen al menos el mismo número de píxeles iluminados que la altura de las imágenes epipolares se consideran líneas válidas. Esto puede aumentar la precisión de los cálculos de pendiente en dispositivos donde las aberraciones han sido prácticamente corregidas (óptica o computacionalmente en una etapa previa).

5 Las mayores aberraciones del lente 102 principal se producen en los extremos del lente (áreas alejadas de su centro en las que la aproximación paraxial ya no es válida). Todos los rayos de luz que pasan a través de estas partes extremas del lente 102 principal son más aberrados que los rayos que atraviesan el lente más cerca de su centro. En una cámara 100 plenóptica, estos rayos son capturados por los píxeles extremos de cada microimagen 112, o píxeles extremos de cada microlente 104, que también son los píxeles 640 extremos (figura 6A) cerca de la parte superior o inferior de las imágenes epipolares. Por tanto, los píxeles 640 extremos de las imágenes epipolares pueden omitirse para reducir los efectos de las aberraciones ópticas así como para aumentar el número de valores de profundidad detectados (aumentando el número de líneas epipolares válidas sin tener en cuenta los píxeles extremos). Por tanto, las líneas epipolares que tienen menos píxeles que la altura en píxeles de las imágenes epipolares pueden considerarse válidas, como por ejemplo las figuras 6B y 6C sin tener en cuenta sus píxeles superiores (620, 630) e inferiores (622, 632).

15 Las figuras 6D-6G representan un ejemplo del proceso de cálculo de la pendiente de la línea 610 epipolar en la imagen 600 epipolar de la figura 6A. En este ejemplo, se han considerado los siguientes valores "I" de intensidad de los píxeles, como se muestra en la tabla 642 de la figura 6D: un valor de 0 para los píxeles negros, un valor de 20 para los píxeles de color gris oscuro, un valor de 60 para píxeles de color gris claro y un valor de 100 para los píxeles blancos. La tabla 644 de la figura 6E representa la segunda derivada numérica en el píxel i de la intensidad I a lo largo de la dimensión lx, de acuerdo con la siguiente ecuación:

$$\frac{\partial^2 I(i)}{\partial lx} = I(i+1) + I(i-1) - 2 \cdot I(i)$$

donde i+1 representa el píxel subsiguiente y i-1 el píxel precedente sobre la dimensión lx. La distancia Δlx entre píxeles consecutivos es siempre la misma (se ha considerado un valor de Δlx = 1).

25 La figura 6F representa un gráfico 646 con los valores de la segunda derivada (eje vertical) para cada píxel px (eje horizontal) a lo largo de la dimensión lx (sub-eje horizontal), mostrando los cruces por cero 650 de la segunda derivada identificada por valores consecutivos positivo-negativo o negativo-positivo. Como se explicó anteriormente, los bordes del objeto en el mundo del objeto se determinan detectando los cruces por cero 650 de la segunda derivada espacial.

30 La figura 6G representa, en la imagen 600 epipolar de la figura 6A, los cruces por cero 650 con precisión de subpíxeles. Para comprender la precisión de subpíxeles, el cruce por cero se produjo para el píxel px=1 entre los microlentes lx=2 (con un valor de segunda derivada de 100) y lx=3 (valor de segunda derivada de -200), se ha ampliado. La línea 652 que conecta ambos valores de la segunda derivada interseca la ordenada cero en el cruce por cero 650, que se encuentra dentro de lx=2 con precisión de subpíxeles. La pendiente de la línea 610 epipolar de la figura 6G se obtiene aplicando una regresión 656 lineal a los cruces por cero 650 detectados y calculando directamente la pendiente de la regresión 656 lineal.

35 Las figuras 6H-6J representan otro ejemplo para el proceso de cálculo de la pendiente de la línea 612 epipolar en la imagen 602 epipolar de la figura 6B. Los valores I de intensidad de los píxeles se muestran en la tabla 653 de la figura 6H, mientras que la tabla 654 de la figura 6I representa los valores de la segunda derivada. Los cruces por cero 650 se computan y se muestran como puntos en la figura 6J. La pendiente de la línea 612 epipolar se computa aplicando una regresión 658 lineal a los cruces por cero 650 detectados. Se tiene en cuenta que la regresión 658 lineal de la línea 612 epipolar en la figura 6J tiene una pendiente más alta que la regresión 656 lineal de la línea 610 epipolar en la figura 6G debido a los cruces por cero 650a y 650b obtenidos de los píxeles 620 y 622 respectivamente.

45 Todos los cruces por cero se consideran en la regresión lineal. Sin embargo, algunos de los cruces por cero pueden descartarse previamente y no considerarse en el proceso de obtención de la pendiente de las líneas epipolares. Los puntos con alta dispersión en comparación con el resto de puntos que se utilizan para aplicar la técnica de regresión lineal se pueden identificar y excluir de este proceso para obtener una estimación de la pendiente más precisa o eliminar valores atípicos. Por ejemplo, en la figura 6J, el cruce por cero 650a originado por el píxel superior 620 y el cruce por cero 650b originado por el píxel inferior 622 de la imagen 602 epipolar pueden descartarse al computar la regresión 658 lineal (obteniendo una línea epipolar con una pendiente similar a la pendiente obtenida para la línea 610 epipolar de la figura 6G), ya que los píxeles superiores 620 e inferiores 622 no están conectados al resto de píxeles que originan la línea 612 epipolar (en este caso los píxeles superiores 620 e inferiores 622 puede haber sido causado por aberraciones del lente 102 principal).

55 Una vez que se computan las segundas derivadas 644, se decide si definen líneas epipolares válidas o no válidas. Para este proceso, se pueden descartar algunos valores de las segundas derivadas correspondientes a algunos píxeles, como se explicó anteriormente. Se aplica una regresión lineal a los cruces por cero válidos para calcular sus pendientes correspondientes. Por el contrario, para todas las líneas epipolares identificadas como no válidas, no es necesario realizar más cálculos.

Es posible utilizar métodos heurísticos, análisis morfológico, inteligencia artificial o cualquier otro método para determinar de antemano a partir de imágenes epipolares si las líneas epipolares son válidas o no válidas y evitar cálculos posteriores, ni siquiera calcular las pendientes para las líneas epipolares que sabemos de antemano que no son válidos.

- 5 Cuando se aplica la regresión lineal a las líneas epipolares válidas detectadas, también se puede calcular una estimación de error. Como ejemplo, la suma de las distancias entre los puntos de la línea epipolar (es decir, los cruces por cero) y la línea de regresión estimada final se puede utilizar como error (es decir, la suma de los valores absolutos de las distancias entre la línea epipolar calculada y los puntos utilizados para calcular esta línea epipolar). Sin embargo, se puede definir cualquier otro tipo de cómputo de errores.
- 10 Se puede utilizar un umbral de error máximo para descartar (y no considerar en el resto del algoritmo) una línea epipolar. Con ese fin, si el error computado es mayor que el umbral de error máximo, la línea epipolar se considera no válida, y si el error computado es menor que el umbral de error máximo, la línea epipolar se considera válida.

15 Una imagen 400 epipolar horizontal puede contener varias líneas epipolares (hasta líneas epipolares de Nix), como se muestra, por ejemplo, en las líneas 510 epipolares válidas detectadas de una imagen 502 epipolar horizontal en la figura 5. De manera similar, una imagen epipolar vertical puede contener varias líneas 511 epipolares (hasta líneas epipolares de Nly). La figura 7A muestra un ejemplo de una imagen 700 epipolar horizontal que incluye dos líneas (710 y 712 en la figura 7E) epipolares diferentes. La figura 7A muestra las regresiones (756, 758) lineales de los cruces por cero 650 correspondientes a ambas líneas epipolares. Este ejemplo representa un escenario más realista que los presentados en las figuras 4 y 6, ya que el patrón de luz ahora es producido por un objeto con un cierto tamaño en lugar de infinitesimal. Es por eso que la alta intensidad (píxeles blancos) registrada por el sensor 106 de imagen ocupa varios microlentes (l_x) en la figura 7A.

20 Los valores "I" de intensidad de los píxeles de la imagen 700 epipolar se muestran en la tabla 720 de la figura 7B, mientras que la tabla 730 de la figura 7C representa los valores de la segunda derivada. El método para considerar si un píxel de una imagen epipolar está etiquetado o detectado como píxel 731 de borde o no comprende encontrar aquellos píxeles (px, l_x) con un valor negativo de la segunda derivada que tienen en su lado derecho o izquierdo un píxel con una segunda derivada positiva (píxeles resaltados de la figura 7C). Alternativamente, como se muestra en la figura 7D (la misma tabla de la figura 7C, valores de la segunda derivada), un píxel de una imagen epipolar puede etiquetarse como píxel 731 de borde para esos píxeles (px, l_x) con un valor positivo de la segunda derivada que tienen en su lado derecho o izquierdo un píxel con una segunda derivada negativa (píxeles resaltados de la figura 7D).

30 Una vez que se han detectado los píxeles 731 de borde, se explica a continuación el procedimiento para identificar líneas epipolares válidas dentro de imágenes epipolares, haciendo referencia al ejemplo de la figura 7E (correspondiente a los valores de la segunda derivada y píxeles 731 de borde de la figura 7C):

- Para cada píxel l_x ($l_x=1$ a $l_x=11$) en las imágenes epipolares horizontales (o l_y en las imágenes epipolares verticales) ubicadas en la fila central px (o py para imágenes epipolares verticales) correspondiente a los píxeles centrales ($px = 5$) y etiquetado como píxel 731 de borde (píxeles de borde A y J):

35 1- Buscar píxeles etiquetados como píxel 731 de borde en las posiciones vecinas de arriba ($l_x, px-1$), ($l_x + 1, px-1$), ($l_x-1, px-1$): píxel de borde B (para la primera iteración que comienza desde el píxel de borde A) y el píxel de borde K (para la primera iteración que comienza desde el píxel de borde J).

40 2- Si se encuentra un píxel 731 de borde, actualizar l_x y px con las coordenadas del nuevo píxel 731 de borde (coordenadas del píxel de borde B: $l_x=4, px=4$ en la primera iteración comenzando desde el píxel de borde A; coordenadas del píxel de borde K: $l_x=8, px=4$ en la primera iteración comenzando desde el píxel de borde J) y se repite el paso 1 (los siguientes píxeles de borde encontrados: píxeles de borde D, F y H cuando se itera desde el píxel de borde A; píxeles de borde M y O cuando iterando desde el píxel de borde J, donde el píxel de borde Q no se considera parte de la línea epipolar ya que está ubicado en l_x+2 con respecto al píxel de borde O). De lo contrario, continuar con el paso 3.

45 3- Buscar píxeles etiquetados como borde en las posiciones vecinas de abajo ($l_x, px+1$), ($l_x+1, px+1$), ($l_x-1, px+1$): píxel de borde C (cuando la iteración comienza desde píxel de borde A) y píxel de borde L (cuando se itera desde el píxel de borde J).

50 4- Si se encuentra un píxel 731 de borde, actualizar l_x y px con las coordenadas del nuevo píxel 731 de borde (coordenadas del píxel de borde C: $l_x=4, px=6$ en la primera iteración comenzando con el píxel de borde A; coordenadas del píxel de borde L: $l_x=8, px=6$ en la primera iteración comenzando desde el píxel de borde J) y repetir el paso 3 (los siguientes píxeles de borde encontrados: píxeles de borde E, G e I al iterar desde el píxel de borde A; píxeles de borde N y P cuando iterando desde el píxel de borde J, donde el píxel de borde R no se considera parte de la línea epipolar ya que está ubicado en l_x-2 con respecto al píxel de borde P). De lo contrario, continuar con el siguiente paso.

55 El resultado de este proceso iterativo es una primera línea 710 epipolar (correspondiente al píxel A del borde central) y una segunda línea 712 epipolar (correspondiente al píxel J del borde central). La primera línea 710 epipolar está

formada por 9 píxeles de borde (H, F, D, B, A, C, E, G, I). La segunda línea 712 epipolar está formada por 7 píxeles de borde (O, M, K, J, L, N, P).

5 • Dependiendo del número de píxeles 731 de borde detectados para un determinado l_x en la fila central p_x en el proceso iterativo descrito, la línea epipolar puede considerarse válida o no válida. El número de píxeles 731 de borde detectados debe ser al menos la altura en píxeles (es decir, 9 en el ejemplo de la figura 7E) de la imagen epipolar. La primera línea 710 epipolar cumple con este criterio ya que tiene 9 píxeles; sin embargo, la segunda línea 712 epipolar no cumple con este criterio ya que está formada por solo 7 píxeles. Los píxeles extremos ($p_x=1$, $p_x=9$) pueden omitirse para reducir los efectos de las aberraciones ópticas del lente 102 principal (en ese caso, el número de píxeles 731 de borde detectados debe ser al menos la altura en píxeles de la imagen epipolar menos 2, es decir, 7 píxeles en la figura 7E). En este último ejemplo, ambas líneas (710, 712) epipolares de la figura 7E se considerarían válidas.

10 • Dependiendo de la consistencia de la dirección de cada píxel 731 de borde dentro de una línea epipolar, la línea epipolar puede considerarse válida o no válida (apuntando hacia la misma dirección dentro de la imagen epipolar). Por ejemplo, en la primera línea 710 epipolar, comenzando desde el punto del borde central A, todos los píxeles del borde de arriba (B, D, F, H) están ubicados en las posiciones l_x-1 o l_x , mientras que los píxeles del borde de abajo (C, E, G, I) están en posiciones $l_x + 1$ o l_x , formando una dirección consistente para la primera línea 710 epipolar. Lo mismo se aplica a la segunda línea 712 epipolar, comenzando desde el punto del borde central J todos los píxeles del borde de arriba (K, M, O) están ubicados en las posiciones l_x-1 o l_x , mientras que los píxeles del borde de abajo (L, N, P) están en las posiciones $l_x + 1$ o l_x .

15 En un ejemplo, estos dos criterios (número de píxeles 731 de borde detectados para una línea epipolar y consistencia de la dirección) deben cumplirse para que la línea epipolar se considere válida.

20 Por tanto, y de acuerdo con la figura 7E, para considerar válida una línea epipolar:

• En primer lugar, los píxeles de la imagen epipolar correspondientes a un borde de objeto (es decir, píxeles 731 de borde) se detectan utilizando los valores de la segunda derivada.

25 • Luego, se obtiene un conjunto de píxeles de borde conectados que forman una línea epipolar. Se pueden emplear diferentes algoritmos, tal como el proceso de iteración previamente definido en la figura 7E, comenzando desde los píxeles de borde A y J (las flechas muestran las direcciones de búsqueda de iteración hacia arriba y hacia abajo buscando píxeles de borde adyacentes para obtener un conjunto de píxeles de borde conectados formando la línea epipolar).

30 • Con base en uno o más criterios (por ejemplo, número de píxeles de borde en el conjunto y dirección coherente de los píxeles de borde en el conjunto), la línea epipolar se considera válida o no válida.

35 Cuando se detecta una línea epipolar válida, se computa la pendiente de esta línea. Este valor de pendiente se puede convertir directamente en un valor de profundidad, ya que existe una relación directa entre pendientes y valores de distancia. Una vez calculadas las pendientes de las líneas epipolares analizadas, el resultado del método es un mapa de profundidad bidimensional disperso que contiene los valores de profundidad (dz) de los bordes de los objetos de la escena capturados por una cámara plenóptica. Las coordenadas (dx , dy) del mapa de profundidad indican la posición lateral de los puntos del objeto correspondientes (es decir, las coordenadas bidimensionales del mundo del objeto), mientras que los valores de profundidad (dz) representan la profundidad de las coordenadas correspondientes (dx , dy) en el mundo de los objetos. La figura 8 ilustra los bordes de un mapa de profundidad dispersa que muestra tres objetos (802, 804, 806) a diferentes profundidades, en donde el color negro representa ningún valor de profundidad asignado y cuanto más blanco es el valor de profundidad, más lejos está el objeto en la escena.

40 El método puede comprender una etapa adicional para generar un mapa de profundidad dispersa considerando la pendiente de las líneas epipolares obtenidas en la etapa anterior. El mapa de profundidad dispersa se obtiene asignando valores de profundidad (dz) de objetos en el mundo real a los bordes calculados antes (dx , dy).

45 La entrada para la generación del mapa de profundidad dispersa son dos matrices (una matriz 516 de profundidad horizontal y una matriz 517 de profundidad vertical que relaciona los valores de profundidad calculados (dz) y la posición correspondiente en la estructura de campo de luz (p_x , p_y , l_x , l_y). La entrada para la generación del mapa de profundidad dispersa también pueden ser las dos matrices de pendientes (512, 513) obtenidas en un paso anterior. En este caso, primero se obtiene un mapa de pendiente dispersa y la conversión a profundidad es solo aplicado a este mapa de pendiente bidimensional, reduciendo así los requisitos computacionales.

50 La matriz 516 de profundidad horizontal se obtiene analizando las imágenes epipolares horizontales mientras que la matriz 517 de profundidad vertical se obtiene a partir de las imágenes epipolares verticales. El tamaño de cada una de estas matrices en el estado de la técnica (516, 517) es $N_{px} \times N_{py} \times N_{lx} \times N_{ly}$, siendo N_{px} y N_{py} el número de píxeles por microimagen en las direcciones horizontal y vertical, y N_{lx} y N_{ly} el número de microlentes horizontales y verticales.

- 5 Cuando se realiza la regresión lineal de una línea epipolar, es posible obtener solo un valor de pendiente. En consecuencia, el tamaño de las matrices de entrada de esta etapa se puede reducir en gran medida para almacenar solo el valor de profundidad/pendiente para cada línea epipolar producida por el método de regresión lineal, de modo que el tamaño de la matriz de profundidad horizontal sea $N_{py} \cdot N_{ly} \cdot N_{lx}$ (una imagen epipolar horizontal puede contener hasta N_{lx} líneas epipolares) y el tamaño de la matriz de profundidad vertical es $N_{px} \cdot N_{lx} \cdot N_{ly}$ (una imagen epipolar vertical puede contener hasta N_{ly} líneas epipolares).
- Las dos matrices de profundidad/pendiente pueden incluir solo los puntos analizados en las imágenes epipolares horizontal-central y vertical-central (o cualquier otra imagen epipolar), de manera que los tamaños de las matrices sean $N_{lx} \times N_{ly}$ para ambas.
- 10 Muchos puntos de estas matrices pueden no tener un valor de profundidad calculado ya que no se ha detectado una línea epipolar válida en la posición correspondiente dentro de las imágenes epipolares (no se detectaron bordes).
- Se puede utilizar una etapa de combinación para asignar cada valor de profundidad obtenido (dz) a las coordenadas bidimensionales del mundo del objeto (dx , dy), obteniendo el mapa de profundidad (dx , dy , dz) en función de la pendiente calculada de los puntos y considerando las coordenadas (px , py , lx , ly) de los puntos (es decir, la posición sobre el sensor). Como se puede observar en las figuras 1 a 4, un punto 110 de objeto produce diferentes patrones sobre el sensor así como diferentes pendientes en las líneas epipolares. Por lo tanto, calculando la pendiente y conociendo la posición (px , py , lx , ly) sobre el sensor, es posible encontrar la posición mundial correspondiente (dx , dy) para cada línea epipolar detectada.
- 15 Se pueden obtener varios valores de dz diferentes para el mismo par (dx , dy), ya que un solo borde en el mundo del objeto puede originar varias líneas epipolares afectadas por ruidos ligeramente diferentes, aberraciones, oclusiones o errores de cuantificación, produciendo líneas epipolares con diferentes pendientes y, por lo tanto, diferentes profundidades. Además, algunas de las líneas epipolares horizontales y algunas verticales pueden producir valores de dz ligeramente diferentes.
- 20 Todos los valores de profundidad redundantes (diferentes valores de dz) se combinan en un único mapa de profundidad para reducir el ruido estadístico al generar el mapa de profundidad bidimensional (un único valor de dz por coordenadas dx , dy).
- 25 Cuando se obtienen todos los valores de profundidad (dz) en el mapa de profundidad (dx , dy , dz), se pueden obtener varios valores de profundidad (dz) para la misma posición (dx , dy). Por tanto, se pueden aplicar varios métodos para obtener el valor final. A modo de ejemplo y no a modo de limitación, se puede aplicar la media aritmética o la mediana o cualquier otra técnica de promediado (con o sin ponderaciones ponderadas) a todos los valores de profundidad (todos los valores de dz) que se obtuvieron para la misma posición del mapa de profundidad (dx , dy).
- 30 Debido a esta redundancia se reduce el ruido estadístico, mejorando la calidad del mapa de profundidad. Además, la estimación del error calculada para las líneas epipolares se puede considerar para elegir el valor de profundidad final (dz) de una determinada posición del mapa de profundidad (dx , dy); por ejemplo, eligiendo el valor con menor error entre todos los valores que se proyectaron en la misma posición (dx , dy) (por ejemplo, considerando como error la suma de todas las distancias entre la línea epipolar y los píxeles que originaron esa línea epipolar o cualquier otra medida).
- 35 Cuantos más valores de profundidad se obtengan, más preciso es el mapa de profundidad ya que se incrementa la redundancia, minimizando los errores de las medidas de profundidad. No obstante, se puede disminuir la redundancia considerada por los algoritmos, reduciendo también la calidad del mapa de profundidad, con el fin de reducir los requisitos computacionales y la complejidad de la implementación.
- 40 El mapa de profundidad dispersa bidimensional se genera directamente tomando una cierta vista plenóptica de la estructura epipolar horizontal (o vertical) que contiene los valores de profundidad estimados, es decir, tomando todos los puntos con px y/o py establecidos en ciertos píxeles (normalmente el píxel central, ya que es la vista menos afectada por las aberraciones). En este caso, la complejidad computacional se reduce a costa de tener mapas de profundidad menos redundantes y posiblemente más dispersos (solo un valor de profundidad para cada microlente).
- 45 La resolución del mapa de profundidad puede ser mayor que el número total de microlentes para aprovechar la precisión de subpíxeles obtenida en la etapa de detección de borde de cruce por cero.
- Dado que los valores de pendiente solo se pueden obtener en los bordes de la imagen epipolar identificados (en las líneas epipolares), el mapa de profundidad disperso obtenido en la etapa anterior contiene muchas posiciones vacías (dx , dy), no solo para un gran número de píxeles, sino también para un gran número de microlentes en los que la homogeneidad del mundo real no produce bordes en las imágenes epipolares. Los valores de profundidad correspondientes para todas estas posiciones vacías se pueden obtener considerando los valores de profundidad de las posiciones vecinas. Este procedimiento para obtener un mapa de profundidad denso se puede llamar "relleno de mapa de profundidad" y se beneficia de muchas técnicas anteriores de relleno de imágenes.
- 50
- 55

Se pueden aplicar varias técnicas para rellenar el mapa de profundidad dispersa con el fin de obtener un mapa de profundidad denso. Por consiguiente, algunos de estos enfoques son meros ejemplos pero no limitaciones: técnicas de crecimiento, división y fusión de regiones y/o agrupamiento, así como algunos otros enfoques conocidos en la técnica anterior para el procesamiento de imágenes. Además, se pueden emplear métodos de regularización para completar el mapa de profundidad.

La figura 9 representa un ejemplo de un mapa de profundidad denso que muestra tres objetos (902, 904, 906) a diferentes profundidades. Esta figura 9 muestra en niveles de gris un mapa de profundidad denso del mapa de profundidad dispersa generado en la figura 8, en donde el color negro representa ningún valor de profundidad asignado y cuanto más blanco es el valor de profundidad, más lejos está el objeto en la escena.

El método de la presente invención se ejecuta en un dispositivo móvil electrónico, tal como un teléfono inteligente, una tableta o un ordenador portátil. Las figuras 10A, 10B y 10C ilustran diferentes ejemplos de dispositivos 1000 móviles electrónicos con una unidad de procesamiento o medios 1004 de procesamiento configurados para ejecutar el método con el fin de obtener mapas de profundidad a partir de imágenes 1002 capturadas por una cámara 100 plenóptica.

Con el fin de obtener mapas de profundidad en tiempo real en dispositivos móviles, se recomienda encarecidamente implementar el presente método de una manera extremadamente eficiente. Para lograrlo, es posible aprovechar los múltiples núcleos incluidos en los actuales procesadores 1006 multinúcleo (figura 10A), incluso en procesadores desde dispositivos móviles, creando varios subprocesos de ejecución de algoritmos de tal forma que cada uno de ellos se encargue de realizar diferentes operaciones.

