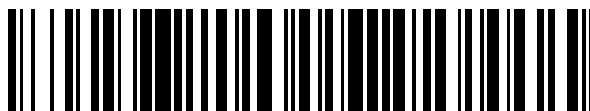


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 562 255**

51 Int. Cl.:

H04N 13/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.06.2010** **E 10729007 (4)**

97 Fecha y número de publicación de la concesión europea: **02.12.2015** **EP 2446635**

54 Título: **Inserción de objetos 3D en una imagen estereoscópica a una profundidad relativa**

30 Prioridad:

24.06.2009 US 220007 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
03.03.2016

73 Titular/es:

**DOLBY LABORATORIES LICENSING
CORPORATION (100.0%)
100 Potrero Avenue
San Francisco, CA 94103-4813, US**

72 Inventor/es:

**WELSH, RICHARD J. y
RALPH, CHRISTIAN B.**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 562 255 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

Inserción de objetos 3D en una imagen estereoscópica a una profundidad relativa

5 Tecnología

La invención se refiere al sector de las imágenes tridimensionales (3D) y, en particular, al procesamiento de objetos para la colocación a una profundidad perceptual.

10 Antecedentes

Para dispositivos bidimensionales (2D), los subtítulos (por ejemplo, subtitulación, subtitulación para sordos y otros) pueden ser renderizados arbitrariamente en una imagen 2D preexistente mediante un dispositivo de reproducción de video, o mediante un software de composición. Los subtítulos 2D insertados aparecerán, mediante oclusión, delante o detrás en la imagen preexistente. Los datos auxiliares proporcionados con los subtítulos pueden especificar ubicaciones y estilos de colocación en 2D simples. Por ejemplo, un dato auxiliar de los subtítulos de una emisión puede especificar subtitulación deslizable hacia arriba (en roll-up), emergente (en pop-up) o pintada (paint-on). Los subtítulos deslizables hacia arriba aparecen típicamente en la parte inferior de la pantalla, mientras que los subtítulos emergentes aparecen en cualquier lugar de la pantalla. Los subtítulos pintados utilizan un bloque estacionario predeterminado en la pantalla.

No obstante, estas técnicas convencionales no resultan adecuadas para una imagen 3D preexistente. La renderización arbitraria de los subtítulos en el espacio 3D resulta en los subtítulos existentes en el fondo (por ejemplo, detrás de un elemento preexistente, pero representados en primer plano (por ejemplo, delante del elemento preexistente), o existente opuestamente en primer plano y renderizados en el fondo. En otras palabras, los subtítulos pueden parecer, mediante oclusión, delante de un elemento mientras aparece, mediante disparidad interocular, detrás del mismo elemento. Este conflicto confunde el subtítulo visual humano, que rompe la ilusión de profundidad en la imagen 3D, y provoca asimismo fatiga ocular para un espectador.

Estos problemas pueden evitarse mediante un análisis visual manual de cada inserción de subtítulos 3D durante la producción. Es decir, un editor humano confirma si cada inserción es apropiada. Como se puede imaginar, este proceso no resulta práctico para contenido pre-grabado y de todo punto imposible para una emisión en directo.

Además, se realiza un análisis visual manual para la dimensión de un dispositivo de visualización específico, y a menudo, produce un resultado no anticipado para un dispositivo de visualización diferente. La profundidad perceptual es no lineal con la disparidad de píxel, y la disparidad de píxel es una función tanto del desfase de separación de píxel como de cada resolución de píxel del dispositivo de visualización. De acuerdo con esto, los subtítulos insertados pueden aparecer en una ubicación no deseada incluso utilizando el mismo desfase de separación de píxel. Por ejemplo, los subtítulos pueden aparecer en el dispositivo de visualización diferente más allá del infinito, produciendo divergencia en los ojos del espectador. Los subtítulos aparecen más allá del infinito cuando su disparidad de píxel entre las imágenes del ojo derecho y el izquierdo es mayor que la separación interocular del espectador, que es en general de aproximadamente 65 milímetros para un adulto. Un resultado no anticipado puede incluir asimismo una colocación ilegal de los subtítulos, tal como detrás del espectador, lo que es imposible de conseguir en un visualizador.

De lo anterior, se ve que los subtítulos 3D, así como la inserción de objetos 3D en general, situados automáticamente y/o independientemente de un dispositivo de visualización puede proporcionar muchos beneficios sobre las técnicas 2D y 3D convencionales.