Se crean dos subprocesos de ejecución de CPU para que una primera CPU 1008a (en la figura 10B) ejecute los pasos descritos (véase figura 5) para las imágenes 502 epipolares horizontales mientras que una segunda CPU 1008b se encarga de realizar las mismas operaciones en las imágenes 503 epipolares verticales.

Pueden usarse técnicas computacionales más avanzadas para aumentar la eficiencia computacional. Por ejemplo, se puede utilizar una unidad de procesamiento de gráficos (GPU 1010 en la figura 10C), incluso las incluidas en dispositivos móviles, ya que una GPU incluye varios cientos o miles de núcleos capaces de ejecutar operaciones simultáneamente. En consecuencia, cada imagen epipolar (vertical y horizontal) se procesa simultáneamente en un núcleo diferente de una GPU 1010 para acelerar aún más la ejecución del algoritmo.

Como ya se explicó, el proceso para transformar los patrones encontrados en imágenes epipolares en información de profundidad requiere la aplicación de algunas técnicas de procesamiento de imágenes. Las imágenes epipolares contienen líneas epipolares, que son píxeles conectados que forman una línea (varios píxeles del sensor correspondientes al mismo punto en el mundo de los objetos). Las pendientes de estas líneas epipolares están directamente relacionadas con la forma del patrón iluminado sobre los microlentes y, lo que es más importante, con la profundidad correspondiente de ese punto en el mundo de los objetos. Resumiendo el proceso, los patrones que se encuentran en las imágenes epipolares, las líneas epipolares, brindan información sobre la profundidad de los objetos en el mundo de los objetos reales. Estas líneas pueden detectarse utilizando algoritmos de detección de bordes y sus pendientes pueden medirse mediante técnicas de regresión lineal. Tanto la detección de bordes como la regresión lineal se pueden realizar con precisión de subpíxeles. Así, el paso de detección de bordes comprende calcular una segunda derivada espacial en dimensiones x y y para imágenes epipolares horizontales 400 y verticales 402, respectivamente, para cada píxel de las imágenes epipolares y detectando los cruces por cero de las segundas derivadas espaciales, determinando las pendientes de las líneas epipolares válidas con precisión de subpíxeles y aplicando una línea que se ajuste a los cruces por cero de aquellos puntos que forman los bordes detectados. En adelante, un píxel se considera un píxel de borde cuando se encuentra un cruce por cero de la segunda derivada espacial dentro del área de dicho píxel.

La pendiente de cada línea epipolar da un valor que, convenientemente procesado como se describe, proporciona la profundidad real del punto en el mundo del objeto que produjo tal patrón. Una de las principales ventajas de esta metodología para la estimación de profundidad es que todos los cálculos se pueden realizar solo en aquellos píxeles del sensor donde se han detectado bordes del mundo del objeto, lo que representa una porción relativamente pequeña de la imagen, evitando realizar cálculos en cada píxel del sensor.

Sin embargo, debido a fenómenos ópticos y físicos, a grandes distancias de la cámara, donde los rayos de luz de cualquier punto del mundo de los objetos llegan todos casi en paralelo entre sí al lente de la cámara (cualquiera que sea el campo del punto del objeto que creó esos rayos), se requiere una cantidad relativamente grande de variación de distancia para producir solo una pequeña variación en la pendiente detectada de las líneas epipolares en una cámara plenóptica, es decir, dos objetos diferentes colocados a diferentes distancias pueden producir prácticamente la misma pendiente (ya que el sensor necesitaría una precisión infinita para detectar esta variación; en otras palabras, solo píxeles infinitesimales y un mundo libre de ruido producirían cambios en la pendiente). En estas situaciones, es extremadamente importante obtener una estimación precisa de la pendiente, de lo contrario, la profundidad estimada diferirá significativamente de la profundidad real de los objetos en el mundo. Nótese en este sentido que, como los sensores están discretizados en píxeles finitos, siempre se introduce un error al medir pendientes en una cámara plenóptica.

Las figuras 11A-11C muestran tres ejemplos diferentes de imágenes 400 epipolares horizontales capturadas por una cámara 100 plenóptica cuando se mide la profundidad de un objeto ubicado a una cierta distancia de la cámara. La diferencia entre las tres imágenes 400 epipolares horizontales es que, incluso si la distancia es exactamente la misma en los tres casos, el hecho de que la mano del operador de la cámara 100 plenóptica produjera variaciones de posición ligeramente diferentes debido a vibraciones provocó tres lecturas ligeramente diferentes. Idealmente, las tres líneas epipolares deberían ser exactamente iguales. Sin embargo, debido al ruido intrínseco al sistema y dado que la luz está iluminando píxeles ligeramente diferentes, los cruces por cero de las líneas epipolares también se encuentran en posiciones ligeramente diferentes, produciendo tres valores de pendiente diferentes después de aplicar una regresión 1102, 1104 y 1106 lineal, respectivamente. Suponiendo que la figura 11A y la figura 11C son los casos extremos (medidas de pendiente máxima y mínima posibles para un determinado objeto a una cierta distancia de la cámara), se puede definir una región de incertidumbre entre estos dos límites. Un ejemplo de esta región se muestra en la figura 11D, en la que se representa la imagen 400 epipolar horizontal de la figura 11B con su correspondiente regresión lineal y las correspondientes a los valores extremos. Solo un valor exacto de pendiente dentro de este rango de incertidumbre produciría el valor exacto correcto de profundidad.

Debido a esta falta de precisión, la precisión de las profundidades estimadas en una cámara 100 plenóptica disminuye a medida que aumenta la profundidad. La figura 12 muestra la incertidumbre de una determinada cámara 100 plenóptica al medir la distancia de un objeto ubicado a una distancia relativamente corta y una distancia relativamente grande. Esta figura 12 muestra dos ejemplos de las distribuciones 1202 y 1204 estadísticas (o posibles fluctuaciones en el eje vertical, originadas por lecturas ligeramente diferentes de la cámara con un objeto siempre ubicado a la misma distancia) de las medidas de profundidad obtenidas con una cámara 100 plenóptica al calcular la profundidad de dos objetos ubicados a diferentes distancias (las dos líneas 1206 y 1208 verticales punteadas) de la cámara (el eje horizontal aumenta con la distancia). El eje horizontal representa la profundidad o distancia desde la cámara 100 plenóptica mientras que el eje vertical es el número de mediciones que proporcionaron el mismo valor de profundidad para un punto en el mundo de objetos siempre a la misma distancia. La curva 1202 de la izquierda muestra la distribución al medir un objeto ubicado a una distancia relativamente corta mientras que la curva 1204 de la derecha del gráfico representa la distribución obtenida con la misma cámara pero ahora con el objeto ubicado a una distancia mayor. Como se ilustra en la figura 12, la incertidumbre de la profundidad estimada aumenta con distancias mayores, aumentando también la dispersión de las profundidades estimadas. Esa es la razón por la que una sola cámara plenóptica proporciona buenas estimaciones de profundidad solo para distancias relativamente cortas.

De acuerdo con otra realización de la presente invención, se proporciona un método y sistema para la obtención de un mapa de profundidad que potencia las capacidades proporcionadas por una sola cámara plenóptica, reduciendo drásticamente la incertidumbre de la medición de grandes distancias introducidas por la variación de pendiente baja proporcionada mediante el uso de la información de las cámaras adicionales que forman un sistema de vista múltiple. Esta realización mejorada se puede aplicar a configuraciones de cámara múltiples y muy complejas que incluyen un gran número de cámaras en una configuración similar a un arreglo, como se describirá más adelante. Usando una o más cámaras 1304 convencionales en combinación con una o más cámaras 100 plenópticas, a una cierta separación D (típicamente unos pocos centímetros cuando se usan cámaras en dispositivos móviles), se reduce la incertidumbre de la medición de grandes distancias.

Las figuras 13A-13E muestran algunos ejemplos de posibles configuraciones del sistema multivista usando una cámara 100 plenóptica y varias cámaras 1304 convencionales a una cierta distancia D aunque las distancias entre la cámara 100 plenóptica y cada cámara 1304 convencional pueden variar para cada cámara 1304 convencional. Las cámaras pueden incorporarse, por ejemplo, como cámaras posteriores de un dispositivo 1300 móvil, tal como un teléfono inteligente o una tableta. La realización de la figura 13A representa una cámara 100 plenóptica alineada horizontalmente con una cámara 1304 convencional separada una distancia D. La figura 13B muestra una cámara 100 plenóptica alineada horizontalmente con una cámara 1304 convencional y también alineada verticalmente con una segunda cámara 1304 convencional. La realización de la figura 13C muestra una cámara 100 plenóptica alineada horizontalmente con dos cámaras 1304 convencionales, una en el lado derecho y otra en el lado izquierdo, y una tercera cámara 1304 convencional alineada verticalmente. El ejemplo de la figura 13D representa una cámara 100 plenóptica alineada horizontal y verticalmente con dos cámaras 1304 convencionales en cada dimensión. Finalmente, la figura 13E muestra un dispositivo 1300 móvil que incorpora una cámara 100 plenóptica alineada horizontalmente con dos cámaras 1304 convencionales, una en el lado derecho y otra en el lado izquierdo.

Sin embargo, en aras de la claridad y la simplicidad, pero no como una limitación, se describe en este documento un método mejorado para una realización en la que el sistema de captura de imágenes o la configuración de la cámara comprende una cámara 100 plenóptica y una cámara 1304 convencional colocada en una cierta distancia D sobre el eje horizontal, como se muestra en el ejemplo de la figura 13A. Una vez que se explica el método para una cámara convencional, sería sencillo replicar el método en una realización usando varias cámaras 1304 convencionales. El método mejorado también se explicará para imágenes 400 epipolares horizontales, aunque también pueden emplearse imágenes 402 epipolares verticales.

El método mejorado para obtener información de profundidad se basa en el procedimiento de estimación de profundidad explicado anteriormente para una cámara 100 plenóptica, en donde se computa la pendiente de las líneas epipolares formadas en las imágenes epipolares y finalmente se relaciona con una cierta profundidad en el mundo del objeto. Sin embargo, una de las principales contribuciones del sistema y método mejorado es el uso de la imagen 2D

proporcionada por una o más cámaras 1304 convencionales como una vista adicional de la cámara 100 plenóptica. Esta nueva vista plenóptica se encuentra a una cierta distancia del cámara 100 plenóptica. Adaptándola adecuadamente con las técnicas requeridas, la imagen 2D se puede utilizar para ampliar drásticamente la línea de base del sistema plenóptico. Por consiguiente, las líneas epipolares de la cámara 100 plenóptica también pueden extenderse con la nueva vista plenóptica proporcionada por la cámara 1304 convencional. Este procedimiento de extensión se usa para mejorar la precisión al medir la pendiente de una línea epipolar.

La información de la cámara 1304 convencional se utiliza para medir las pendientes con mayor precisión. Sin embargo, para utilizar esta información adicional, es necesario encontrar la equivalencia entre la imagen 2D de la cámara 1304 convencional y las vistas plenópticas de la cámara 100 plenóptica. Para lograr esto, se debe tener en cuenta la separación entre las cámaras así como las diferencias en el campo de visión, el tamaño de píxel, el tamaño del sensor, el tamaño de las microlentes, etc. El proceso se explica en la figura 14, en donde un punto ubicado a cierta distancia de una cámara 100 plenóptica y a partir de una cámara 1304 convencional ilumina un píxel de borde (cx, cy) 1402 en el sensor 1400 de imagen de una cámara 1304 convencional y varios píxeles y microimágenes 112 en el sensor 106 de imagen de la cámara 100 plenóptica. Esto permite incluir una línea adicional a la imagen 400 epipolar horizontal capturada por el sensor de imagen de la cámara plenóptica 106, siendo dicha línea adicional una línea 1406 de extensión del sensor 1400 de imagen de la cámara convencional. Esta línea 1406 de extensión se considera como una vista plenóptica adicional de la cámara 100 plenóptica. Como se ilustra en la figura 15, la ubicación de la línea 1406 de extensión está determinada por dos desplazamientos, una separación vertical B (directamente relacionada con la distancia D entre la cámara 100 plenóptica y la cámara 1304 convencional) y un desplazamiento horizontal H (mediante el cual se hacen coincidir los campos comunes de ambas cámaras). Además, también es necesario obtener la fila cy (correspondiente a la línea 1406 de extensión) de la imagen 1412 de cámara convencional que corresponde a la fila (py, ly) (que forma la imagen 400 epipolar) de la cámara 100 plenóptica. Todas estas equivalencias solo dependen de los parámetros intrínsecos de ambas cámaras y de su alineación (posiciones relativas). Los procedimientos para obtenerlos se explican a continuación.

Siguiendo el proceso que se acaba de describir, se utiliza una línea 1406 de extensión obtenida de la cámara 1304 convencional como vista extra de la cámara 100 plenóptica a cierta distancia D, como se explica en la figura 15. Así, la regresión 1506 lineal de la línea 1404 epipolar obtenida exclusivamente de la imagen 400 epipolar de la cámara 100 plenóptica se prolonga hasta la línea 1406 de extensión. Alrededor de este punto 1504 de intersección (que corresponde a un píxel (cx', cy') del sensor 1400 de cámara convencional) una región 1512 está definida para buscar el píxel de borde correspondiente (cx, cy) 1402 de la cámara 1304 convencional. Este píxel de borde (cx, cy) 1402 corresponde al mismo borde de objeto en el mundo que los píxeles que forman la línea 1404 epipolar de la imagen 400 epipolar. En este paso del proceso se realizan algoritmos de correspondencia para encontrar el píxel 1402 de borde entre todos los píxeles dentro de la ventana 1512, dicho píxel 1402 de borde corresponderá al píxel más similar a los píxeles que forman la línea 1404 epipolar.

Una vez que se ha encontrado el píxel 1402 de borde correspondiente, los píxeles 1404 de línea epipolar y el píxel 1402 de borde forman una línea 1408 epipolar extendida. Los píxeles de línea 1408 epipolar extendida se utilizan para realizar un nuevo procedimiento de regresión lineal para que obtenga una regresión 1508 lineal y una pendiente recalculada. Para calcular la nueva pendiente, se puede utilizar un procedimiento para calcular los bordes correspondientes a la línea 1408 epipolar extendida con precisión de subpíxeles, por ejemplo, obteniendo los cruces por cero de la segunda derivada de la línea 1408 epipolar extendida a lo largo de la dirección lx (o la dirección ly para las líneas epipolares verticales). La segunda derivada de los puntos de la cámara convencional en la dirección de la línea 1406 de extensión también se puede aplicar a lo largo de la dirección Cx .

Debido a la información proporcionada por la cámara convencional, la nueva pendiente de la regresión 1508 lineal asociada a la línea 1408 epipolar extendida está mucho más cerca de la pendiente 1514 ideal que se habría obtenido en un mundo libre de ruido y con píxeles infinitesimales, y mucho mejor que la primera estimación 1506 en la que solo se utilizaron los píxeles proporcionados por la cámara 100 plenóptica.

El procedimiento completo para mejorar la precisión de la estimación de profundidad de una cámara plenóptica se puede resumir en dos etapas, como se muestra en los diagramas de flujo de las figuras 16A y 16B. El procedimiento para extender las imágenes epipolares se describe solo para imágenes epipolares horizontales; sin embargo, es sencillo extender este análisis a un escenario en el que la cámara convencional se coloca en el eje vertical y se consideran las imágenes epipolares verticales en lugar de las imágenes epipolares horizontales:

- Una primera etapa 1600 correspondiente a la calibración de cámaras:
 - Determinar la separación B 1602 vertical donde las líneas 1406 de extensión de la imagen 1412 de cámara convencional deben agregarse en las imágenes 400 epipolares horizontales (la separación vertical B depende de la distancia D entre la cámara 100 plenóptica y la cámara 1304 convencional).
 - Determinar el desplazamiento horizontal H 1604 que se debe aplicar a las líneas 1406 de extensión de la cámara 1304 convencional cuando están incluidas en las imágenes 400 epipolares horizontales de la cámara 100 plenóptica (para igualar el campo común de ambas cámaras en la dirección horizontal).

- 5
- Obtener la relación 1606 entre las posiciones espaciales de la cámara plenóptica y las posiciones espaciales de la cámara convencional, teniendo en cuenta el campo de visión, el tamaño de los píxeles y la ubicación de ambas cámaras. En particular, cuando se aplica a imágenes 400 epipolares horizontales, se obtiene la relación entre la dimensión espacial l_y de la cámara 100 plenóptica y la dimensión espacial c_y de la cámara 1304 convencional de acuerdo con el campo de visión vertical, el tamaño de los píxeles y la ubicación de ambas cámaras.
- Una segunda etapa 1610 correspondiente a los cálculos de pendiente, mediante la cual se mejora la precisión de la pendiente. Para cada línea 1404 epipolar encontrada en las imágenes 400 epipolares horizontales de la cámara 100 plenóptica:
 - Calcular una primera estimación de la pendiente 1612 considerando solo los píxeles 1404 de línea epipolar, y la correspondiente regresión 1506 lineal, en la imagen 400 epipolar de la cámara 100 plenóptica.
 - Obtener 1614 la línea c_y (línea 1306 de extensión) de la imagen 1412 de cámara convencional que corresponde a la coordenada vertical (l_y , p_y) de la imagen 400 epipolar horizontal de la cámara 100 plenóptica. Esta fila c_y contiene el píxel 1402 de borde particular en la imagen de la cámara convencional correspondiente al mismo objeto en el mundo que los píxeles del borde de la línea epipolar en la imagen de la cámara plenóptica.
- 10
- Extender 1616 la imagen 400 epipolar colocando la línea 1406 de extensión de acuerdo con el desplazamiento horizontal H y la separación vertical B obtenida previamente en la primera etapa 1600.
 - Calcular 1618 el punto 1504 de intersección de la regresión lineal de la cámara plenóptica 1506 con la línea 1406 de extensión y obtener el píxel correspondiente (c_x' , c_y') de la imagen de la cámara convencional.
- 15
- Definir 1620 una ventana 1512 (que puede ser una ventana 1512 unidimensional o una ventana 1512' bidimensional, como se muestra respectivamente en las figuras 17A y 17B; en el ejemplo de la figura 17B, la ventana 1512' bidimensional está formada por tres filas de píxeles: fila c'_y-1 , fila c'_y y fila c'_y+1) alrededor del píxel (c_x' , c_y') (punto 1504 de intersección) donde se buscará el píxel 1402 de borde de la línea 1406 de extensión de la cámara convencional.
- 20
- Aplicar 1622 un método de correspondencia para encontrar el píxel 1402 de borde (c_x , c_y) en la imagen 1412 de cámara convencional que mejor coincida con el punto del objeto del mundo representado por la línea 1404 epipolar. El píxel 1402 de borde es un punto en la imagen 1412 de la cámara convencional que corresponde al mismo punto en el mundo de los objetos que el punto de borde representado por la línea 1404 epipolar.
- 25
- Finalmente, aplicar una técnica 1624 de regresión lineal a la línea 1408 epipolar extendida formada por los píxeles 1404 de línea epipolar y el píxel 1402 de borde, y calcular la pendiente de la regresión 1508 lineal. Para ello, en una realización una regresión lineal se aplica al conjunto de puntos formado por los cruces por cero de la segunda derivada de los píxeles de borde de la línea 1404 epipolar de la imagen 400 epipolar horizontal en la dirección l_x (o l_y para imágenes epipolares verticales) y el cruce por cero de la segunda derivada de los valores alrededor del píxel 1402 de borde a lo largo de la dirección c_x (a lo largo de la línea 1406) de la imagen 1412 de cámara convencional.
- 30
- Es sencillo extender el procedimiento a otras configuraciones de cámara multivista, tales como las que se muestran en las figuras 13B-13E o la figura 18 que representan otra posible realización del sistema multivista de acuerdo con la presente invención, que comprende un arreglo bidimensional de cámaras 100 plenópticas y cámaras 1304 convencionales. Como ejemplo, pero no como limitación, si una cámara plenóptica está rodeada por cuatro cámaras convencionales (una en la parte superior, una en la parte inferior, una a la izquierda y otra a la derecha de la cámara plenóptica, como en el ejemplo de la figura 13D), las imágenes epipolares horizontales 400 y verticales 402 proporcionadas por la cámara 100 plenóptica se pueden extender agregando líneas 1406 de extensión horizontales en la parte superior e inferior de las imágenes epipolares horizontales 400 y verticales 402, dichas líneas 1406 de extensión correspondientes a líneas horizontales/verticales (para imágenes epipolares horizontales/verticales) de la imagen 1412 capturada por el sensor 1400 de imagen de cámara convencional; por lo tanto, se puede obtener una redundancia adicional al tener cuatro en lugar de una sola cámara convencional, lo que reduce los efectos del ruido al aumentar el número de mediciones. Las imágenes epipolares horizontales se extienden con líneas 1406 de extensión (líneas horizontales colocadas a la distancia adecuada en la parte inferior como en las figuras 14 y 15 para la primera cámara 1304 convencional en el lado derecho, y líneas horizontales colocadas a la distancia adecuada en la parte superior para la segunda cámara 1304 convencional en el lado izquierdo), líneas proporcionadas por las cámaras convencionales derecha e izquierda, mientras que las imágenes epipolares verticales se extienden con líneas 1406 de extensión verticales (líneas verticales de la imagen 1412 de cámara convencional ubicadas a la distancia adecuada) de las cámaras convencionales superior e inferior. Para extender adecuadamente las imágenes (400, 402) epipolares con múltiples cámaras 1304 convencionales, se debe calcular la separación vertical B y el desplazamiento horizontal H para cada cámara 1304 convencional individual adicional dependiendo de su ubicación y sus parámetros físicos.
- 35
- 40
- 45
- 50
- 55
- Se tiene en cuenta que, dado que la separación entre la cámara plenóptica y la convencional es mucho mayor que la línea de base de una sola cámara plenóptica, a distancias relativamente grandes se requieren variaciones de profundidad más pequeñas para producir cambios de pendiente notables en la nueva línea 1408 epipolar extendida,

formada por un píxel 1402 de borde de la cámara convencional además del conjunto de píxeles de borde de la línea 1404 epipolar de la cámara plenóptica. La nueva pendiente de la regresión 1508 lineal de la línea 1408 epipolar extendida se puede utilizar para determinar una profundidad muy precisa de los puntos del borde en el mundo de los objetos. Esto implica que la incertidumbre en la pendiente de la línea 1404 epipolar de la cámara plenóptica (o de manera similar la incertidumbre definida por las pendientes 1102 y 1106 alrededor de la pendiente 1104 en el ejemplo de las figuras 11A-11D) se reduce drásticamente cuando se identifica el píxel 1402 de borde de la cámara convencional, generando la línea 1408 epipolar extendida (y la correspondiente regresión 1508 lineal), como se muestra en las figuras 14 y 15. Esta drástica reducción de la incertidumbre se muestra en las figuras 19A y 19B. La figura 19A representa la distribución 1902 obtenida para una cierta profundidad grande usando solo una cámara plenóptica, mientras que la figura 19B muestra la distribución 1904 obtenida para el mismo objeto pero considerando la cámara plenóptica y también la información de la cámara convencional. Se observa cómo la dispersión (y por lo tanto la incertidumbre) es mucho más estrecha cuando se utiliza la información proporcionada por la cámara convencional.

La primera etapa 1600 del proceso para extender las imágenes epipolares requiere el conocimiento de algunos parámetros físicos del sistema óptico con el fin de utilizar la imagen 2D de la cámara convencional como una vista plenóptica adicional. En primer lugar, la separación B entre la línea 1406 de extensión de la cámara convencional y el centro (en particular, la línea 1516 horizontal central) de la imagen 400 epipolar está directamente relacionada con la distancia D entre la cámara 1304 convencional y la cámara 100 plenóptica (relacionado con la línea de base entre las dos cámaras). Dado que cada fila 1510 en una imagen 400 epipolar corresponde a una vista plenóptica diferente de la cámara plenóptica y las vistas están distribuidas a lo largo de la abertura de la cámara plenóptica, es sencillo obtener la posición de la línea 1406 de extensión de la imagen 1412 de cámara convencional. (solo una vista adicional en una separación vertical B en píxeles correspondiente a la distancia D entre la cámara 100 plenóptica y la cámara 1304 convencional, véase la figura 15). A modo de ejemplo, y sin limitar la generalidad de la invención, una separación típica entre la cámara plenóptica y la cámara convencional puede ser de alrededor de 5 cm, el número de vistas plenópticas (igual al número de píxeles debajo de cada microlente 106) suele ser alrededor de 10x10 y la abertura 200 de una minicámara puede ser típicamente de alrededor de 1 mm (por lo tanto, el paso entre vistas plenópticas es de 1/10 mm en este ejemplo). Así, la separación B entre la línea 1516 horizontal central de la imagen 400 epipolar horizontal (entre la vista plenóptica central de la cámara plenóptica) y la línea 1406 de extensión de la cámara convencional es la fracción entre la línea base de 5 cm (la separación entre ambas cámaras) y el paso entre las vistas plenópticas de la cámara plenóptica:

$$B = \frac{D}{\text{paso_de vistas_plenópticas}} = \frac{50mm}{\frac{1}{10} mm} = 500 \text{ vistas plenópticas}$$

Dado que cada vista plenóptica se representa como una fila en las imágenes epipolares, la separación B en la dimensión vertical entre la fila central (línea 1516 horizontal central) de la imagen 400 epipolar y la línea 1406 de extensión de la cámara convencional sería en el ejemplo 500 píxeles (teniendo en cuenta que la separación d entre filas 1510 consecutivas de una imagen epipolar es un píxel), como se muestra en la figura 15.