Sumario de la descripción

Existen métodos y aparatos para renderizar al menos un objeto en una imagen estereoscópica para un dispositivo de visualización. La imagen objeto puede incluir dos imágenes (por ejemplo, una imagen para el ojo derecho y una imagen para el ojo izquierdo) manipuladas separadamente o compuestas entre sí. En una realización, un método incluye recibir datos de profundidad perceptual como fracción de la distancia del espectador para el objeto. Un desfase de separación de píxel para un dispositivo de visualización particular se calcula utilizando la profundidad perceptual. Las imágenes primera y segunda (tales como las imágenes de ojo izquierdo y derecho) del objeto se insertan respectivamente en la imagen estereoscópica con el desfase de separación de píxel.

En una modificación de la invención, se proporcionan datos de profundidad como fracción de la distancia del espectador, para al menos una porción de un subtítulo. Los datos de profundidad se utilizan para determinar un desfase de separación de píxel apropiado entre las imágenes primera y segunda del subtítulo para obtener substancialmente una profundidad perceptual deseada. Las imágenes primera y segunda del subtítulo son insertadas en una imagen estereoscópica o en un flujo de bits de imagen con el desfase de separación de píxel determinado. En una realización específica, los datos de profundidad proporcionados están normalizados para un dispositivo de visualización de representación (por ejemplo, una resolución de un píxel por milímetro, un ancho de 1 m).

En otra realización más, existe un método para la inserción de objetos en una imagen estereoscópica. La imagen estereoscópica es transmitida directa o indirectamente, a un primer dispositivo de reproducción (tal como un teléfono móvil, un ordenador, una televisión un proyector de cine u otros similares). Los datos de profundidad perceptual como fracción de la distancia del espectador para al menos un objeto son asimismo transmitidos al primer dispositivo de reproducción. El primer dispositivo de reproducción es capaz de calcular un primer desfase de separación de píxel a partir de los datos de profundidad perceptual. La imagen estereoscópica, así como los mismos datos de profundidad perceptual, se transmiten directa o indirectamente a un segundo dispositivo de reproducción. El segundo dispositivo de reproducción es capaz de calcular un segundo desfase de separación de píxel a partir de los datos de profundidad perceptual. En el caso de que los dispositivos de reproducción primero y segundo tengan dimensiones de ancho de pantalla o resoluciones de píxel diferentes, el primer desfase de separación de píxel es distinto del segundo desfase de separación de píxel.

En otra realización de la invención, un método proporciona generar datos de profundidad normalizados para la representación de un objeto en una profundidad estereoscópica 3D. El método incluye la identificación de un elemento dentro de una imagen estereoscópica mediante, por ejemplo, segmentación. Se determina un desfase espacial (tal como un desfase de separación de píxel) en la imagen estereoscópica para el elemento. Pueden utilizarse opcionalmente una pluralidad de desfases espaciales para formar un mapa de disparidad de la imagen estereoscópica. El desfase espacial se correlaciona con una profundidad percibida y, de este modo, pueden determinarse la posición y el volumen ocupados del elemento en el espacio 3D. El método incluye asimismo determinar la geometría de una imagen estereoscópica del objeto para ser insertada en la imagen estereoscópica. Uno o más desfases espaciales de objeto se calculan para colocar el objeto estereoscópico en el espacio 3D evitando el volumen ocupado, así como cumpliendo cualquier otra restricción de colocación. Cada uno de los uno o más desfases espaciales del objeto pueden ser normalizados para ser independientes del dispositivo de visualización y expresados como su correspondiente profundidad perceptual como fracción de la distancia del espectador. Puede obtenerse un flujo de datos de cada profundidad perceptual. En una realización específica, se utiliza también un mapa de profundidad de la imagen estereoscópica para ajustar de manera más precisa la profundidad o profundidades percibidas y, de acuerdo con esto, la posición y el volumen ocupados del elemento.

En otra modificación de la invención, un aparato para generar una señal de salida incluye un terminal de entrada, un terminal de salida y circuitos de procesamiento de señal acoplados al terminal de entrada y al terminal de salida. Los circuitos de procesamiento de señal están adaptados para recibir datos de profundidad perceptual como fracción de la distancia del espectador para al menos un objeto y calcular un desfase de separación de píxel apropiado. Los circuitos de procesamiento de señal insertan, respectivamente, una imagen del objeto para el ojo izquierdo y una imagen del objeto para el ojo derecho de una imagen estereoscópica. La imagen del objeto para el ojo izquierdo y la imagen del objeto para el ojo derecho están desfasadas en el desfase de separación de píxel.