Otro parámetro físico requerido en la primera etapa 1600 del procedimiento de extensión epipolar es el desplazamiento horizontal H, que se calcula para asegurar que la parte común del campo de ambas cámaras coincida para considerar la línea 1406 como una cámara adicional alineada con las vistas plenópticas (las cámaras de 10x10 en el ejemplo anterior). Por lo general, el sensor de la cámara convencional recibe luz de partes del mundo real ligeramente diferentes a las de la cámara plenóptica debido a la separación entre ellas. Además, el campo de visión (FOV) de ambas cámaras puede ser diferente (incluso si en el caso más habitual tiene sentido diseñar ambos FOV exactamente iguales), lo que implica capturar también diferentes partes del mundo real.

El desplazamiento horizontal H del sensor de imagen de la cámara convencional debe tener en cuenta todos estos factores para estimar correctamente la pendiente de la regresión 1508 lineal de la línea 1408 epipolar extendida. Como se puede observar en la figura 15, un desplazamiento horizontal incorrecto H aplicado a la línea 1406 de extensión produciría una estimación incorrecta de la pendiente de la línea 1408 epipolar extendida. En una realización preferida, el desplazamiento horizontal H se calcula experimentalmente como se muestra en la figura 20A. El experimento consiste en colocar un punto 2002 luminiscente alineado en el eje 2004 óptico de la cámara 100 plenóptica. Además, este punto 2002 luminiscente se coloca a una distancia 2006 del lente 102 principal de la cámara 100 plenóptica que corresponde al plano conjugado del arreglo 104 de microlentes de la cámara 100 plenóptica. En ese escenario específico, todas las vistas plenópticas capturan exactamente el mismo patrón y las imágenes epipolares contienen líneas 430 epipolares perfectamente verticales (como se muestra en la figura 4D y la figura 20B).

Por otro lado, dado que la cámara 1304 convencional está separada una cierta distancia D de la cámara 100 plenóptica, el punto 2002 ilumina un cierto píxel 2010 que es diferente del centro del sensor de la cámara 1304 convencional (ya que están asumiendo que el sensor 1400 de imagen de la cámara convencional está alineado con el eje óptico de la cámara convencional). La figura 20B muestra la imagen 400 epipolar horizontal de la cámara

plenóptica que contiene la línea 430 epipolar completamente vertical así como la línea horizontal de píxeles 2012 ($c_x=1, \dots, c_x=c_{x_m\acute{a}x}$) del sensor 1400 de imagen de la cámara convencional que contiene el píxel 2010 iluminado. Nótese que si ambas cámaras están alineadas horizontalmente, esta línea 2012 correspondería a la línea central del sensor, de lo contrario se debe realizar una simple búsqueda para encontrar el píxel 2010 sobre el sensor 1400 de imagen de la cámara 1304 convencional. Una vez ubicados el píxel 2010 y la línea 2012 horizontal, el desplazamiento horizontal H que se debe aplicar para obtener el píxel 2010 perfectamente alineado con la línea 430 epipolar se puede calcular directamente en la figura 20B.

Una vez determinada la separación vertical B y el desplazamiento horizontal H de la línea 2012 de la imagen 1400 de cámara convencional, se establece una relación entre las coordenadas espaciales de la cámara plenóptica (l_x, l_y) y las coordenadas espaciales de la cámara convencional (c_x, c_y) deben encontrarse para extender adecuadamente las líneas epipolares que se encuentran dentro de una imagen (400, 402) epipolar. En una imagen 400 epipolar horizontal (p_y, l_y) las filas representan la misma posición espacial l_y capturada por las diferentes vistas plenópticas a lo largo de la dimensión p_x (en el ejemplo de la figura 14, la primera fila de la imagen 400 epipolar corresponde a $p_x=1$, y la última fila corresponde a $p_x=9$). Por tanto, para elegir correctamente la línea 1406 de extensión de la imagen 1412 de cámara convencional que debe añadirse a la imagen 400 epipolar horizontal, se debe encontrar una equivalencia entre l_y y la dimensión vertical c_y de la imagen 1400 de cámara convencional.

En una realización preferida, esta relación se puede encontrar usando una configuración similar a la presentada en la figura 20A pero en lugar de usar un punto 2002 luminiscente, un patrón más grande que ilumina más píxeles 2010 de la cámara 1304 convencional y se utilizan más microlentes de la cámara 100 plenóptica. La correspondencia entre (c_x, c_y) y (l_x, l_y) se obtiene identificando el tamaño del patrón S_c producido sobre el sensor de la cámara convencional (el número de píxeles iluminados) y el tamaño del patrón S_p producido sobre el del arreglo de microlente de la cámara plenóptica (el número de microlentes que se han iluminado). Comparando ambos tamaños se obtiene la relación entre ambas coordenadas espaciales, sin embargo también se debe considerar el desplazamiento horizontal H. En consecuencia, cuando una cámara 1304 convencional se alinea sobre el eje horizontal de una cámara 100 plenóptica, se pueden obtener las siguientes relaciones:

$$c_x = l_x \cdot S_{cx}/S_{px} + \text{desplazamiento_hor}$$

$$c_y = l_y \cdot S_{cy}/S_{py} + \text{desplazamiento_ver}$$

Donde S_{px} y S_{cx} son los tamaños en la dimensión x de los patrones producidos en la cámara 100 plenóptica y la cámara 1304 convencional, respectivamente. De manera similar, S_{py} y S_{cy} tienen el tamaño en la dimensión y de los patrones producidos en la cámara 100 plenóptica y la cámara 1304 convencional, respectivamente. El parámetro desplazamiento_hor es el desplazamiento horizontal H obtenido previamente. Por otro lado, el parámetro desplazamiento_ver es cero si la cámara 100 plenóptica y la cámara 1304 convencional están perfectamente alineadas en el eje horizontal. De lo contrario, se debe utilizar un experimento similar al explicado en las figuras 20A y 20B (pero con imágenes 402 epipolares verticales) para obtener un desplazamiento vertical para compensar la desalineación en el eje vertical.

En al menos una realización, los procedimientos para encontrar estas relaciones se realizan utilizando precisión de subpíxeles al calcular los tamaños S_{px} , S_{cx} , S_{py} y S_{cy} , ya que es posible utilizar los cruces por cero de la segunda derivada de los bordes de los patrones para calcular los tamaños correspondientes. De manera similar, el desplazamiento horizontal H se puede obtener con precisión de subpíxeles alineando el cruce por cero del píxel 2010 con la línea 430 epipolar.

En este punto del proceso, se ha completado la primera etapa 1600 del diagrama de la figura 16A y se han calibrado las diferentes cámaras. Luego se inicia la segunda etapa 1610 para mejorar la precisión de las pendientes de la cámara plenóptica.

Para cada línea 1404 epipolar detectada en la imagen epipolar, se debe encontrar el píxel 1402 de borde correspondiente en la línea 1406 de extensión de la imagen de la cámara convencional. En al menos una realización, la pendiente de cada línea 1404 epipolar se calcula primero mediante técnicas de regresión lineal considerando solo los puntos identificados como bordes en el mundo del objeto (utilizando los correspondientes cruces por cero de la segunda derivada calculados con precisión de subpíxeles en la cámara plenóptica) obteniendo la línea 1506 de regresión lineal. Después de eso, la línea 1406 de extensión correspondiente debe identificarse a partir de la imagen 1412 de cámara convencional utilizando el razonamiento explicado anteriormente: la relación entre l_y y c_y y el desplazamiento vertical, y la relación entre l_x y c_x y el desplazamiento horizontal H.

Entonces, considerando el desplazamiento horizontal H y la separación B, se extiende la línea 1406 y se calcula la intersección 1404 con la línea 1306 de extensión de la cámara convencional. Se debe aplicar la relación entre c_x e l_x para obtener el píxel correspondiente (c_x', c_y'). Este píxel se utilizará para determinar la región 1412 dentro de la línea 1306 de extensión de la cámara convencional en la que buscaremos el píxel 1302 de borde. Este punto corresponde al mismo borde en el mundo de los objetos que esos píxeles de la línea epipolar 1304 plenóptica.

En al menos una realización, se usa una ventana 1512 unidimensional con un número arbitrario de píxeles como la región considerada dentro de la línea 1406 para buscar el píxel 1402 de borde de la cámara 1304 convencional que corresponde al mismo borde en el mundo de los objetos que los píxeles que forman la línea 1404 epipolar de la cámara plenóptica.

- 5 También es posible, en al menos una realización, utilizar una ventana 1512' bidimensional considerando líneas adyacentes a la línea c_y 1406 de la imagen 1412 de cámara convencional. En al menos una realización, el ancho (y la altura en una ventana 1512' bidimensional) de esta ventana 1512 se elige de acuerdo con la dispersión obtenida al estimar una cierta profundidad con solo la cámara plenóptica (véase líneas 1102 y 1106 de puntos en la figura 11D y curvas 1202 y 1204 de dispersión en la figura 12). Esta ventana puede ser asimétrica, es decir, el número de píxeles
10 considerados a la izquierda del píxel (c_x' , c_y') puede ser diferente al número de píxeles considerados a la derecha de dicho píxel.

Una vez que se define una ventana 1512 1D o una ventana 1512' 2D de un cierto número de píxeles dentro de la imagen 1412 de la cámara 1304 convencional alrededor del píxel 1504, es necesario identificar qué píxel de los varios
15 candidatos posibles es el píxel 1402 de borde, es decir, el píxel en la cámara 1304 convencional que ha sido generado por la misma fuente de luz en el mundo de los objetos, que corresponde al píxel que mejor coincide con los píxeles del borde que forman la línea 1404 epipolar (es decir, el píxel más similar). Se pueden utilizar varias técnicas para hacer coincidir las imágenes de las dos cámaras (SAD-Suma de diferencias absolutas, correlaciones, entropías o cualquier otra medida cuantitativa de desviación). Una posible realización utiliza como referencia para la comparación en la cámara plenóptica el píxel central de la línea 1404 epipolar ya que los rayos que producen dicho píxel central
20 atraviesan la parte central del lente 102 principal de la cámara 100 plenóptica y atraviesan los microlentes 104 correspondientes en ángulos bajos, sufriendo las más bajas aberraciones.

Una forma sólida de realizar la identificación es hacer coincidir patrones en lugar de comparar píxeles individuales, de modo que una determinada parte del mundo de los objetos se pueda identificar más fácilmente. Estos patrones se
25 pueden formar tomando píxeles adyacentes del píxel central que corresponden a partes adyacentes del mundo de objetos. En una cámara plenóptica, estas partes adyacentes del mundo de objetos son muestreadas por los microlentes adyacentes. Como ejemplo, supongamos que el píxel 1402 de borde que se encuentra en la cámara convencional corresponde al mismo objeto del mundo que el píxel central de la línea 1404 epipolar, que se encuentra en el centro ($p_x=5$, $p_y=5$) de la microimagen (i_x , i_y). Luego, para identificar correctamente el píxel 1402, se define un patrón que se emparejará alrededor de este píxel central de la línea 1404 epipolar considerando los cuatro píxeles
30 centrales circundantes ($p_x=5$, $p_y=5$) de las cuatro microimágenes ($i_x + 1$, i_y), ($i_x - 1$, i_y), (i_x , i_y), (i_x , $i_y + 1$), (i_x , $i_y - 1$) adyacentes. De esta manera, solo se consideran los puntos menos aberrados de las cinco vistas ($p_x=5$, $p_y=5$ en cada microimagen con 9x9 píxeles por microimagen) centrales plenópticas. Una vez que se ha definido el patrón de referencia de la cámara 100 plenóptica, se definen patrones del mismo número de píxeles en el sensor 1412 de imagen de la cámara 1304 convencional. En particular, se define un patrón para cada píxel de la línea 1406 de extensión
35 dentro del ventana 1512. Este patrón de referencia podría tener un tamaño diferente o incluso ser un patrón 1D.

En una realización, los píxeles de la cámara 1304 convencional son mucho más pequeños que los microlentes de la cámara 100 plenóptica, de modo que un único microlente integra luz procedente de una parte mucho más grande del mundo de objetos que la integrada por un único píxel de la cámara 1304 convencional (véase figura 15). En tal
40 escenario, los patrones definidos en el sensor de imagen de la cámara convencional deben incluir más píxeles para identificar adecuadamente la misma región espacial (como una región del mundo que proyecta su luz sobre un único microlente 105 de tamaño relativamente grande de una cámara 100 plenóptica proyectará su luz sobre un número mayor de píxeles relativamente pequeños de una cámara 1304 convencional).

Si los sensores en ambas cámaras fueran del mismo tamaño (por ejemplo, ambos sensores de 10 megapíxeles), y el número de píxeles por microimagen 112 fuera, por ejemplo, 10x10 (100 píxeles) en el ejemplo anterior, tendríamos
45 que igualar el patrón formado por estos cinco píxeles (el píxel de referencia (i_x , i_y) y sus cuatro vecinos conectados) en la cámara plenóptica con patrones de 500 píxeles (un patrón por cada píxel dentro de la región 1512) en la cámara convencional. Cada uno de estos patrones de la cámara convencional está formado por cinco cuadrados de 10x10 píxeles cada uno. De esta manera, se mejora la robustez de la solución frente a una mera comparación de 1 píxel frente a 10x10 píxeles en la cámara convencional. Sin embargo, en el ejemplo anterior, si el sensor de imagen de la cámara 1304 convencional es un sensor de 40 megapíxeles, tendríamos que hacer coincidir un patrón de cinco píxeles
50 centrales de la vista plenóptica central en la cámara plenóptica con un patrón de 2000 píxeles (cinco cuadrados de 20x20 píxeles del sensor 1412 de imagen de cámara convencional) en la cámara convencional, mejorando en esta vista la precisión del resultado de profundidad.

Los algoritmos de coincidencia de patrones (entropía, correlación, SAD, ...) finalmente producirán el píxel en la cámara
55 1304 convencional que mejor coincida con el píxel central de la línea 1404 epipolar. Si, por ejemplo, estuviéramos usando la SAD (suma de diferencias absolutas) en el ejemplo anterior, el valor de intensidad de los cinco píxeles (i , j) del patrón de referencia en la cámara plenóptica se resta de, por ejemplo, la media de los valores de intensidad de los 500 píxeles (k , l) de patrones definidos alrededor de cada píxel candidato de la cámara convencional. La suma de todos los valores absolutos de estas sustracciones se computa para cada píxel candidato dentro de la ventana 1512,
60 obteniendo un valor único para cada píxel candidato dentro de la región 1512 de la cámara convencional. El píxel

finalmente elegido es el de menor diferencia (menor valor de SAD), y se utiliza para extender la línea epipolar de la cámara plenóptica.

Por lo tanto, aplicando algoritmos de correspondencia se encuentra el píxel 1402 de borde de la cámara convencional. Con este píxel adicional (que funciona como una vista plenóptica extra), la pendiente de la regresión lineal de la línea 1408 epipolar extendida se calcula aplicando una técnica de regresión lineal (o cualquier otro método de ajuste) al conjunto de puntos formado por la cruce por cero de la segunda derivada del píxel 1402 de borde encontrado en la imagen 1412 de cámara convencional y los cruces por cero de la segunda derivada de los píxeles que forman la línea 1404 epipolar encontrada en la imagen 400 epipolar de la cámara 100 plenóptica. Nótese que tanto la regresión lineal como la detección de bordes realizadas en la imagen 400 epipolar de la cámara 100 plenóptica y en la imagen 1412 de la cámara 1304 convencional pueden calcularse usando precisión de subpíxeles. En al menos una realización, cuando se calcula la regresión lineal de la línea 1408 epipolar extendida, se puede asignar más importancia al cruce por cero de la segunda derivada espacial del píxel 1402 de borde detectado en la cámara 1304 convencional que a los cruces por cero de la segunda derivada de los píxeles de borde de la línea 1404 epipolar detectados en la cámara 100 plenóptica (por ejemplo, asignando un peso del cruce por cero del píxel 1402 de borde valores más altos que el peso asignado al resto de los puntos de cruce por cero de la línea 1404 epipolar).

En al menos una realización, la cámara convencional puede diseñarse para que sea ópticamente equivalente a una vista plenóptica. Idealmente, esto implicará que el desplazamiento horizontal H es cero y las coordenadas espaciales (l_x, l_y) de las vistas plenópticas son directamente equivalentes a las de la cámara convencional (c_x, c_y) . Este sistema puede estar compuesto por una cámara plenóptica y una cámara convencional que es funcionalmente equivalente a una vista plenóptica trasladada, es decir, el número de píxeles de la cámara convencional es igual al número de microlentes de la cámara plenóptica y al tamaño de los píxeles de la cámara convencional tiene el mismo tamaño que los microlentes de la cámara plenóptica. Además, la distancia 2016 entre el lente 102 principal de la cámara 100 plenóptica y el arreglo 104 de microlentes es la misma que la distancia 2014 entre el lente 2020 principal y el sensor 1400 de imagen de la cámara 1304 convencional, así como ambas distancias focales que también son iguales para ambas cámaras. En esta realización, la relación entre los píxeles de las cámaras convencionales y los microlentes de las cámaras plenópticas es sencilla (relaciones entre (l_x, l_y) y (c_x, c_y)) y solo la distancia D entre ambas cámaras y la diferencia en el campo capturado por ambas cámaras se debe considerar antes de extender las líneas epipolares (la cámara convencional puede capturar diferentes partes de la escena que la cámara plenóptica por lo que se debe calcular el desplazamiento horizontal H , por ejemplo, de acuerdo con el experimento de las figuras 20A y 20B comentado anteriormente).

Esta invención también mejora el rendimiento de reenfoque que se puede lograr usando solo un par estéreo o solo una cámara plenóptica. Como se indicó anteriormente, un requisito para poder estimar las profundidades en cualquier sistema de formación de imágenes es que las áreas de interés de una escena deben estar enfocadas, de lo contrario, la información de diferentes ubicaciones espaciales en el mundo se mezcla en el sensor. En una cámara plenóptica, el rango de profundidad en el que se enfoca una escena aumenta de manera efectiva ya que la apertura se divide en múltiples vistas equivalentes (con aperturas muy pequeñas y, por lo tanto, grandes profundidades de campo).

Este requisito también se aplica cuando queremos volver a enfocar la imagen a un cierto plano de profundidad. El proceso de reenfoque consiste esencialmente en mezclar adecuadamente las diferentes vistas capturadas de la escena en función del plano de profundidad que se desee enfocar para combinar las vistas como si los sensores del sistema estuvieran ubicados a la distancia conjugada de la profundidad deseada (por ejemplo, en el caso particular de una cámara plenóptica, la película virtual del arreglo de microlentes se puede propagar hacia adelante y/o hacia atrás para obtener "películas virtuales" antes o más allá del plano del arreglo de microlentes). Por lo tanto, cuanto más diferentes sean las vistas, más realista se puede lograr el efecto de reenfoque. Otra posibilidad para realizar el efecto de reenfoque es desenfocar (por ejemplo con un filtro gaussiano) aquellas partes de la escena que no están ubicadas en el mismo plano de profundidad que la profundidad deseada (en realidad esto es desenfocar haciendo borrosas las áreas a ciertas profundidades conocidas que deseamos que estén fuera de foco). Esto se puede realizar directamente considerando el mapa de profundidad conocido de la escena.

A partir de lo anterior, está claro que el rendimiento de reenfoque está directamente relacionado con la capacidad de estimación de profundidad, de modo que cuanto más fiables se estiman las distancias, mejor se consigue el efecto de reenfoque. Este principio es válido para cualquier sistema de formación de imágenes.

Las realizaciones mostradas en las figuras 13A-13E mejoran el rendimiento de la estimación de profundidad (reduciendo drásticamente la incertidumbre de los valores de profundidad estimados) y también permiten estimar distancias mayores en comparación con una sola cámara plenóptica o sistemas multivista convencionales. Además, también se mejora el rendimiento de reenfoque.

Dado que las cámaras plenópticas comienzan a perder precisión en las mediciones de profundidad a distancias relativamente pequeñas de la cámara, el proceso de reenfoque en las cámaras plenópticas ya no es efectivo a medida que aumenta la distancia (aproximadamente un metro para minicámaras de unos pocos mm para teléfonos inteligentes, no mucho más con lentes prácticos más grandes en cámaras de mano). Debido a la línea de base relativamente pequeña entre vistas plenópticas adyacentes de una cámara 100 plenóptica (una décima de mm en el ejemplo anterior), más allá de ciertas distancias, los rayos llegan a la cámara casi tan paralelos como llegarían si

vinieran de una distancia infinita; por tanto, es imposible diferenciar entre profundidades de digamos dos o tres metros y una distancia infinita, y el reenfoque se vuelve imposible más allá de dos o tres metros, funcionando bien sólo para distancias inferiores a un metro.

5 El sistema multivista descrito a partir de la figura 13A en adelante combina eficazmente el buen rendimiento de las cámaras 100 plenópticas para distancias pequeñas y también aprovecha las líneas de base más grandes de los sistemas de cámaras múltiples para distancias grandes. Para maximizar el efecto de reenfoque es necesario extender la profundidad de campo del sistema de vista múltiple tanto como sea posible.

10 Para lograr esto, en al menos una realización, la distancia hiperfocal de la cámara 1304 convencional está diseñada de manera que entre el infinito y una distancia T (usualmente la mitad de la distancia hiperfocal, por ejemplo un metro) todo tenga una nitidez aceptable, y de manera complementaria diseña la cámara 100 plenóptica de manera que pueda medir distancias con una incertidumbre aceptable desde la distancia T hasta distancias muy cercanas a la cámara, llegando incluso al límite para estimar la distancia de objetos que se acercan a la EFL (distancia focal efectiva) de la cámara 100 plenóptica (unos pocos milímetros).

15 De acuerdo con una realización, la presente invención permite obtener imágenes de escenas totalmente enfocadas incluso con muchos objetos diferentes ubicados a muchas distancias diferentes en el mundo de los objetos (desde objetos muy cercanos, aumentando gradualmente la distancia de los objetos y eventualmente alcanzando una distancia infinita para algunos objetos). A modo de ejemplo y nunca como limitación, una realización aplica métodos de reenfoque estándar de cámaras plenópticas para obtener imágenes enfocadas en los objetos que se ubican entre la cámara y T. La imagen totalmente enfocada final se compone tomando los objetos más nítidos de cada imagen reenfocada, mientras que para los objetos ubicados a distancias mayores que T simplemente los tomamos de la imagen de la cámara 1304 convencional ya que todos los objetos dentro del rango T e infinito están enfocados.

20 De manera similar, la presente invención también se puede usar para reenfocar una foto a un cierto plano de profundidad después de que se haya tomado la foto, superando sistemas anteriores como cámaras plenópticas únicas o pares estéreo. Como se explicó anteriormente, la presente invención aumenta la capacidad de discriminación de distancia de una cámara plenóptica mediante el uso de un sistema de vista múltiple. Esto permite calcular distancias con buena precisión para objetos ubicados a distancias mayores que T; así, es posible crear efectos de fotografía artística como, por ejemplo, enfocar un rango pequeño entre distancias de A hasta B (siendo A y B elegidos por el usuario para definir el rango enfocado). La imagen para distancias mayores que A y menores que B se puede generar desenfocando la cámara 100 plenóptica (si A y B son menores que T) o simplemente filtrando digitalmente (un filtro de desenfoque) la imagen de la cámara convencional, lo que es posible en nuestro sistema multivista ya que disponemos de un mapa de profundidad fiable incluso para largas distancias y podemos optar por desenfocar en un rango de distancias elegido por el usuario.

25 Todas las aplicaciones descritas también son compatibles con métodos de superresolución, que se pueden aplicar para aumentar la resolución de las vistas plenópticas. Además, es posible aumentar la baja resolución de las vistas plenópticas mezclando las imágenes de la cámara plenóptica y la cámara convencional a través de rutinas de procesamiento de imágenes adecuadas.

30 Además de las ventajas ya descritas, la presente invención puede evitar el uso de motores (o MEMS) necesarios para enfocar cámaras modernas de píxeles de micra (ya que el enfoque se realiza digitalmente: para grandes distancias con cámara convencional y para pequeñas distancias con cámara plenóptica), reduciendo costes, mejorando la confiabilidad y proporcionando imágenes totalmente enfocadas si los usuarios así lo desean después de que se haya tomado la foto. Sin embargo, esto no es una limitación de la presente invención y también puede ser utilizado por cámaras con enfoque variable.

La invención ofrece un mejor reenfoque para distancias cortas, para largas distancias y para áreas fuera de foco en pares estereoscópicos, así como mejores cálculos de distancia que permiten imágenes 3D de mayor calidad.

45 Por simplicidad y claridad, la descripción de la invención divulgada para mejorar el rendimiento para la estimación de profundidad de cámaras plenópticas individuales y de pares estéreo se ha explicado considerando solo una cámara convencional alineada horizontalmente con una cámara plenóptica. No obstante, un experto en el campo puede extender fácilmente esta invención a un sistema multivista formado por múltiples cámaras siendo al menos una de ellas una cámara plenóptica. Además, la distribución espacial de estas cámaras puede ser arbitraria (sin necesidad de costosos procedimientos de calibración o alineación) y solo deben realizarse cambios pequeños y sencillos para adaptar los métodos propuestos en este documento.

50 Como se explicó anteriormente, cuando una cámara convencional se alinea horizontalmente con una cámara plenóptica, las imágenes epipolares horizontales se pueden extender como se muestra en las figuras 14 y 15 considerando la separación entre ambas cámaras y la equivalencia entre ambos píxeles del sensor. Está claro que se puede aplicar un proceso similar a las imágenes epipolares verticales cuando ambas cámaras están alineadas en el eje vertical.

En al menos una realización, el procedimiento descrito anteriormente se puede extender trivialmente a un escenario donde una cámara plenóptica se alinea verticalmente con dos cámaras convencionales (una arriba y la otra debajo de

la cámara plenóptica) y también se alinea horizontalmente con dos cámaras convencionales más (una a la derecha y otra a la izquierda de la cámara plenóptica), como en el ejemplo de la figura 13D, componiendo un sistema formado por un total de cinco cámaras, una cámara plenóptica y cuatro cámaras convencionales a su alrededor. En esta realización, tanto las imágenes epipolares horizontales 400 como las verticales 402 se extienden con una línea 1406 de extensión arriba y otra línea 1406 de extensión debajo de la imagen 400 epipolar de la cámara 100 plenóptica usando las metodologías descritas anteriormente. En las imágenes 402 epipolares verticales, las líneas 1406 de extensión se toman de las imágenes 1412 de las cámaras 1304 convencionales alineadas verticalmente, y las imágenes 400 epipolares horizontales se extienden con las imágenes 1412 de las cámaras 1304 convencionales alineadas horizontalmente. La ventaja obvia de esta configuración, además de las ventajas para los objetos muy cerca de la cámara que trae la cámara plenóptica, es que mejora la línea de base, agregando capacidades para distancias mucho mayores y evitando las pesadas tareas de procesamiento de los sistemas multivista formados por cámaras convencionales para igualar patrones de un gran número de imágenes (cinco en la realización de ejemplo) con búsquedas mucho más fáciles de identificación de patrones (el procedimiento de ventana en la figura 17A y 17B). Otra realización simplificada puede usar una cámara plenóptica, una cámara convencional alineada horizontalmente y una segunda cámara convencional alineada verticalmente, como se muestra en la figura 13B, aumentando las líneas de base para las imágenes epipolares horizontales 400 y verticales 402.

La presente invención se puede aplicar a escenarios más generales. Supongamos una matriz de posiciones espaciales tal que en cada posición de la matriz se pueda colocar una cámara. La figura 18 muestra esta matriz donde las balas pueden ser sustituidas indistintamente por una cámara plenóptica, una cámara convencional o por ninguna cámara. La presente invención puede utilizar cualquier configuración posible de esta matriz. Como es evidente para un experto en el campo, solo es necesario encontrar los desplazamientos y equivalencias adecuados entre los píxeles del sensor para las diferentes cámaras convencionales y las cámaras plenópticas para extender adecuadamente las imágenes epipolares de las cámaras plenópticas con las nuevas vistas de las cámaras convencionales que ofrecen líneas de base más grandes. Las imágenes epipolares de las cámaras plenópticas se pueden extender en la dimensión vertical con tantas líneas como las cámaras convencionales estén en esa dirección. Por ejemplo, si se coloca una cámara plenóptica a la izquierda de cuatro cámaras convencionales alineadas horizontalmente, la imagen epipolar de la cámara plenóptica se ampliará con cuatro líneas adicionales, cada una correspondiente a las cuatro cámaras convencionales y separadas entre sí una distancia que depende de la separación física entre ellas. Al usar tales configuraciones, se incrementa la redundancia de las estimaciones de profundidad, así como la línea de base, lo que reduce la incertidumbre de las mediciones de profundidad para grandes distancias.

Además de esto, también es posible ampliar una imagen epipolar de una cámara plenóptica con varias cámaras plenópticas y/o varias cámaras convencionales. En estos casos, las imágenes epipolares se extienden no solo con líneas simples de cámaras convencionales sino con imágenes epipolares de diferentes cámaras plenópticas, como se muestra en el ejemplo de la figura 21, en donde la línea 1408 epipolar extendida está formada por una imagen 400a epipolar horizontal obtenida de la imagen 2112a capturada por un sensor 106 de imagen de una primera cámara 100a plenóptica, un píxel 1402 de borde obtenido de la imagen 1412 capturada por un sensor 1400 de imagen de una cámara 1304 convencional y una imagen 400b epipolar horizontal obtenida de la imagen 2112b capturada por un sensor 106b de imagen de una segunda cámara 100b plenóptica, en donde la línea 1406 de extensión tiene un desplazamiento horizontal H1 y una separación vertical B1 con respecto a la primera imagen 400a epipolar y la segunda imagen 400b epipolar tiene un desplazamiento horizontal H2 y una separación vertical B2 con respecto a la primera imagen 400a epipolar. A continuación, se calcula una regresión 1508 lineal de la línea 1408 epipolar extendida, y su pendiente correspondiente, para estimar una distancia de alta precisión. Esto se puede utilizar para aumentar aún más la precisión de las mediciones al tener varias primeras estimaciones de pendiente (las calculadas con las imágenes epipolares de una de las cámaras plenópticas) para identificar las regiones o ventanas 1512 del resto de imágenes 1412, 2112b de cámara donde se realiza la búsqueda de los píxeles de borde correspondientes tanto de la cámara 1402 convencional como del píxel central de la línea 1404b epipolar de la cámara plenóptica adicional. En realizaciones de sistemas multivista que incluyen más de una cámara plenóptica, las imágenes epipolares se amplían con la información del resto de imágenes de cámara plenóptica (como la imagen 2112b que se muestra en la figura 21), y la segunda etapa 1610 para identificar que cierta línea 1404a epipolar corresponde al mismo borde del mundo de objetos que la línea 1404b epipolar del resto de cámaras 100b plenópticas debe realizarse. En una realización, esta etapa es equivalente a la explicada anteriormente cuando las imágenes epipolares se amplían con la información de una cámara convencional pero considerando la vista plenóptica central de las cámaras 100b plenópticas adicionales.

En configuraciones donde las cámaras plenópticas no están alineadas con las cámaras convencionales ni en el eje horizontal ni en el vertical, todavía es posible extender las imágenes epipolares de las cámaras plenópticas. Simplemente, se deben aplicar desplazamientos verticales y horizontales para corregir estas desalineaciones y hacer coincidir adecuadamente las diferentes imágenes. Este es un procedimiento bien conocido en la literatura de vista múltiple cuando se deben rectificar varias vistas no alineadas.

En al menos una realización, el sistema de vista múltiple consiste en una matriz $M \times N$ de cámaras distribuidas equidistantemente de manera que las cámaras en las posiciones diagonales son todas ellas cámaras plenópticas como se muestra en la figura 18. Esta realización tiene la ventaja de permitir la posibilidad de extender las imágenes epipolares verticales y horizontales de cada cámara plenóptica con el mismo número de líneas de extensión de las

cámaras convencionales. Esto asegura la misma precisión en el eje horizontal y vertical extendiendo cada imagen epipolar de todas las cámaras plenópticas.

5 En al menos una realización, las cámaras se pueden distribuir irregularmente en la matriz. En al menos una realización, las cámaras se pueden distribuir formando cualquier tipo de figura (por ejemplo, un círculo) o cualquier otra distribución si las dimensiones de la matriz son lo suficientemente altas.

10 El procedimiento de extensión de imagen epipolar propuesto en esta invención se aplica para mejorar el proceso de estimación de profundidad de una cámara plenóptica con la ayuda de cámaras convencionales adicionales y eventualmente generar un mapa de profundidad más preciso. Por tanto, la metodología de la presente invención se puede aplicar a cualquier técnica de generación de mapas de profundidad existente para cámaras plenópticas con base en el análisis de imágenes epipolares y estimaciones de pendientes de las líneas epipolares asistidas con cámaras convencionales.

15 En todavía otra realización, que no puede tomarse como una limitación, se explica un procedimiento de generación de mapas de profundidad de ejemplo. La configuración del ejemplo consiste en una cámara plenóptica en el centro, una cámara convencional en su lado derecho y otra cámara convencional encima de la cámara plenóptica. Una vez que la cámara plenóptica ha capturado el campo de luz y las cámaras convencionales las imágenes 1412 correspondientes, se analizan las imágenes epipolares del campo de luz de la cámara plenóptica. El procedimiento de extensión de la imagen epipolar se aplica a las imágenes epipolares horizontales y verticales, obteniendo una estimación precisa de la pendiente para cada línea epipolar.

20 Es importante señalar que en una cámara plenóptica varias líneas epipolares pueden contener información del mismo punto en el mundo de los objetos. Así, todas las pendientes que corresponden a las mismas posiciones espaciales deben combinarse para aprovechar la información redundante y reducir aún más la incertidumbre, obteniendo una pendiente final única por posición espacial (dx , dy). Este mapa de pendientes se obtiene calculando todos los valores de pendiente en función de su posición en el sensor (px , py , lx , ly), más concretamente calculando los puntos dx y dy en el mundo de objetos que pertenecen a una o varias líneas epipolares proyectando los puntos de esas líneas epipolares (producidas por el mismo punto en el mundo de los objetos) en el mismo plano en el mundo de los objetos y asignando un valor de pendiente para cada par dx , dy .

25 Dependiendo de la configuración del sistema multivista, es posible que no todas las líneas epipolares de las cámaras plenópticas puedan mejorarse con las imágenes 1412 de las cámaras convencionales (por ejemplo, en un escenario donde el sistema contiene una cámara plenóptica alineadas horizontalmente con dos cámaras convencionales, solo las imágenes epipolares horizontales pueden aprovechar el uso de la imagen 1412 de cámara convencional como vistas adicionales). Por tanto, en al menos una realización, durante el proceso de combinación las pendientes calculadas en las imágenes epipolares que se han extendido pueden tener más peso que las pendientes obtenidas exclusivamente a partir de las imágenes epipolares de la cámara plenóptica. Es decir, cuando una pendiente obtenida exclusivamente de una cámara plenóptica, la línea epipolar se proyecta a una determinada posición del mapa de pendiente (dx , dy) y una pendiente cuya precisión se ha mejorado utilizando al menos una imagen 1412 de una cámara convencional. también se proyecta a la misma posición (dx , dy), el valor de pendiente final para dicha posición se puede calcular con cualquier valor medio aritmético ponderado o no. En el caso de que se aplique un promedio ponderado, en al menos una realización, las pendientes mejoradas tienen más peso ya que son más precisas.

30 Una vez obtenido el mapa de pendientes se aplica una relación entre pendiente y profundidad (que depende de los parámetros físicos y configuración de la cámara plenóptica) para obtener el mapa de profundidad. Dado que las líneas epipolares solo se encuentran en los bordes del mundo de los objetos, este mapa de profundidad no está completo y contiene posiciones (dx , dy) sin valores de profundidad (mapa de profundidad dispersa). Para obtener un mapa de profundidad denso, se pueden aplicar métodos de relleno. En la literatura se pueden encontrar diferentes estrategias de relleno, tales como las basadas en la segmentación de imágenes (crecimiento de regiones, división y fusión, y/o técnicas de agrupamiento), interpolación/aproximación de superficies a partir de puntos dispersos tridimensionales o reconstrucción tridimensional mediante técnicas estéreo multivista, para nombrar algunos. En una realización, los valores de profundidad correspondientes para todas estas posiciones vacías se pueden obtener considerando los valores de profundidad de las posiciones vecinas.

35 En una realización, la resolución del mapa de profundidad puede ser mayor que el número total de microlentes para aprovechar la precisión de subpíxeles obtenida en la etapa de detección de bordes. Como se dijo, los valores de pendiente solo se pueden obtener en los bordes de la imagen epipolar identificados (en las líneas epipolares) y el mapa de profundidad disperso obtenido en la etapa anterior contiene muchas posiciones vacías (dx , dy), no solo para un gran número de píxeles, sino también para un gran número de microlentes en los que la homogeneidad del mundo real no produce bordes en las imágenes epipolares. Así, en este mapa de profundidad de mayor resolución, se aplicarían igualmente las técnicas de relleno anteriores para obtener un mapa de profundidad denso.

40 La figura 5 anterior ilustra un diagrama de flujo para determinar el mapa de profundidad, en el que se detectan líneas epipolares válidas (pasos 510 y 511). Cuando se considera solo una cámara 100 plenóptica, las líneas epipolares válidas son las líneas 610 epipolares (figura 6A) obtenidas a partir de imágenes (400, 402) epipolares de dicha cámara 100 plenóptica, como se ilustra en la figura 22A. En este caso, el paso de detectar líneas (510, 511) epipolares válidas

comprende la detección 2202 de líneas epipolares de la cámara 100 plenóptica y considerar o asignar 2204 las líneas epipolares detectadas como líneas epipolares válidas (asumiendo que se consideran válidas).

Alternativamente, como se muestra en el diagrama de flujo de la figura 22B, el paso de detectar (510, 511) líneas epipolares válidas de la figura 5 puede incluir detectar 2212 líneas 1404 epipolares de las imágenes (400, 402) epipolares de la cámara 100 plenóptica y extendiendo 2214 estas líneas 1404 epipolares con información adicional incluida en imágenes capturadas por uno o más dispositivos de adquisición de imágenes adicionales, para obtener 2216 líneas 1408 epipolares extendidas y asignar 2218 estas líneas 1408 epipolares extendidas como las líneas epipolares válidas. De esta manera, la línea 1408 epipolar extendida de la figura 14 se consideraría la línea epipolar válida en los pasos 510 y 511 de la figura 5, y el paso en la figura 5 de determinar (512, 513) la pendiente de la línea epipolar válida incluiría el cálculo de la pendiente de la línea 1408 epipolar extendida.

Como se explicó anteriormente (por ejemplo, en el ejemplo de la figura 21), y se muestra en la figura 22B, la información 2222 adicional usada para extender las líneas epipolares de la cámara 100 plenóptica puede incluir píxeles 1402 de borde contenidos en imágenes 1412 capturadas por una o más cámaras 1304 convencionales. Alternativamente, la información 2224 adicional puede incluir líneas 1404b epipolares de imágenes 2112b capturadas por una o más cámaras 100b plenópticas adicionales. La información 2226 adicional también puede ser una combinación de píxeles 1402 de borde de una o más cámaras 1304 convencionales y líneas 1404b epipolares de una o más cámaras 100b plenópticas adicionales.

De acuerdo con una realización preferida, el método del sistema de vista múltiple se ejecuta en un dispositivo móvil electrónico, tal como un teléfono inteligente, una tableta o un portátil. Las figuras 23A, 23B y 23C ilustran diferentes realizaciones de dispositivos 2300 móviles electrónicos con una unidad de procesamiento o medios 2306 de procesamiento configurados para ejecutar el método con el fin de obtener mapas de pendiente y/o profundidad a partir de imágenes 2304 capturadas por el sistema 2302 de vista múltiple inventado.

Con el fin de obtener mapas de profundidad en tiempo real en dispositivos móviles, es muy recomendable implementar el presente método de una manera extremadamente eficiente. Para lograr esto, es posible aprovechar los múltiples núcleos incluidos en los actuales procesadores 2308 multinúcleo (figura 23A), incluso en procesadores desde dispositivos móviles, creando varios subprocesos de ejecución de algoritmos de tal forma que cada uno de ellos se encargue de realizar diferentes operaciones. En una realización, se crean dos subprocesos de ejecución de CPU para que una primera CPU 2310 (en la figura 23B) ejecute los pasos descritos para las imágenes epipolares horizontales mientras que una segunda CPU 2310b se encarga de realizar las mismas operaciones en las imágenes epipolares verticales. Se pueden utilizar técnicas computacionales más avanzadas para aumentar la eficiencia computacional. Por ejemplo, se puede utilizar una unidad de procesamiento de gráficos (GPU 2312 en la figura 23C), incluso las incluidas en dispositivos móviles, ya que una GPU incluye varios cientos o miles de núcleos capaces de ejecutar operaciones simultáneamente. En consecuencia, en una realización, cada imagen epipolar (vertical y horizontal) se extiende (si es posible dependiendo del sistema de vista múltiple) y se procesa simultáneamente en un núcleo diferente de una GPU 2312 para acelerar aún más la ejecución del algoritmo.

REIVINDICACIONES

1. Un método para obtener información de profundidad de un campo de luz, que comprende:
generar (502, 503) una pluralidad de imágenes (400, 402) epipolares a partir de un campo (501) de luz capturado por un dispositivo (100) de adquisición de campo de luz;
- 5 un paso (508, 509) de detección de bordes para detectar, en las imágenes (400, 402) epipolares, bordes de objetos en la escena capturados por el dispositivo (100) de adquisición de campo de luz;
para cada imagen (400, 402) epipolar, detectar (510, 511) líneas (610, 1408) epipolares válidas formadas por un conjunto de bordes;
determinar (512, 513) las pendientes de las líneas (610, 1408) epipolares válidas;
- 10 caracterizado porque la detección (510, 511) de líneas epipolares válidas comprende extender las líneas (1404) epipolares de las imágenes (400, 402) epipolares desde el dispositivo (100) de adquisición de campo de luz con información (1402, 1404b) adicional de imágenes (1412, 2112b) capturada por al menos un dispositivo (1304, 100b) de adquisición de imágenes adicional para obtener una línea (1408) epipolar extendida.
- 15 2. El método de la reivindicación 1, en donde las imágenes (400, 402) epipolares se extienden agregando, arriba y/o abajo, la información (1402, 1404b) adicional dependiendo de las posiciones relativas del al menos un dispositivo (1304, 100b) de adquisición de imágenes adicional al dispositivo (100) de adquisición de campo de luz.
3. El método de la reivindicación 2, en donde la información (1402, 1404b) adicional se agrega a una cierta distancia por encima y/o por debajo de las imágenes (400, 402) epipolares de acuerdo con desplazamientos horizontales (H_1 , H_2) y verticales (B_1 , B_2) computados previamente en un proceso de calibración.
- 20 4. El método de la reivindicación 2 o 3, en donde las imágenes (400) epipolares horizontales se extienden agregando la información (1402, 1404b) adicional del al menos un dispositivo (1304, 100b) de adquisición de imágenes adicional que está alineado horizontalmente con el dispositivo (100) de adquisición de campo de luz.
5. El método de cualquiera de las reivindicaciones 2 a 4, en donde las imágenes (402) epipolares verticales se extienden agregando la información (1402, 1404b) adicional del al menos un dispositivo (1304, 100b) de adquisición de imágenes adicional que está alineado verticalmente con el dispositivo (100) de adquisición de campo de luz.
- 25 6. El método de cualquier reivindicación precedente, en donde la información (2222) adicional comprende píxeles (1402) de borde contenidos en imágenes (1412) capturadas por al menos una cámara (1304) convencional como al menos un dispositivo de adquisición de imágenes adicional, en donde dichos píxeles (1402) de borde corresponden al borde del objeto representado por la línea (1404) epipolar.
- 30 7. El método de la reivindicación 6, que comprende determinar una región (1512, 1512') de búsqueda en las imágenes (1412) capturadas por las cámaras (1304) convencionales donde se buscan los píxeles (1402) de borde correspondientes a la línea (1404) epipolar.
8. El método de cualquier reivindicación precedente, en donde la información (2224, 2226) adicional comprende líneas (1404b) epipolares contenidas en imágenes (2112b) capturadas por al menos un dispositivo (100b) de adquisición de campo de luz adicional como el al menos un dispositivo de adquisición de imágenes adicional, en donde dichas líneas (1404b) epipolares corresponden al borde del objeto representado por la línea (1404) epipolar.
- 35 9. El método de la reivindicación 8, que comprende determinar una región de búsqueda en las imágenes (2112b) capturadas por los dispositivos (100b) de adquisición de campo de luz adicionales donde se encuentra el píxel del borde central de las líneas (1404b) epipolares de los dispositivos (100b) de adquisición de campo de luz adicionales correspondientes al borde del objeto representado por la línea (1404) epipolar.
- 40 10. El método de la reivindicación 6, que comprende:
calcular (1612) una regresión (1506) lineal de la línea (1404) epipolar desde el dispositivo (100) de adquisición de campo de luz;
obtener (1614) una línea (1406) de extensión a partir de la imagen (1412) capturada por la cámara (1304) convencional;
- 45 extender (1616) la imagen (400, 402) epipolar del dispositivo (100) de adquisición de campo de luz con la línea (1406) de extensión de la cámara (1304) convencional;
calcular (1618) el punto (1504) de intersección de la línea (1404) epipolar y la línea (1406) de extensión;
definir (1620) una región (1512, 1512') de búsqueda alrededor del punto (1504) de intersección.