En otra realización de la invención, un aparato para generar datos de profundidad para la renderización de un objeto incluye un terminal de entrada, un terminal de salida y circuitos de procesamiento de señal acoplados al terminal de entrada y al terminal de salida. Los circuitos de procesamiento de señal están adaptados para identificar elementos dentro de una imagen estereoscópica y determinar desfases espaciales para los elementos. Los desfases espaciales pueden ser utilizados opcionalmente por los circuitos de procesamiento de señal para formar un mapa de disparidad. Los circuitos de procesamiento de señal están adaptados para determinar además la geometría de una imagen de objeto estereoscópica para ser insertada en la imagen estereoscópica y calcular uno o más desfases espaciales del objeto para colocar el objeto estereoscópico en el espacio 3D evitando volúmenes ocupados, así como satisfacer cualquier otra restricción de colocación. Cada uno de los uno o más desfases espaciales del objeto puede ser normalizado para ser independiente del dispositivo de visualización y expresados como su correspondiente profundidad perceptual como fracción de la distancia del espectador. Puede extraerse un flujo de datos de cada profundidad perceptual a través del terminal de salida.

Como otra realización de la invención, existe un medio de almacenamiento que graba un programa de instrucciones. El programa es ejecutable mediante un dispositivo de visualización para llevar a cabo un método para generar una salida. El método incluye recibir datos de profundidad perceptual normalizada como fracción de la distancia del espectador para al menos un objeto. Un desfase de separación de píxel es calculado para el dispositivo de visualización a partir de los datos de profundidad perceptual normalizada y de una dimensión de visualización. Imágenes de objeto primera y segunda del al menos un objeto se insertan en las imágenes primera y segunda de la imagen estereoscópica. Las imágenes de objeto primera y segunda están desfasadas por el desfase de separación de píxel. El dispositivo de visualización visualmente produce la imagen estereoscópica junto con el objeto insertado.

Varios objetos, elementos y ventajas adicionales de la invención pueden apreciarse de manera más completa con referencia a la descripción detallada y a los dibujos que siguen.

Breve descripción de los dibujos

Una apreciación más completa de la invención y muchas de las ventajas que se esperan de la misma se obtendrán fácilmente, dado que se comprenden mejor con referencia a la siguiente descripción detallada cuando se considera junto con los dibujos que se acompañan, en los cuales:

la Fig. 1 ilustra una disposición de colocación 3D de acuerdo con una realización de la invención;

la Fig. 2A ilustra, como ejemplo, una relación trigonométrica entre la disparidad de imagen y la separación interocular;

la Fig. 2B ilustra la profundidad perceptual en función de la disparidad de píxel para una pantalla de 7 metros de ancho de ejemplo de acuerdo con una realización de la invención;

la Fig. 3 ilustra una disposición de colocación 3D para una pantalla normalizada de acuerdo con una realización de la invención;

la Fig. 4 ilustra un diagrama de flujo simplificado para la colocación de un objeto 3D de acuerdo con una realización de la invención;

las Figs. 5A – 5F ilustran resultados de ejemplo para el diagrama de flujo simplificado para la colocación del objeto 3D de acuerdo con una realización de la invención;

las Figs. 6A – 6B ilustran un mapa de profundidad y su combinación con un mapa de disparidad de segmentación; y

la Fig. 7 ilustra un diagrama simplificado de un sistema de proyector 3D de acuerdo con una realización de la invención.

Descripción detallada

La siguiente descripción y dibujos son ilustrativos de la invención y no deben ser considerados como limitativos de la invención. Se describen numerosos detalles específicos para proporcionar una profunda comprensión de la invención. No obstante, en ciertos casos, detalles bien conocidos o convencionales no se describen, con el fin de evitar oscurecer la descripción de la invención. Las referencias a uno o de una realización en la presente descripción no son necesariamente referencias para la misma realización; y, tales referencias significan al menos una.

La Fig. 1 ilustra un sistema de colocación 3D 100 de acuerdo con una realización de la invención. Se utilizan datos de posición para determinar dónde colocar un objeto cuando se renderiza en una imagen 3D objetivo. El objeto puede incluir cualquiera de: texto de subtítulo, subtitulación para sordos, barra de información en pantalla, noticias deslizantes, guía de programación de televisión, marcador, controles del dispositivo de reproducción en pantalla (controles táctiles de pantalla mostrados opcionalmente), logo de canal superpuesto, información de pantalla de datos (HUD, Heads-Up Display) u otros objetos gráficos. Con el fin de permitir la independencia del dispositivo durante la reproducción, se utiliza un sistema de colocación basado en porcentajes. La posición en el espacio 3D se expresa como un porcentaje del ancho de pantalla (eje x), altura de pantalla (eje y) y distancia del espectador desde la pantalla (eje z).