11. El método de la reivindicación 10, que comprende además aplicar (1622) un proceso de correspondencia para encontrar el píxel (1402) de borde en la imagen (1412) de cámara convencional que coincide con el borde del objeto representado por la línea (1404) epipolar.
- 5 12. El método de cualquier reivindicación precedente, que comprende además obtener imágenes totalmente enfocadas de un sistema (2302) de vista múltiple que comprende el dispositivo (100) de adquisición de campo de luz y al menos una cámara (1304) convencional como al menos un dispositivo de adquisición de imágenes adicional, en donde el paso de obtener imágenes totalmente enfocadas comprende:
- para objetos ubicados a una distancia más allá de un umbral T del sistema (2302) de vista múltiple, obtener imágenes enfocadas de la al menos una cámara (1304) convencional;
- 10 para objetos ubicados a una distancia por debajo de un umbral T del sistema (2302) de vista múltiple, obtener imágenes reenfoadas desde el dispositivo (100) de adquisición de campo de luz;
- componer una imagen totalmente enfocada final tomando, para distancias por debajo del umbral T, los objetos más nítidos de las imágenes reenfoadas del dispositivo (100) de adquisición de campo de luz y, para distancias más allá del umbral T, tomando las imágenes enfocadas de la menos una cámara (1304) convencional.
- 15 13. El método de cualquier reivindicación precedente, que comprende además un paso de reenfoque de imágenes de un sistema (2302) de vista múltiple que comprende el dispositivo (100) de adquisición de campo de luz y al menos una cámara (1304) convencional como al menos un dispositivo de adquisición de imágenes adicional, en donde el paso de reenfoque de imágenes comprende:
- calcular un mapa de profundidad;
- 20 para objetos ubicados a una distancia por debajo de un umbral T del sistema (2302) de vista múltiple, usando imágenes reenfoadas del dispositivo (100) de adquisición de campo de luz;
- para objetos ubicados a una distancia más allá de un umbral T del sistema (2302) de vista múltiple:
- seleccionar un rango enfocado de distancias de la al menos una cámara (1304) convencional, y
- desenfocar objetos en la imagen colocados a una distancia más allá del rango de enfoque seleccionado.
- 25 14. Un dispositivo para generar un mapa de profundidad a partir de un campo de luz, que comprende medios (1004, 2306) de procesamiento configurados para llevar a cabo todos los pasos del método de cualquiera de las reivindicaciones 1 a 13.
15. Un producto de programa informático para generar un mapa de profundidad a partir de una imagen capturada por una cámara plenóptica, que comprende instrucciones de código informático que, cuando son ejecutadas por un procesador, hacen que el procesador realice el método de cualquiera de las reivindicaciones 1 a 13.
- 30

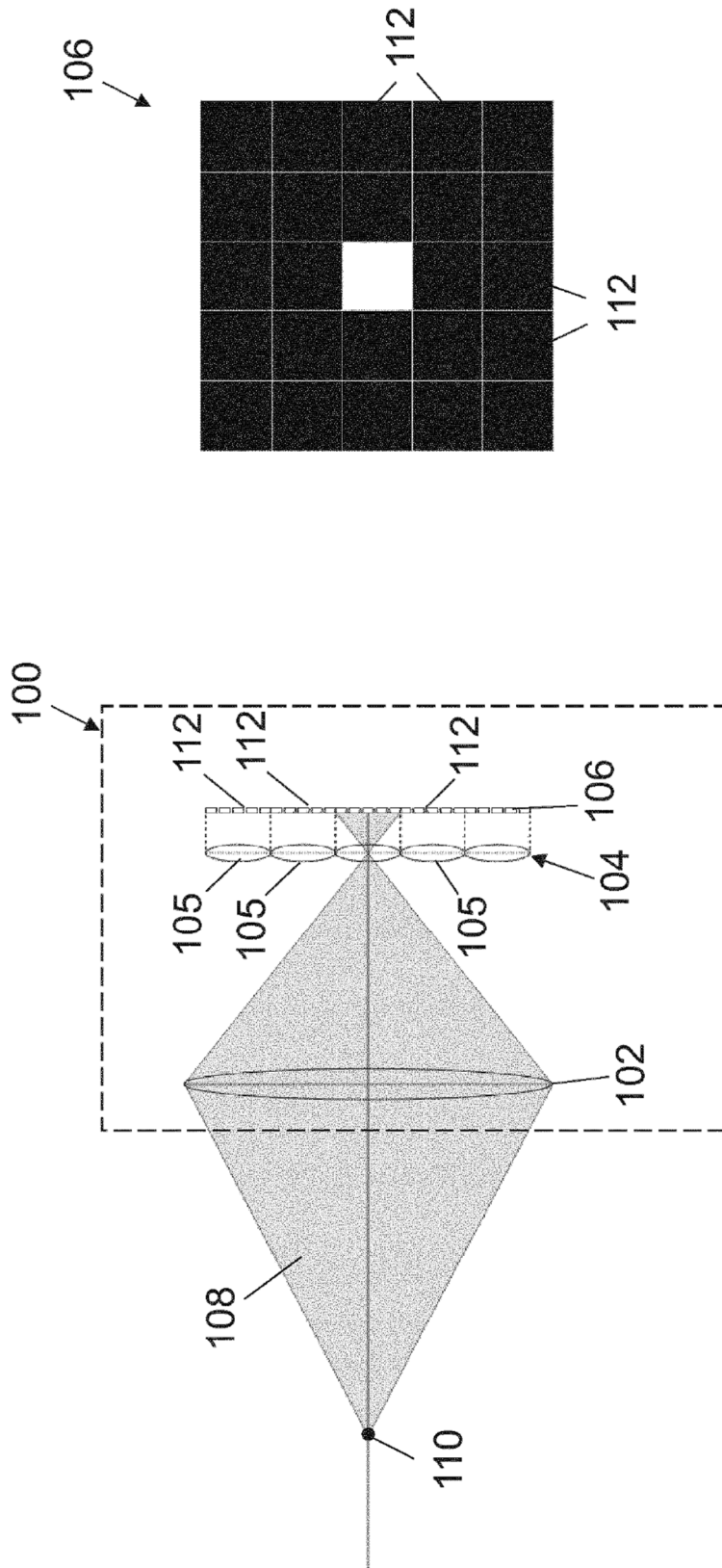


Fig. 1A

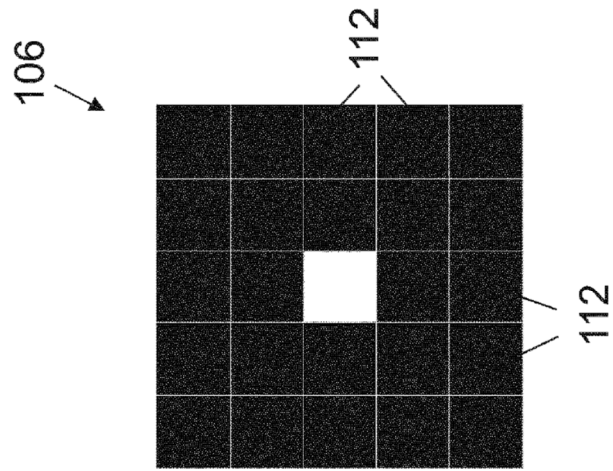


Fig. 1B

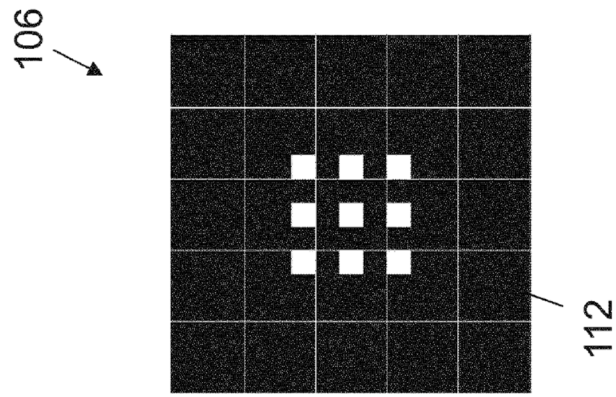


Fig. 2B

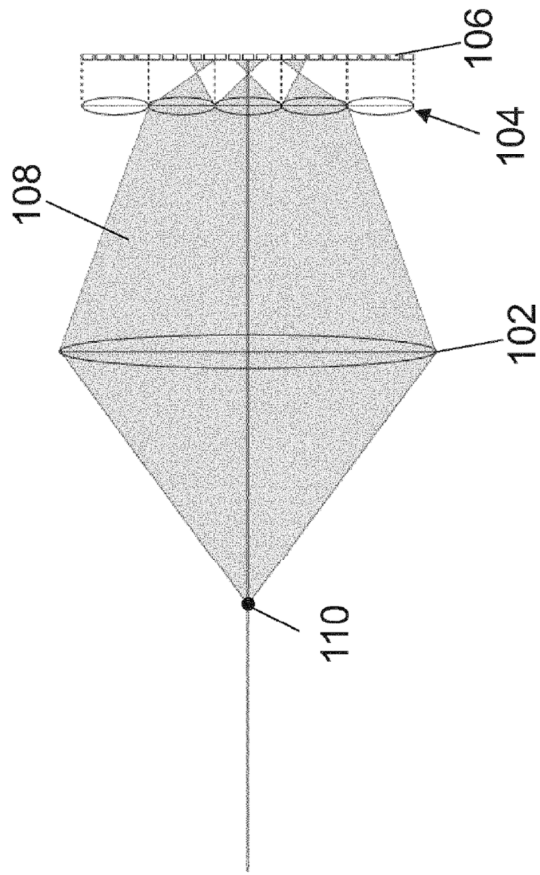


Fig. 2A

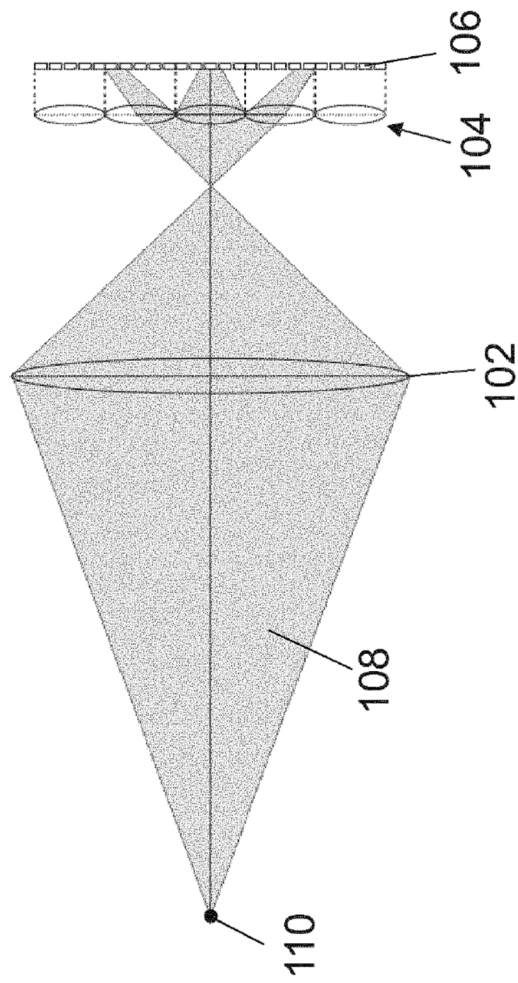


Fig. 3A

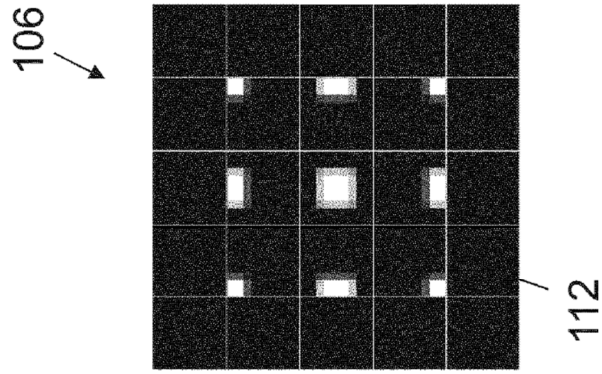


Fig. 3B

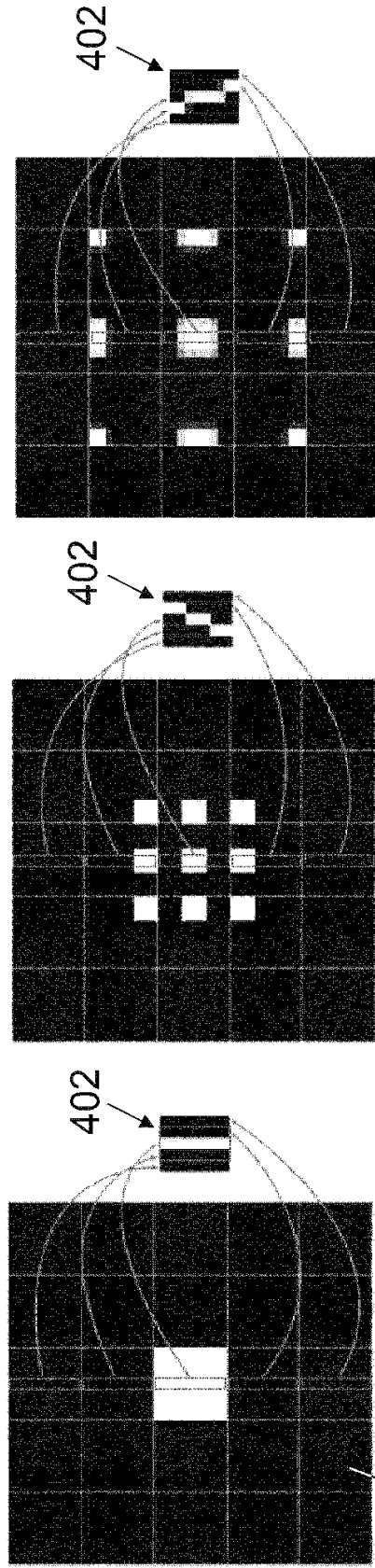
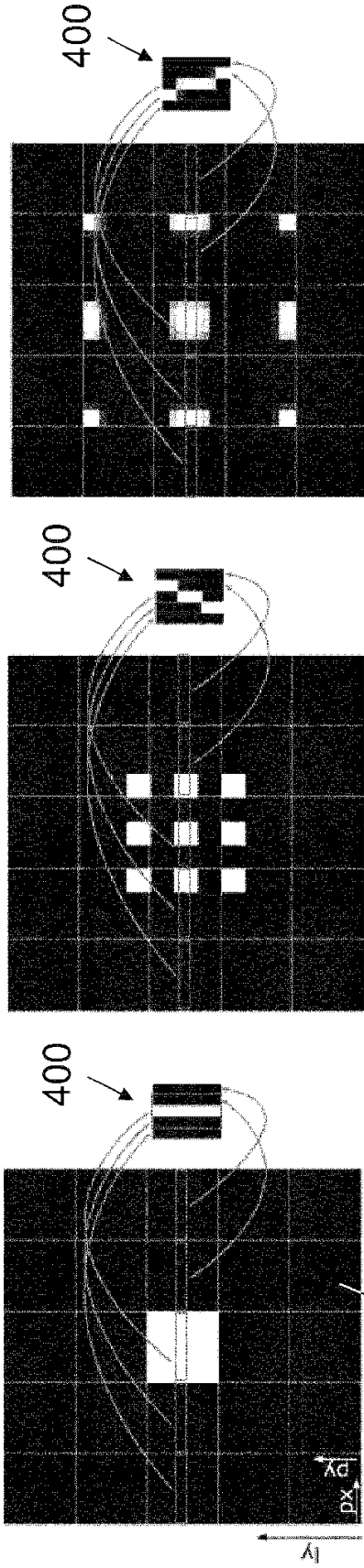


Fig. 4A

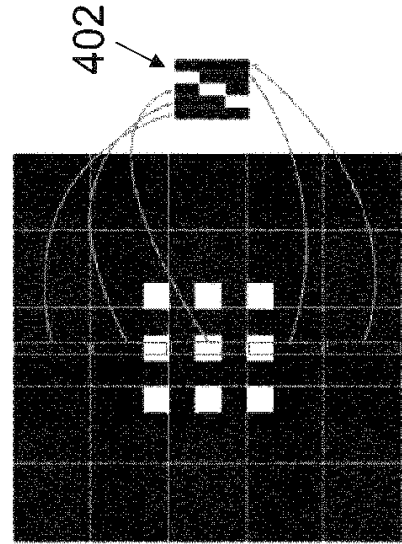
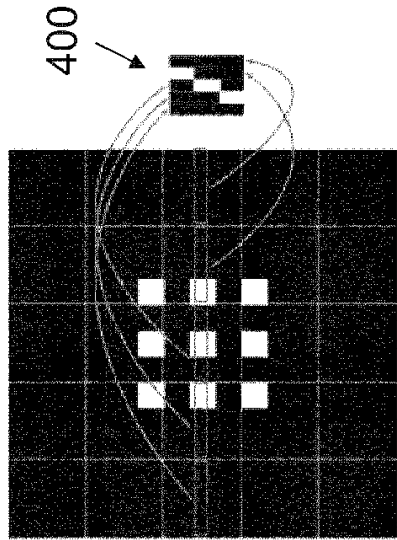