Por ejemplo, un objeto estereoscópico con disparidad positiva de eje x de 19 píxeles entre las imágenes para ojo izquierdo y para ojo derecho en una imagen de un ancho de 2048 píxeles representa aproximadamente el 0,927% (19/2048) del ancho de la imagen. El resultado, por ejemplo, es que el objeto estereoscópico parece estar delante del plano de pantalla, a medio camino entre el espectador y la pantalla (por ejemplo, aproximadamente a 50% de la distancia del espectador desde el plano de pantalla). Para una pantalla de 7 metros de ancho, el objeto estereoscópico puede estar separado de acuerdo con una realización de la invención un ancho de pantalla del 0,927%, 64,94 milímetros (7 metros * 0,927%) o aproximadamente 27 píxeles (suponiendo un ancho horizontal de 1998 píxeles). Esto resulta de nuevo en que el objeto estereoscópico parece estar aproximadamente a medio camino entre el espectador y la pantalla, como se indica mediante la relación trigonométrica de la Fig. 2A y el punto 201 de la Fig. 2B. En contraste, una disparidad de píxel de 19 positivo en la pantalla de 7 metros de ancho, que representa 0,951% (19/1998) del ancho de pantalla, resulta en una profundidad perceptual apreciablemente más cercana al plano de pantalla, como se muestra mediante el punto 203 de la Fig. 2B.

Como puede verse a partir del ejemplo anterior, la colocación proporcional mantuvo la percepción de profundidad para diferentes dispositivos, mientras que la disparidad absoluta de píxel no lo hizo. Resulta deseable que la profundidad perceptual en el eje z se mantenga constante independientemente de la dimensión del dispositivo de visualización. En otras palabras, la imagen 3D debe ser consistente cuando se ve desde diferentes dispositivos de reproducción, aunque la disparidad de píxel puede cambiar. En realidad, los cambios en la disparidad de píxel típicamente no serán lineales de un dispositivo de reproducción a otro.

La independencia del dispositivo de la profundidad percibida resulta ventajosa dado que la distancia de un espectador en el sistema de posición 3D 100 puede variar significativamente. Por ejemplo, la distancia del espectador puede variar de aproximadamente 15 centímetros aproximadamente a 1 metro para un dispositivo manual (por ejemplo, un teléfono móvil, un asistente personal digital (PDA, Personal Digital Assistant, un reproductor de medios portátil – iPod de video de Apple, y otros). La distancia del espectador puede aumentar para los ordenadores portátiles, ordenadores de sobremesa, televisiones y pantallas de juegos arcade aproximadamente a 0,5 metros, aproximadamente a 5 metros. En espectáculos públicos, las distancias del espectador pueden variar de aproximadamente 3 metros, aproximadamente a 100 metros o más para proyectores de cine, pantallas de estadios y vallas publicitarias. Los dispositivos de reproducción, en

consecuencia, tienen una gran variación en los tamaños de pantalla, sobre la base de la distancia prevista del espectador. Un ancho de pantalla para un dispositivo de reproducción, comparado con otro dispositivo de reproducción puede ser mayor, 5 veces mayor, 10 veces mayor o 100 veces mayor, o más.

De acuerdo con una realización de la invención, los valores de posición en el eje z, (profundidad) pueden definirse como un porcentaje de la distancia del espectador desde el plano de pantalla, Z_{pv} . De esta manera, los valores de la profundidad de posición pueden ser independientes del dispositivo. Por ejemplo, con referencia de nuevo a la Fig. 1, un primer objeto a + 50% (punto 101) será percibido siempre a medio camino entre el espectador y el plano de pantalla, independientemente de la posición absoluta del espectador o de la dimensión de la pantalla. Segundos y terceros objetos a 0% (punto 103) y -50% (punto 105) siempre estarán al doble de distancia (en el plano de pantalla) y triple de la distancia (detrás del plano de pantalla) de la distancia al primer objeto, respectivamente. A medida que un espectador se acerca al plano de pantalla comparando un dispositivo de reproducción con otro, un objeto parecerá que está más cerca de ella en términos absolutos. A medida que se aleja del plano de pantalla, el objeto parece estar más alejado. No obstante, el objeto aparecerá en gran manera en la misma proporción de la distancia del espectador.