Fig. 4B

Fig. 4C

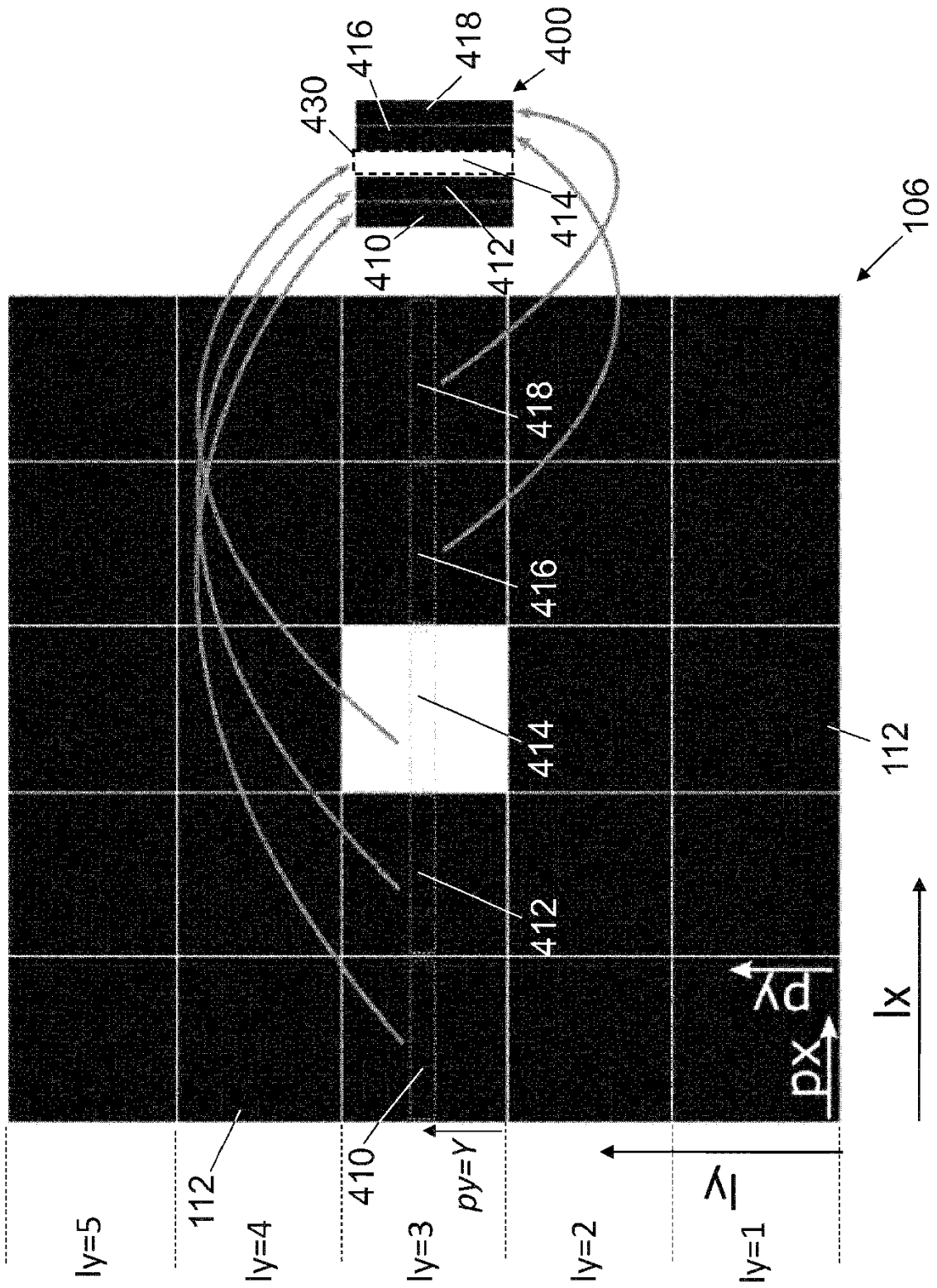


Fig. 4D

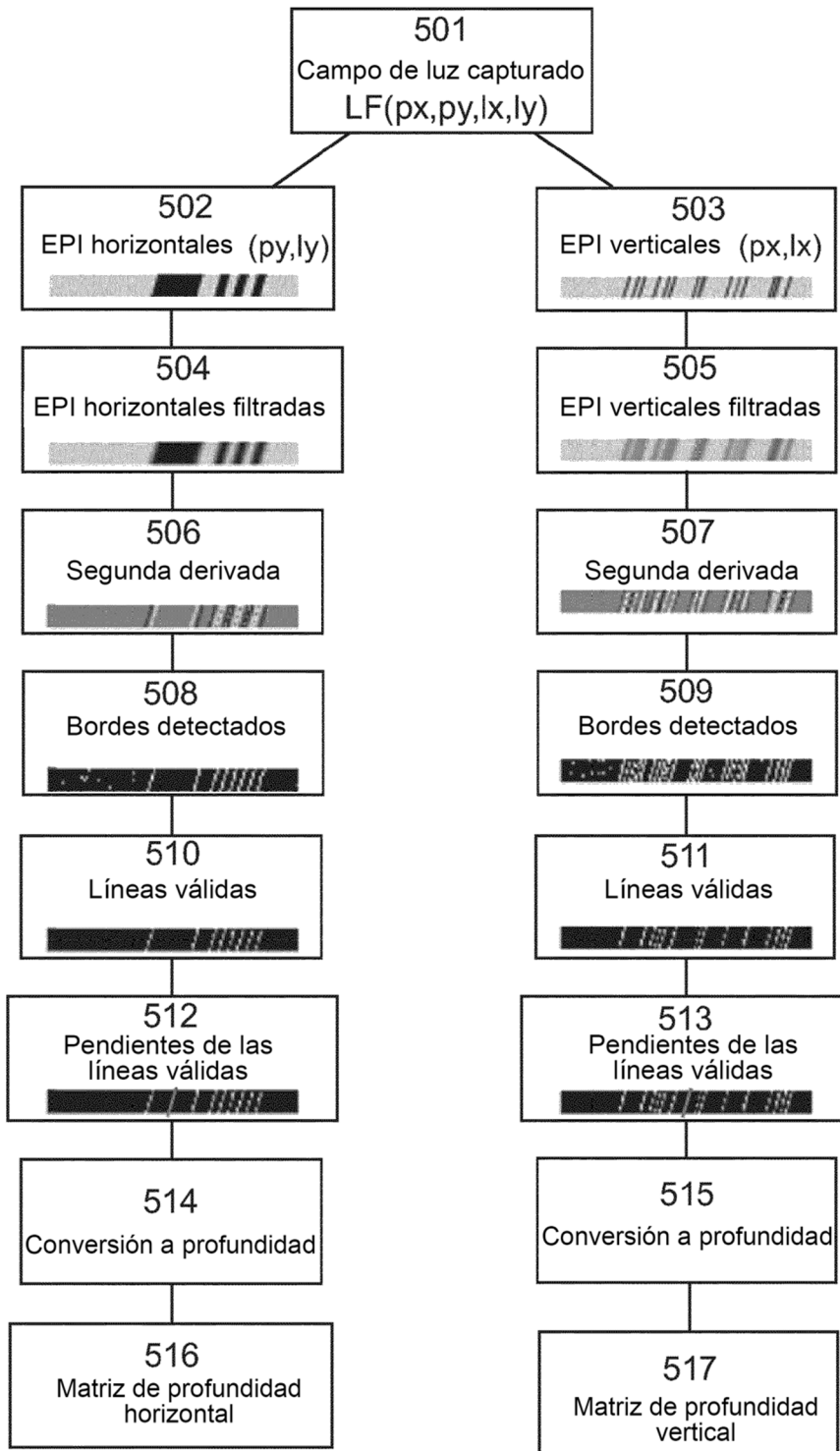


Fig. 5

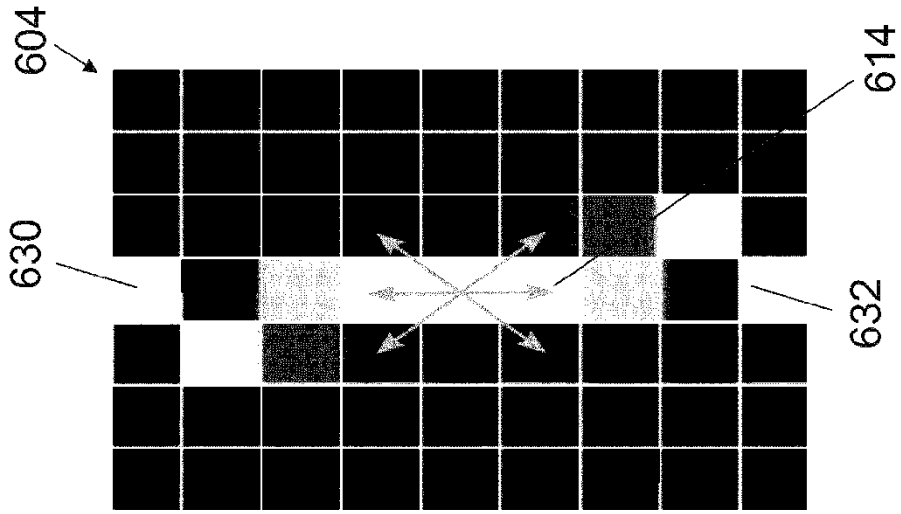


Fig. 6A

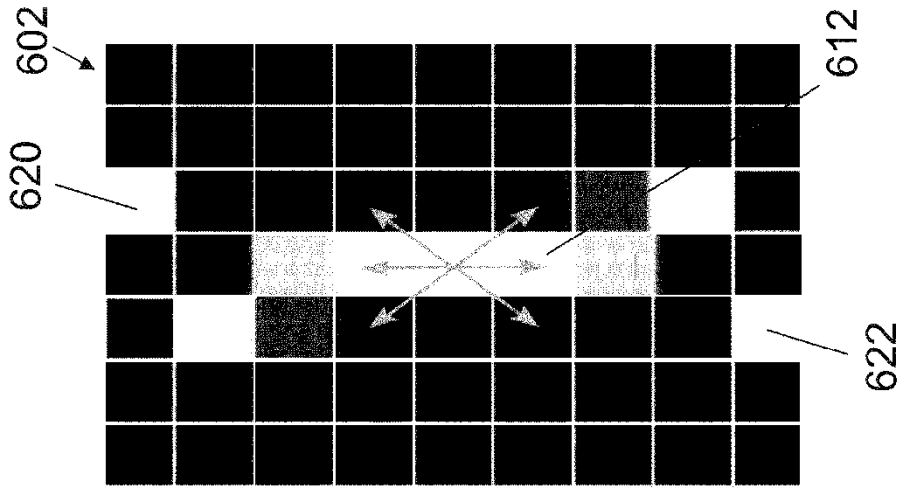


Fig. 6B

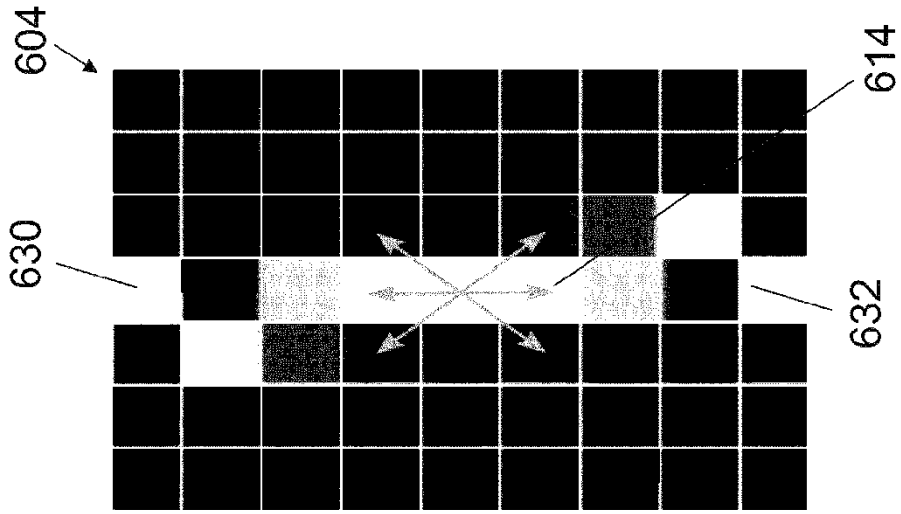


Fig. 6C

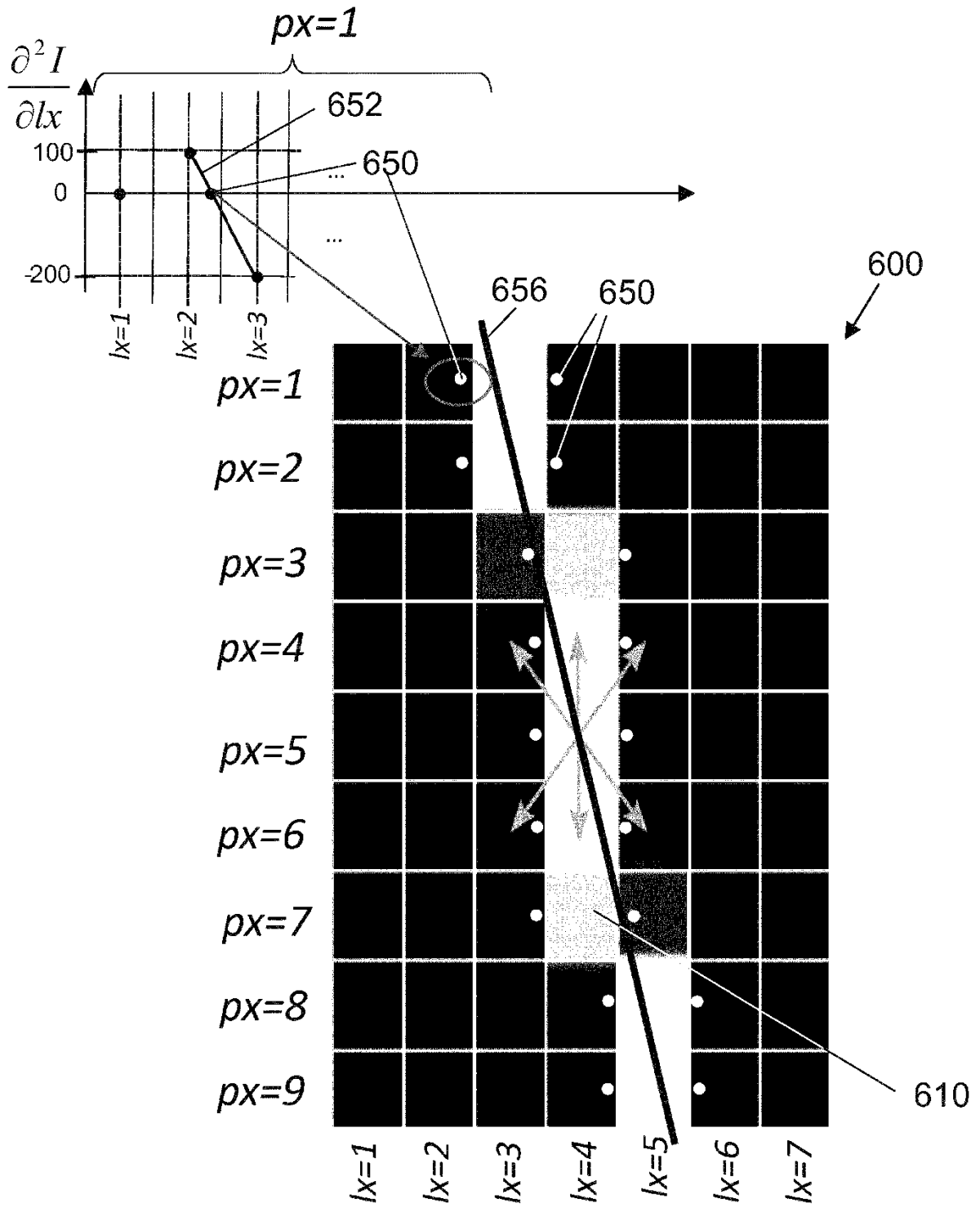
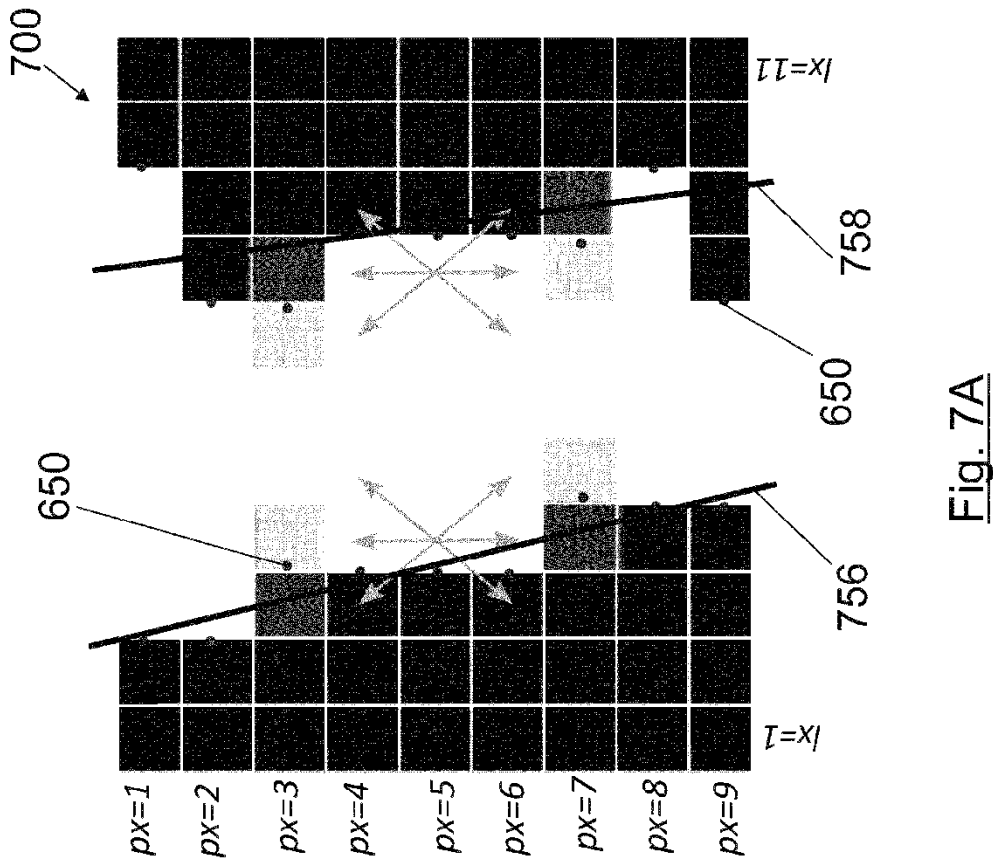
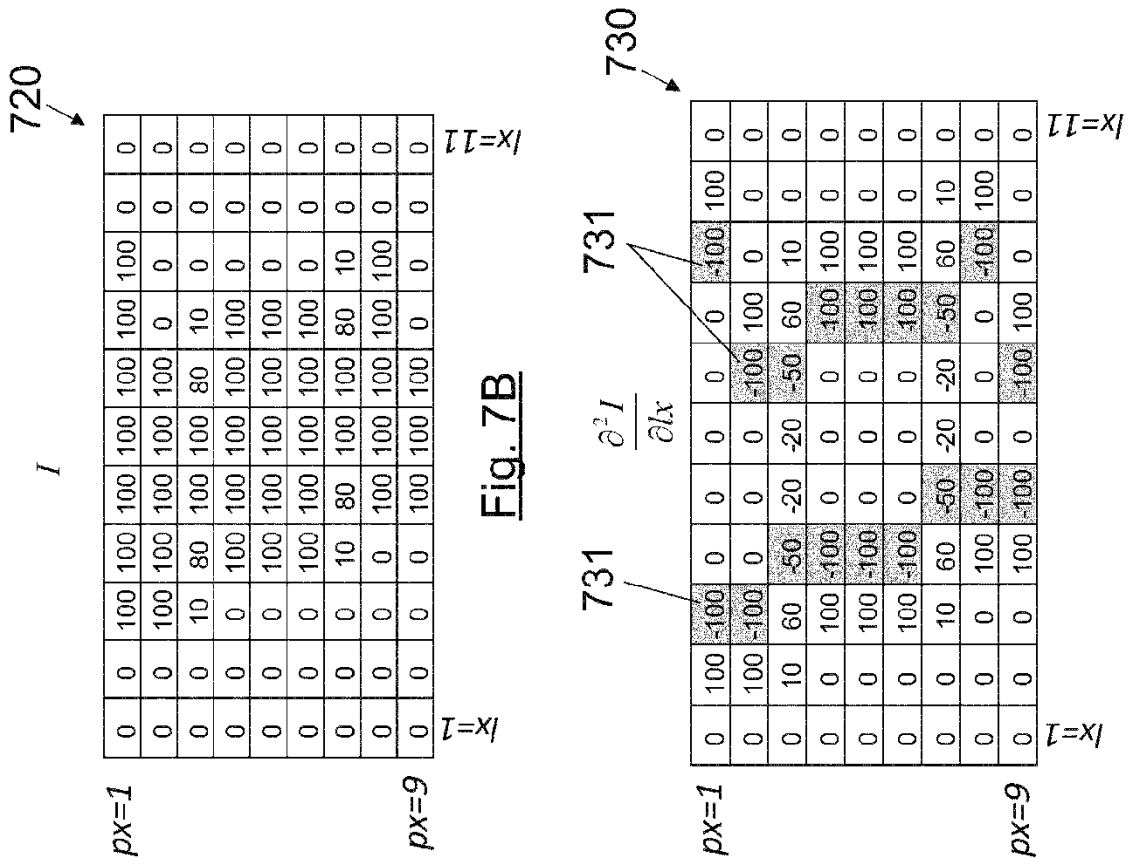


Fig. 6G



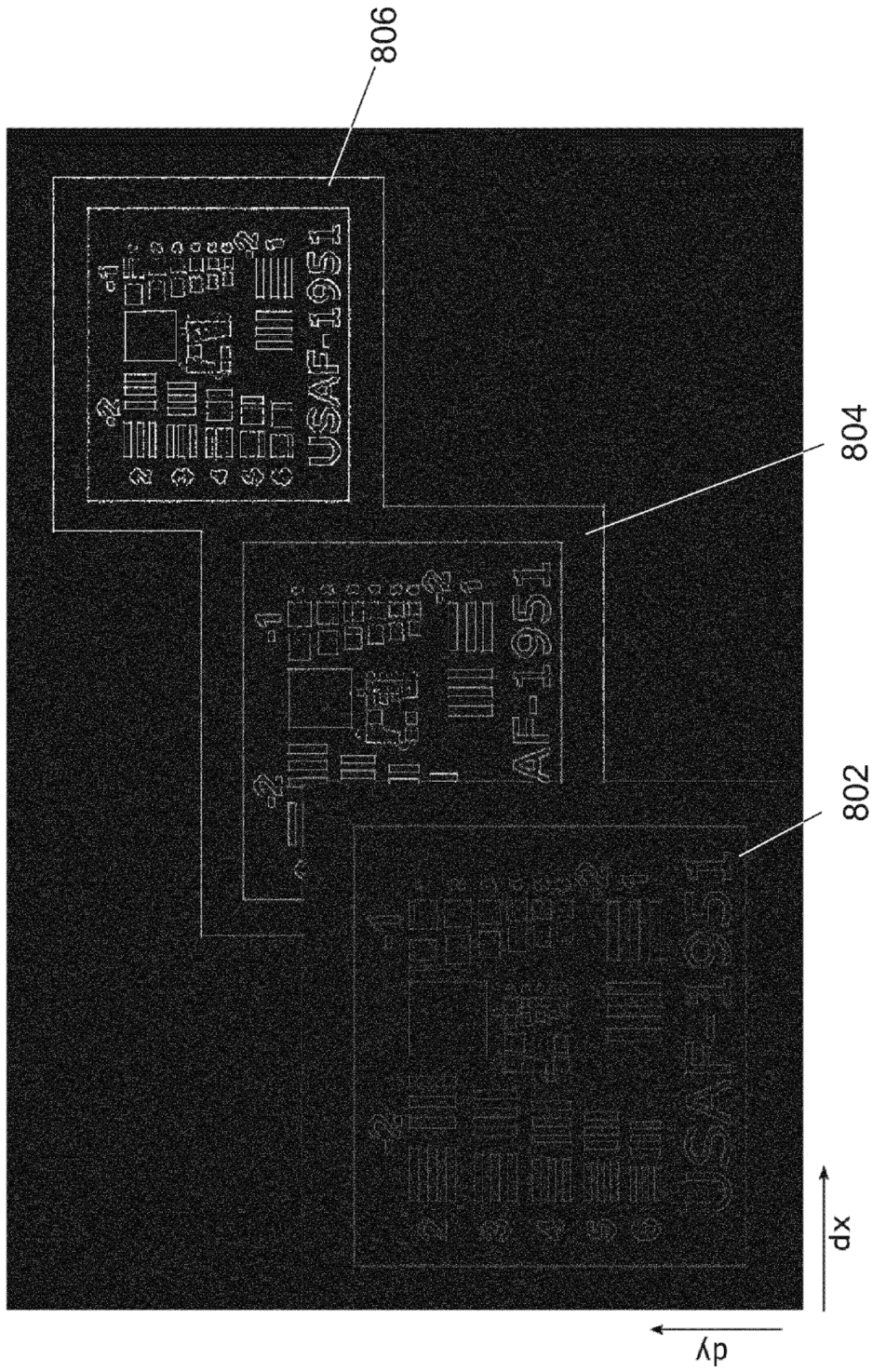


Fig. 8

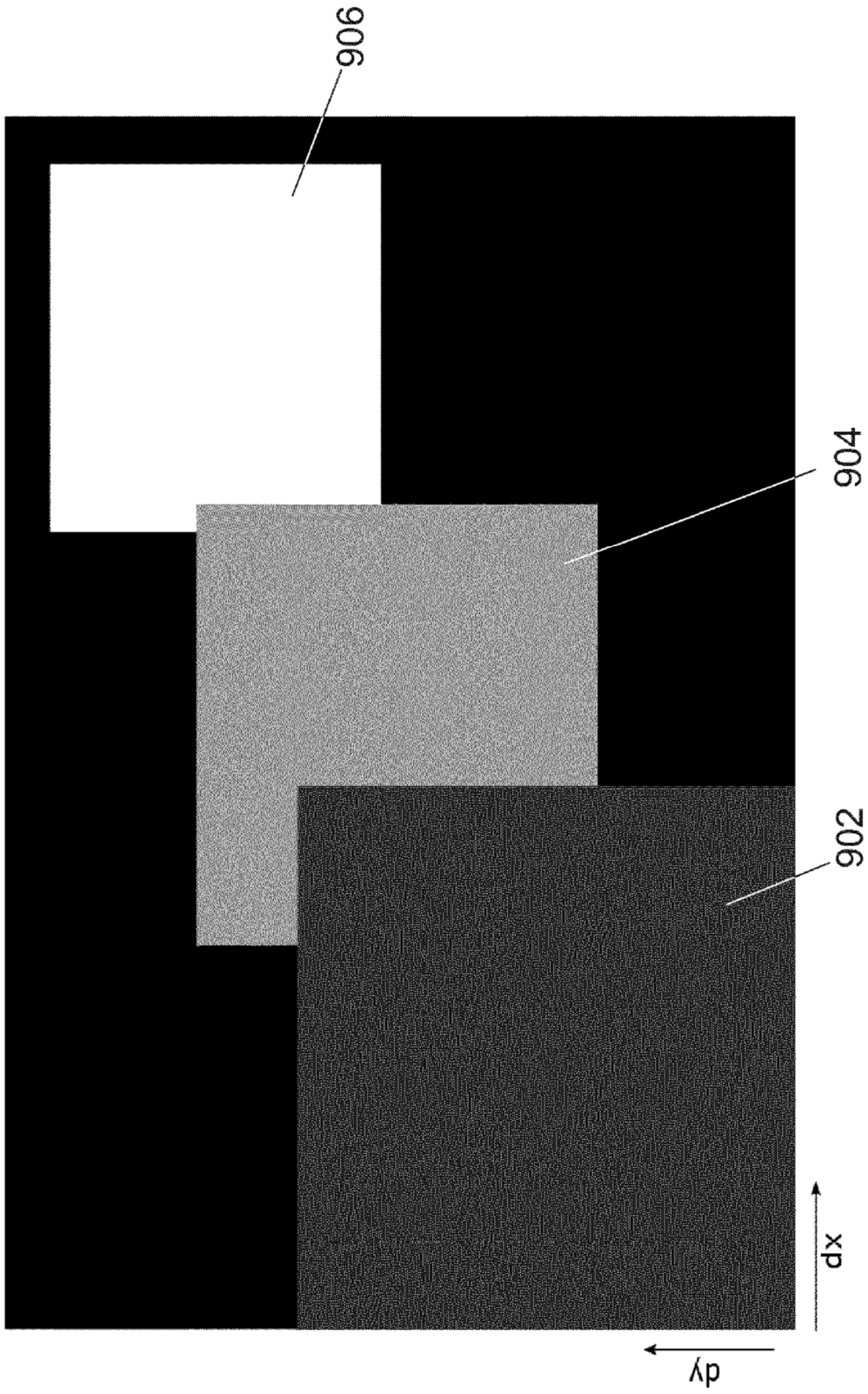


Fig. 9

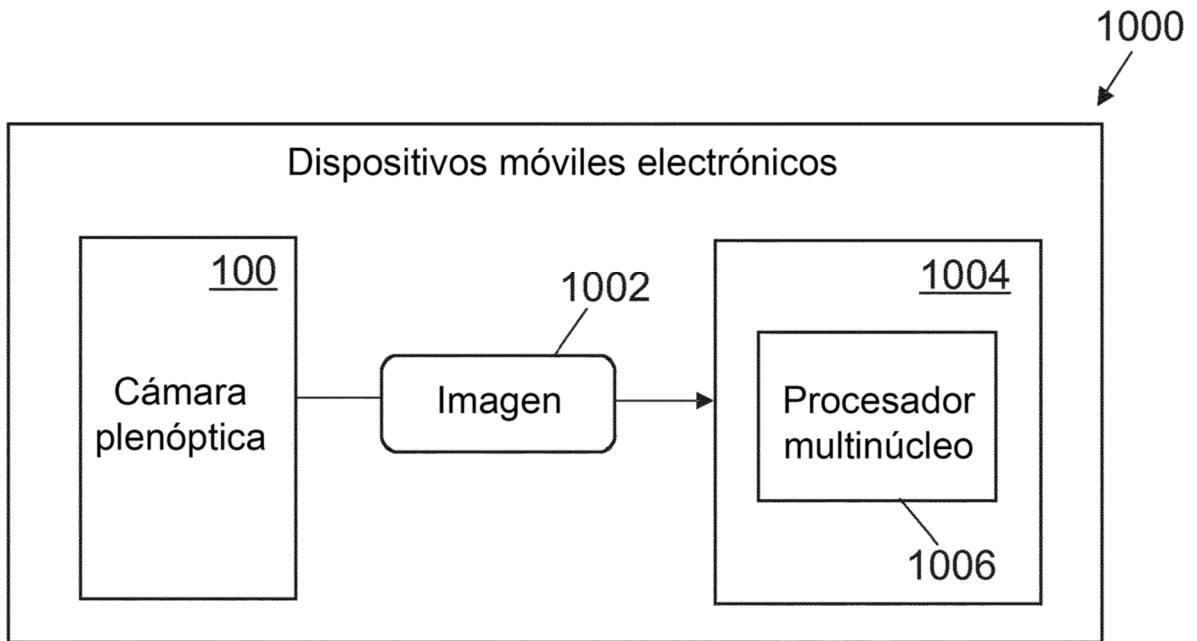


Fig. 10A

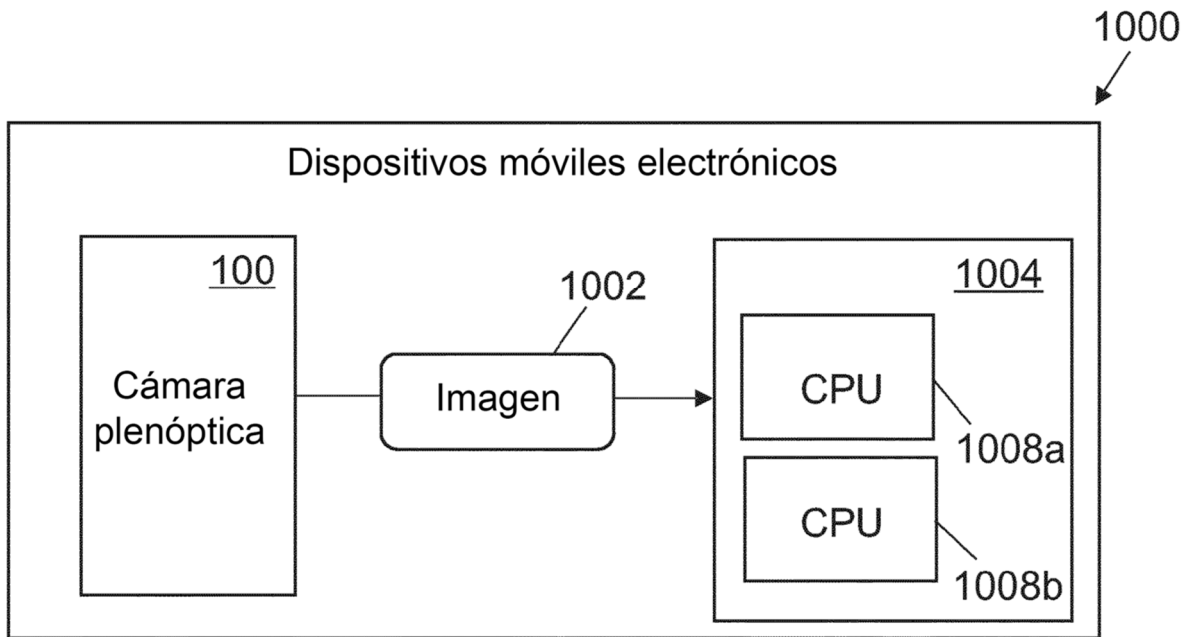


Fig. 10B

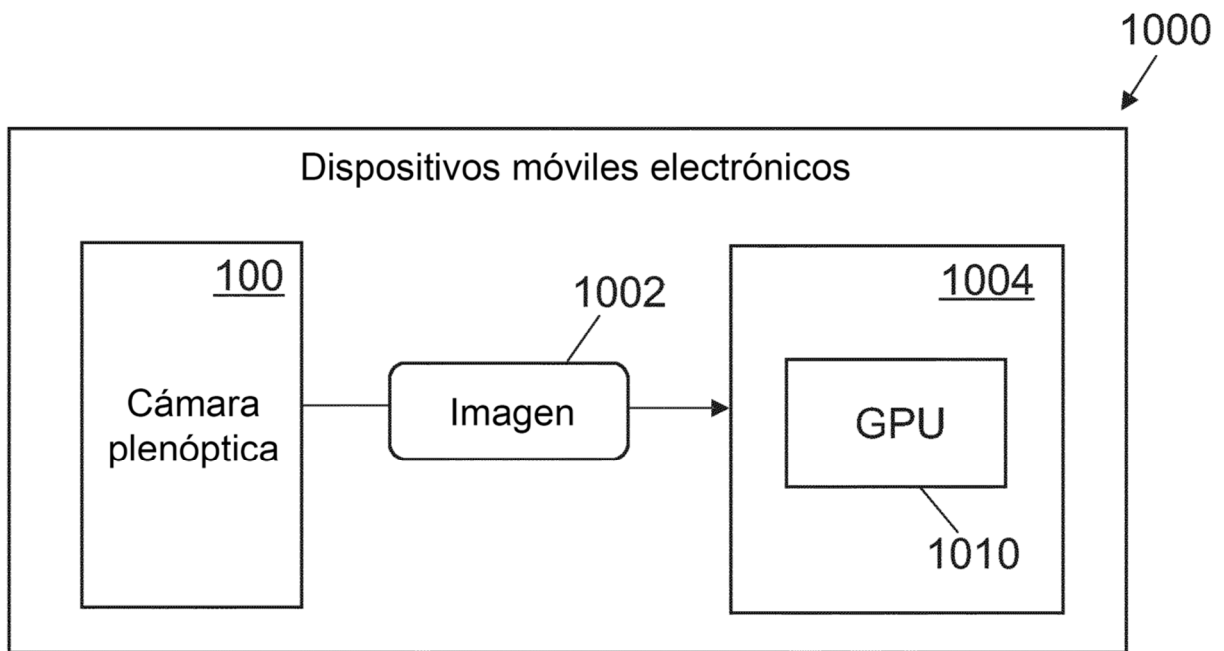


Fig. 10C

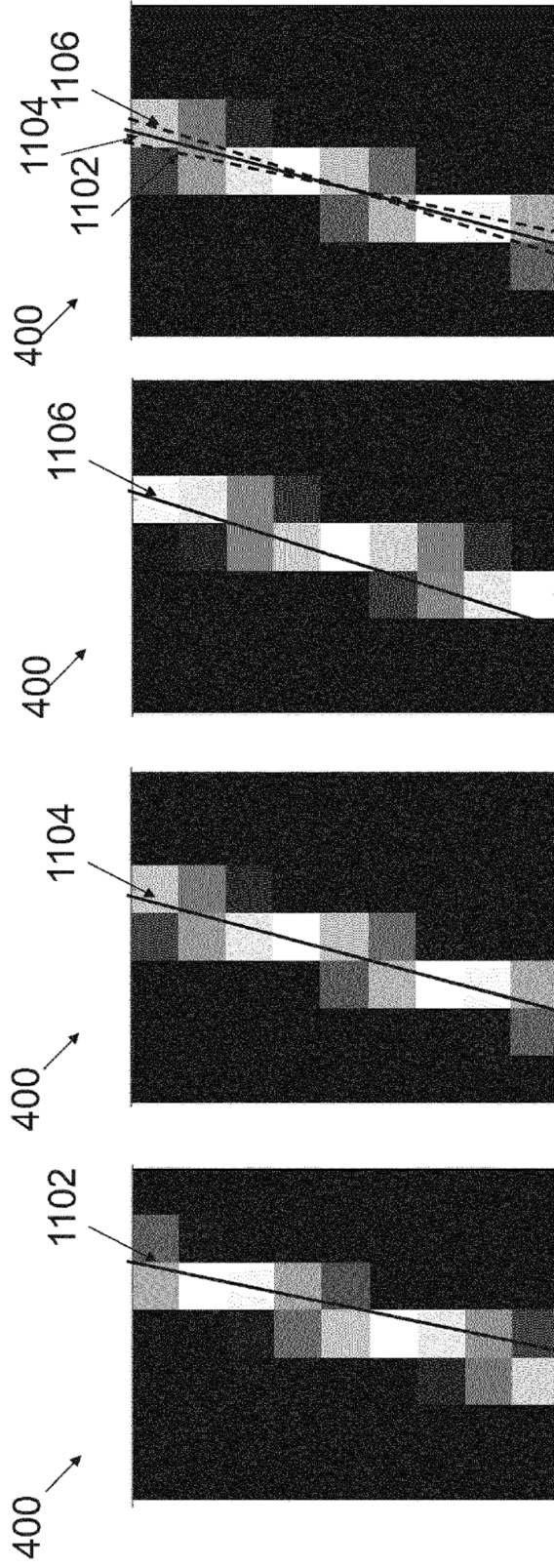


Fig. 11D

Fig. 11C

Fig. 11B

Fig. 11A

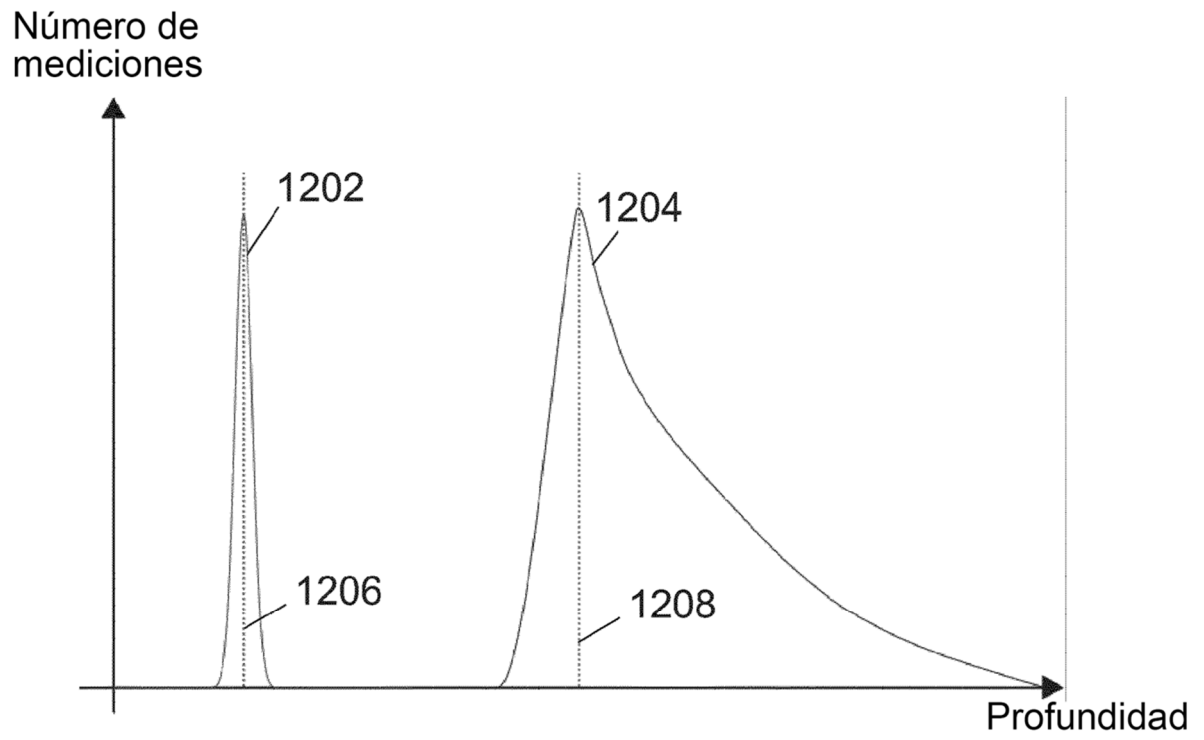


Fig. 12

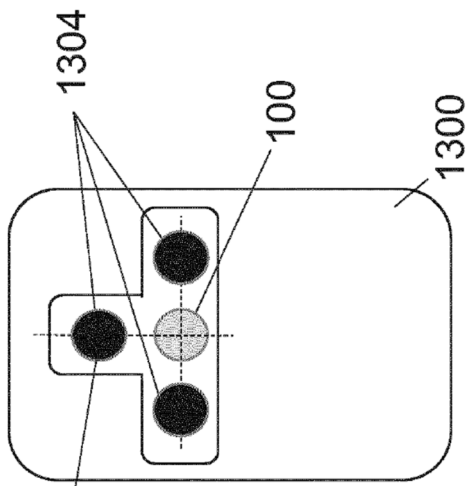


Fig. 13A

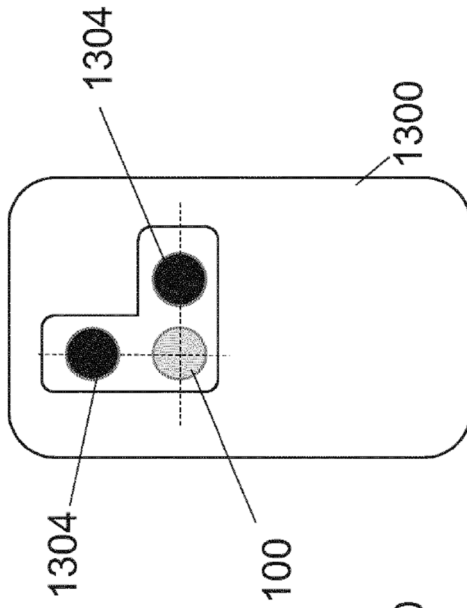


Fig. 13B

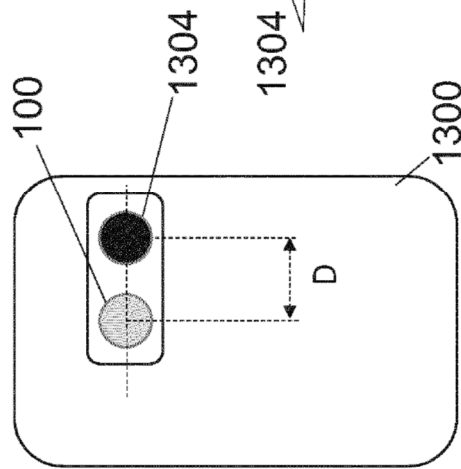


Fig. 13C

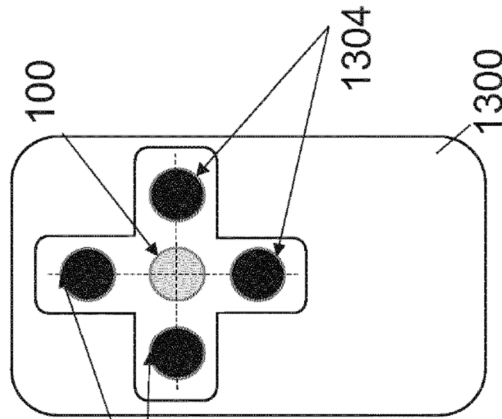


Fig. 13D

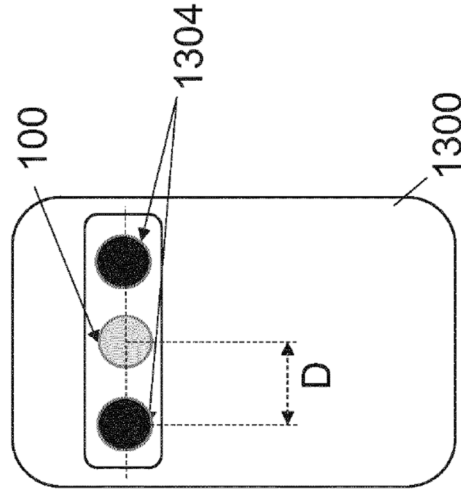


Fig. 13E

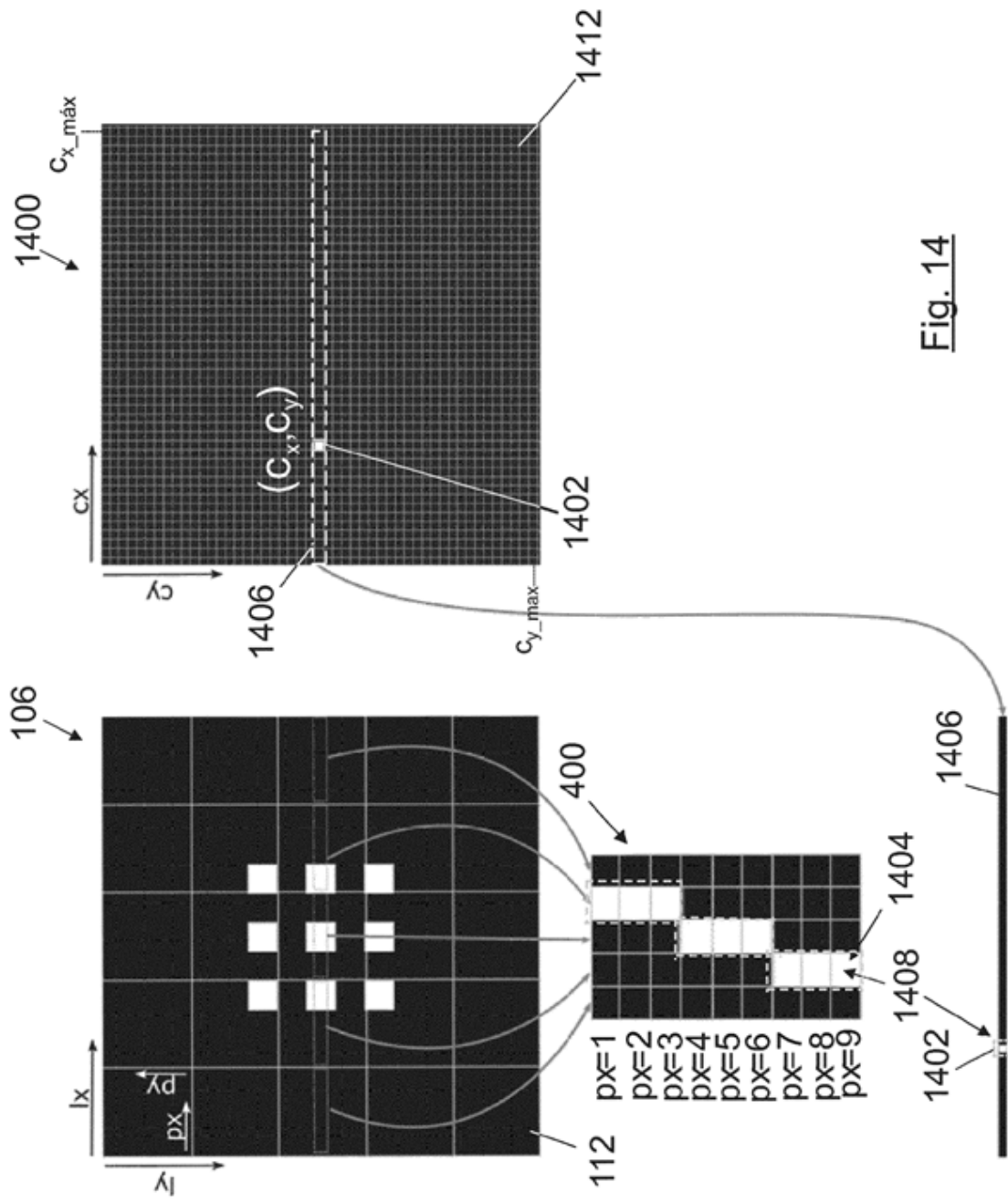


Fig. 14

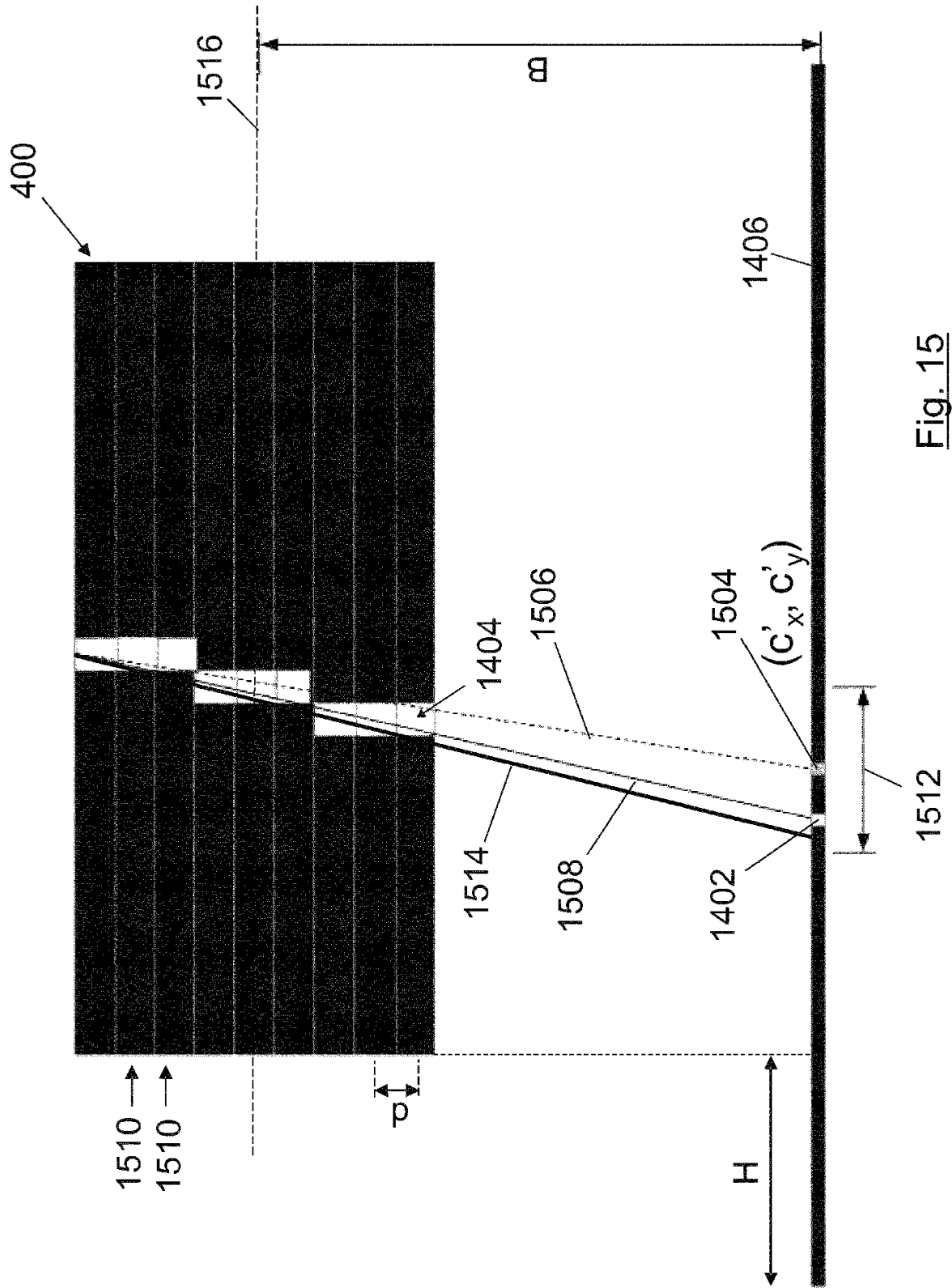


Fig. 15

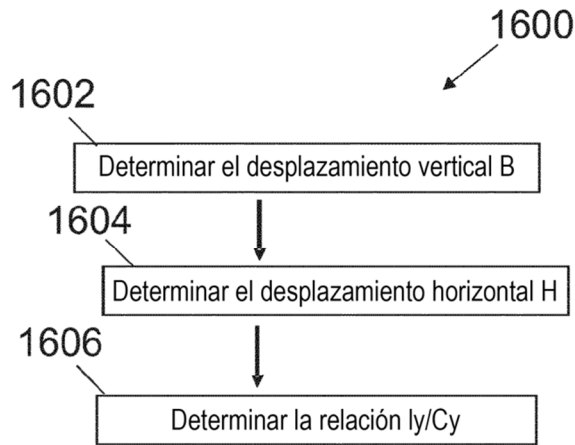


Fig. 16A

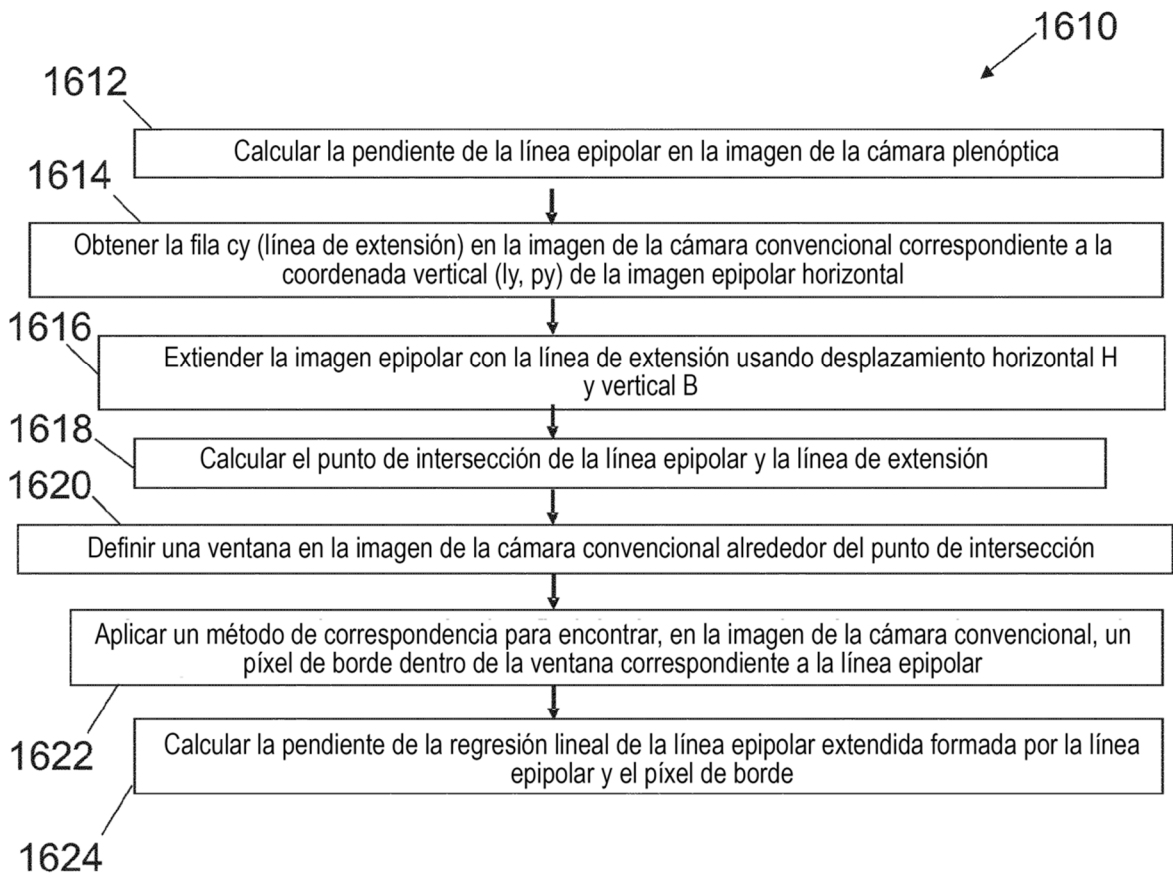


Fig. 16B

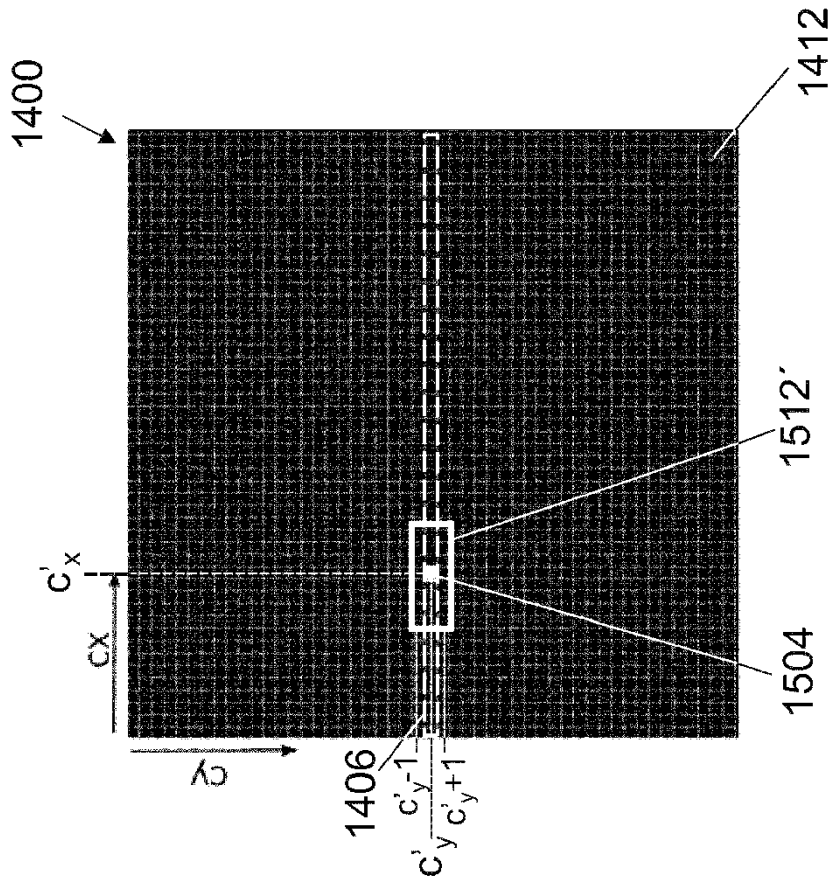


Fig. 17B

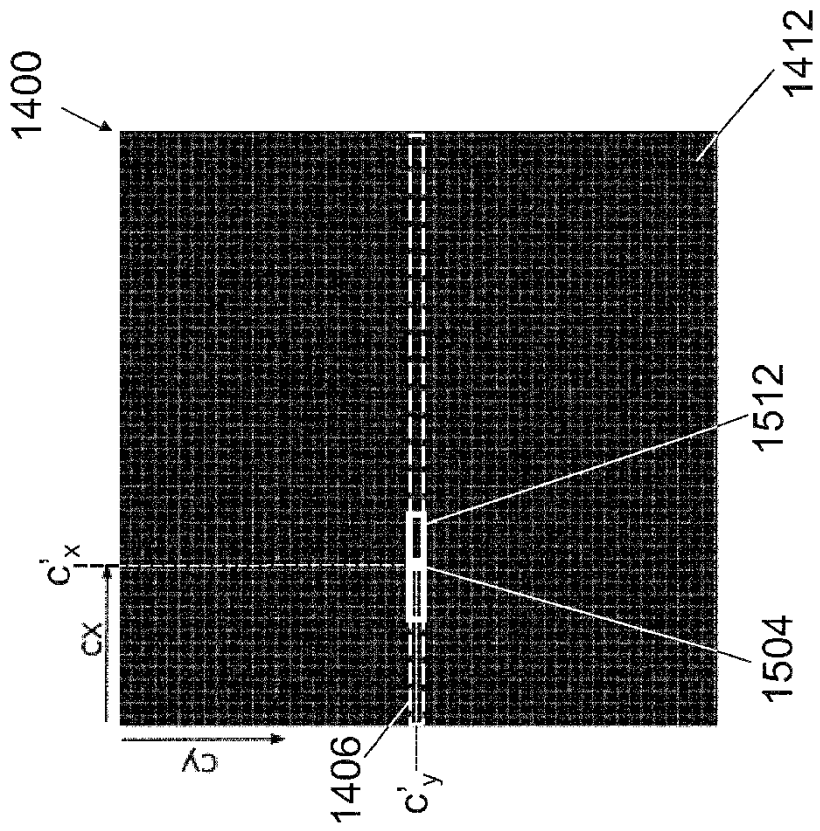


Fig. 17A

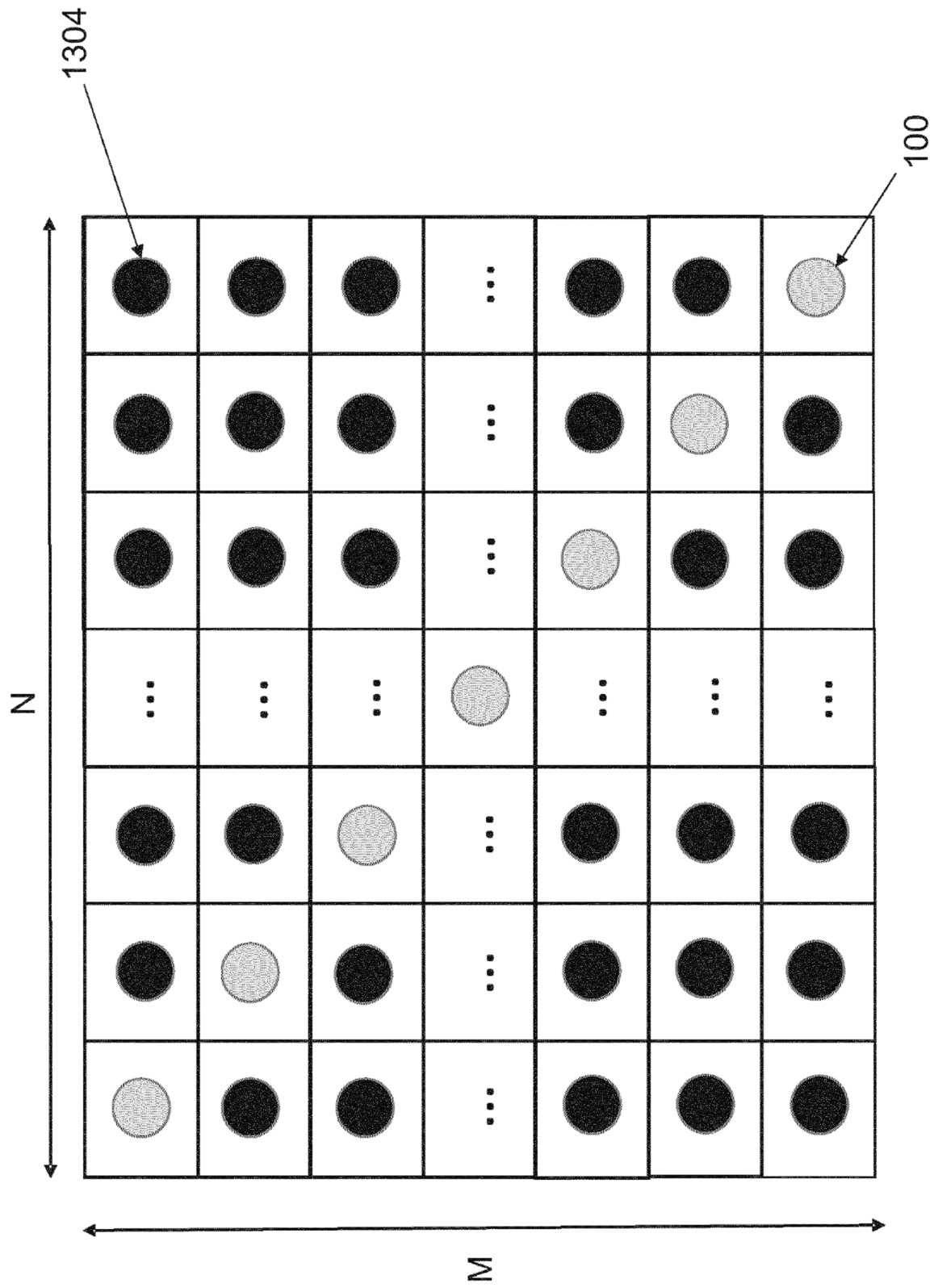


Fig. 18

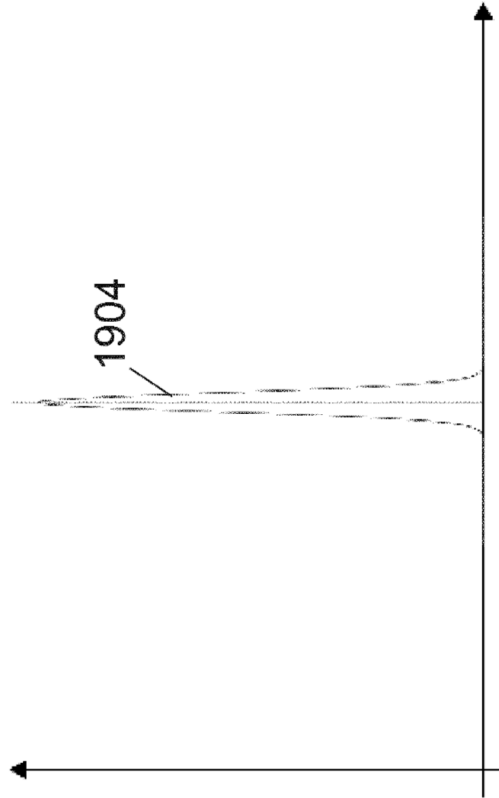


Fig. 19B

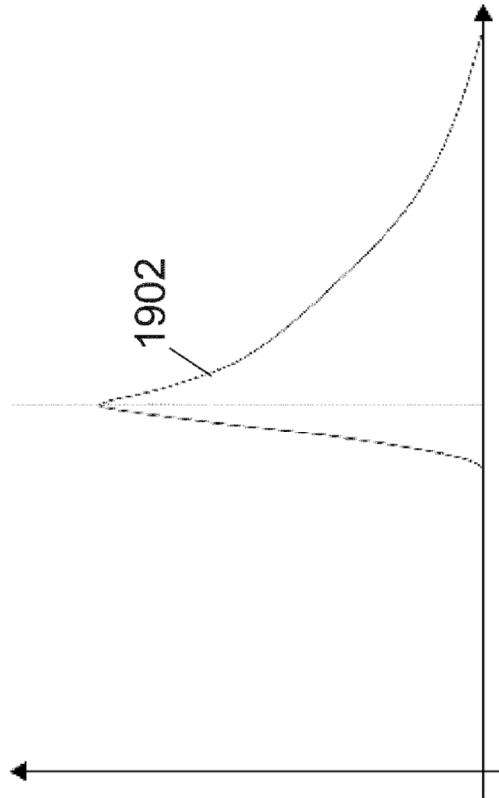


Fig. 19A

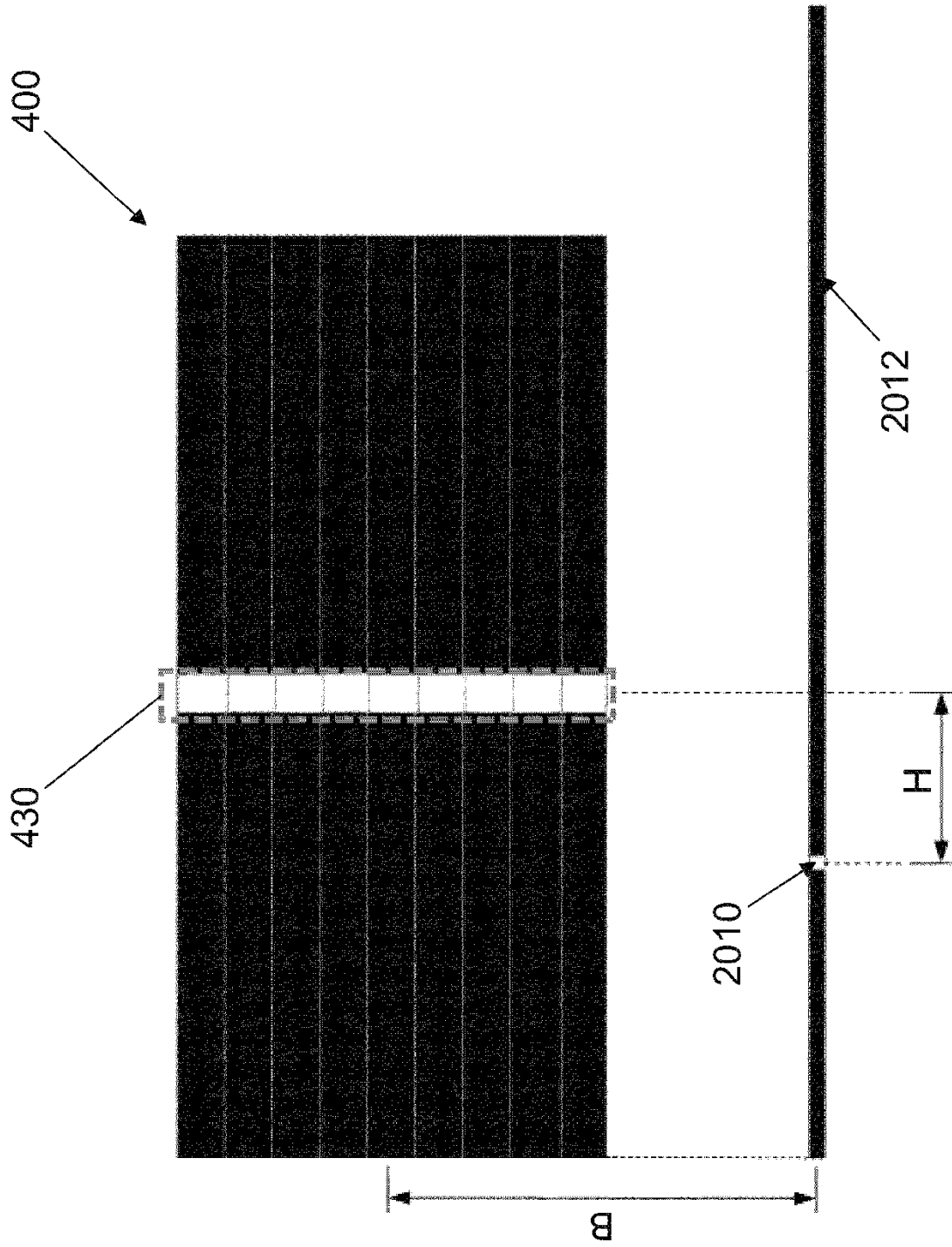


Fig. 20B

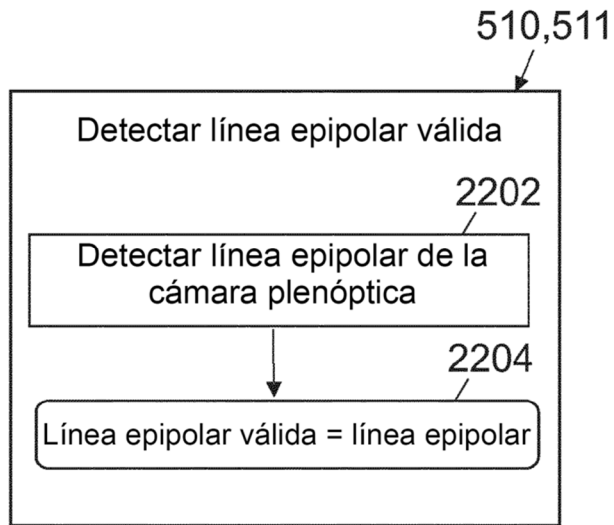


Fig. 22A

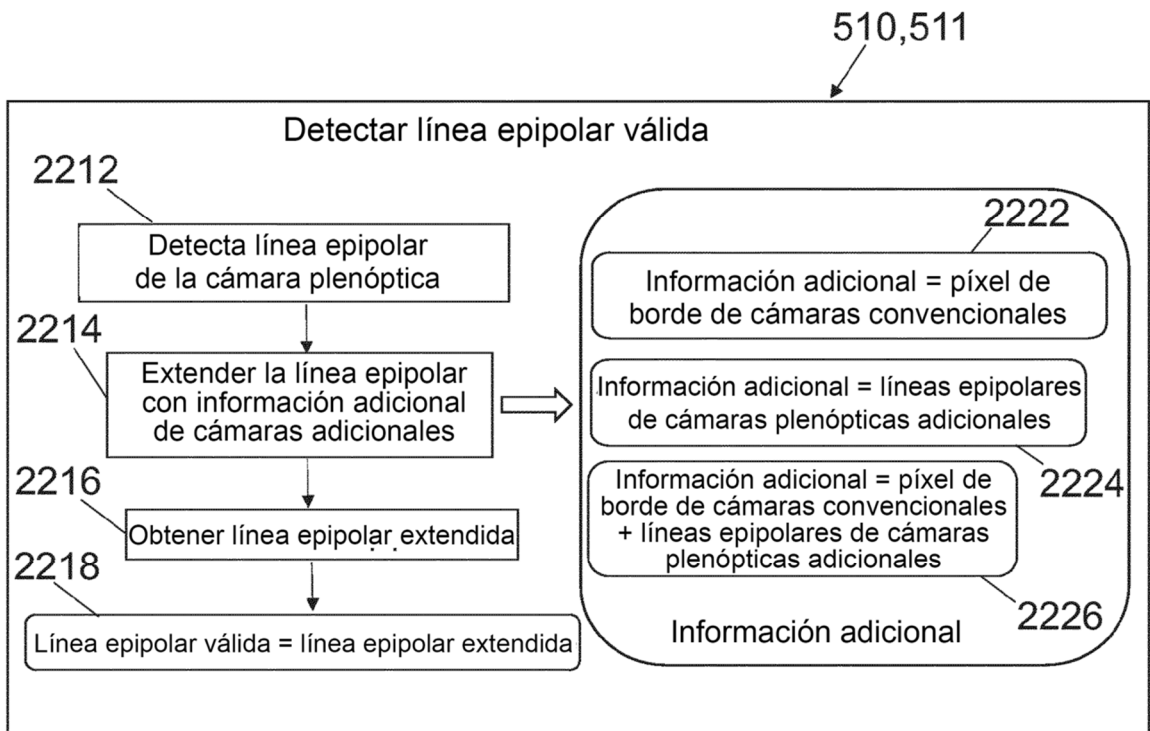


Fig. 22B

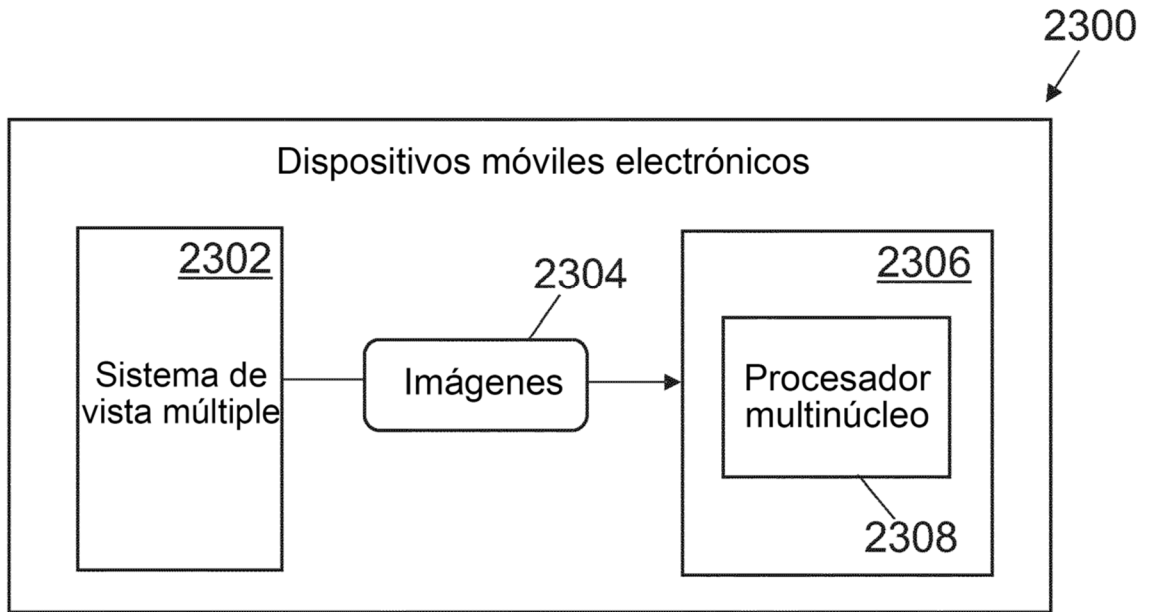


Fig. 23A

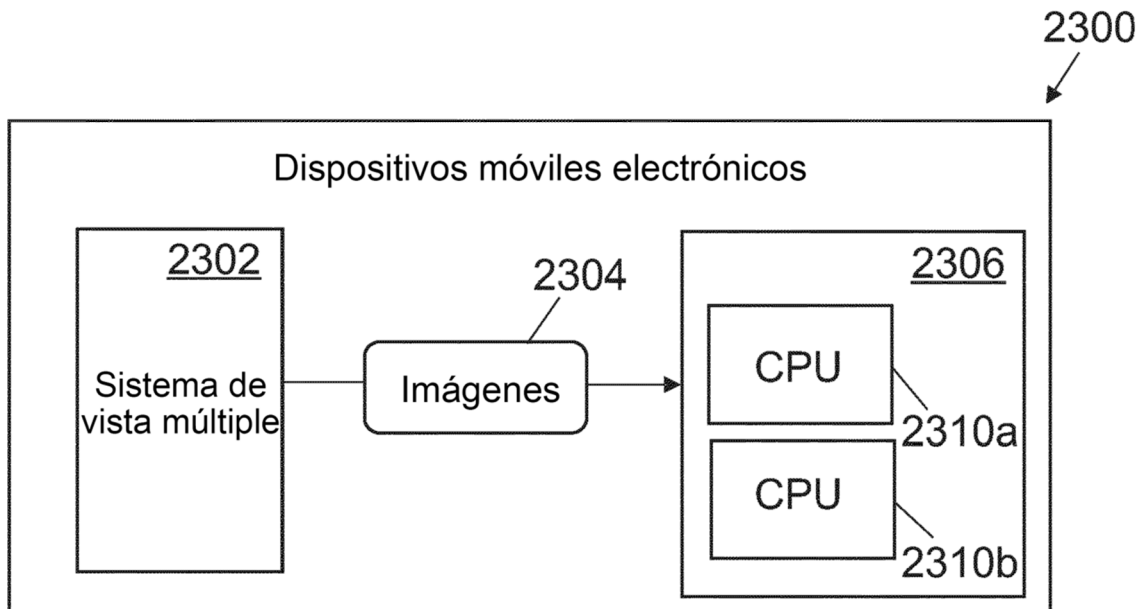


Fig. 23B

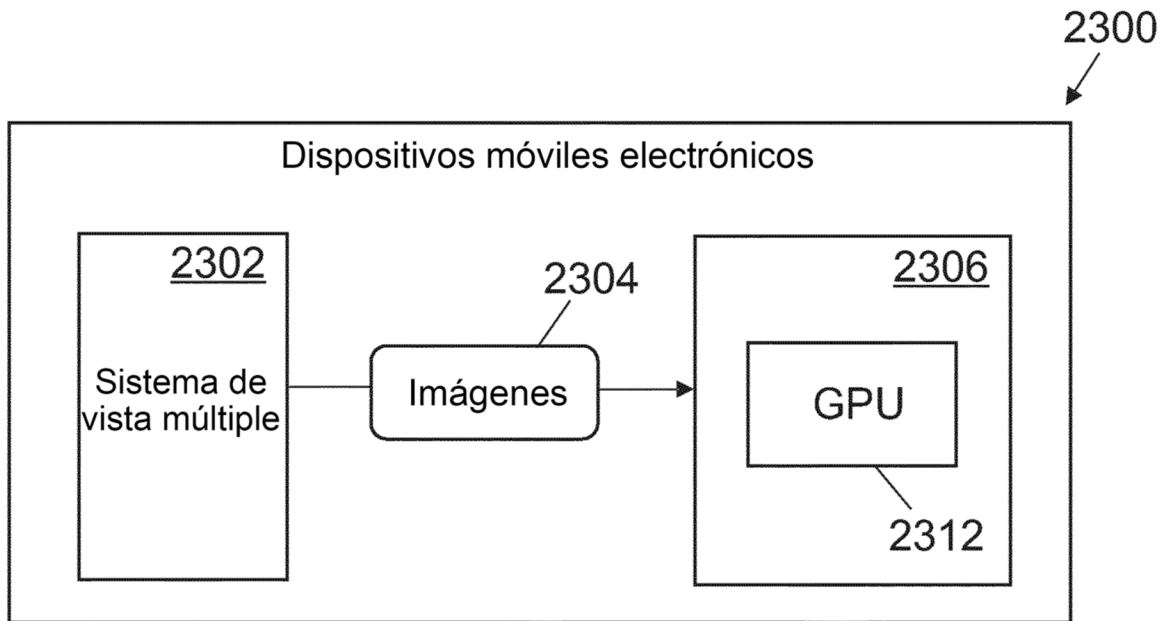


Fig. 23C