En una realización particular, para traducir Z_{pv} en un valor utilizable por un dispositivo de reproducción, el ancho de pantalla físico, w_s del dispositivo de visualización supuesto utilizado para la colocación de la profundidad es conocido o comunicado. En otras palabras, el dispositivo de reproducción recibirá datos de posición como al menos valores x, y, Z_{pv} y w_s . Con la información proporcionada, el dispositivo de reproducción puede calcular la disparidad de pixel apropiada en el eje x para sus propias dimensiones de pantalla, para conseguir una profundidad percibida Z_{pv} .

En otra realización, puede utilizarse una pantalla normalizada de 1 metro de ancho, tal como una pantalla de 1000 píxeles de ancho con una resolución de 1 píxel por milímetro. La normalización proporciona la ventaja de que el dispositivo de reproducción solo necesita conocer el ancho de su propia pantalla para disponer apropiadamente un objeto en profundidad, y el software de composición puede renderizar virtualmente (por ejemplo, no se utiliza ninguna pantalla física para tomar la decisión de la profundidad del objeto) utilizando el ancho normalizado de la pantalla. Es decir, no es necesario comunicar un valor de w_s puesto que es conocido a priori.

Debe entenderse que esta presentación numérica de Z_{pv} puede producir valores que superan la capacidad de un dispositivo de reproducción particular, dispositivos heredados de baja resolución, principalmente pequeños. Por ejemplo, las pantallas de video de los teléfonos móviles pueden ser de tan solo 26 milímetros de ancho (o incluso menores) y están así limitadas a una profundidad máxima de +29% de profundidad. Los dispositivos de reproducción contemporáneos ofrecen tamaños mayores de pantalla con una resolución mucho mayor y son capaces de renderizar a mayor profundidad – aunque estos dispositivos no pueden alcanzar aún +100% de la distancia del espectador. Esta limitación no supone en la práctica un obstáculo, dado que +100% de profundidad es casi siempre no deseable. La disparidad de pixel de un objeto situado muy cerca de un espectador hace difícil que el espectador enfoque y haga converger el objeto.

Adicionalmente, la presentación numérica de Z_{pv} , un porcentaje de la distancia del espectador, no puede expresar adecuadamente la profundidad en o más allá del plano de infinito. Esta carencia se obvia apreciando que parecerá que un objeto se encuentra en el infinito cuando los ejes visuales de los ojos de un espectador son paralelos. De este modo, puede especificarse que el plano de infinito esté en o aproximadamente en un valor negativo de separación interocular (aproximadamente -65 milímetros para un adulto). Para una pantalla normalizada de 1 píxel / milímetro, puede establecerse que el plano de infinito tenga un desfase de separación de píxel de o aproximadamente de -65 píxeles.

Utilizando una disposición de la colocación 3D 300 para una pantalla normalizada como se muestra en la Fig. 3, un dispositivo de reproducción o de composición puede insertar apropiadamente un objeto, tal como subtítulos, en una imagen 3D cuando dispone de tres valores de posición: x como porcentaje del ancho de la pantalla, y como porcentaje de la altura de la pantalla y z_n como porcentaje de la profundidad normalizada percibida. Un desfase de separación de píxel normalizado, S_p , puede entonces calcularse al menos como sigue, sin limitación:

$$S_p = z_n \left(\frac{65}{100 - z_n} \right),$$
 en la que $0 \leq z_n < 100$ (es decir, el objeto se encuentra en o delante del plano de pantalla hacia la posición del espectador); y

$S_p = 0,65z_n$, en la que $z_n < 0$ (es decir, el objeto se encuentra detrás del plano de pantalla lejos de la posición del espectador).

El desfase de separación de píxel normalizado permite la colocación de un objeto en el espacio 3D con respecto a una profundidad percibida por el espectador independientemente del tamaño de la pantalla o de la distancia del espectador. Un dispositivo de reproducción puede utilizar el desfase de separación de píxel normalizado (S_p) recibido, por ejemplo como metadatos en un flujo de bits, para calcular el desfase de separación de píxel específico para un dispositivo (S_d) ajustándose con un factor de su propia distancia entre píxeles. Si la distancia entre píxeles del dispositivo de reproducción es de 0,5 milímetros en lugar de 1 milímetro de la pantalla normalizada, $S_d = S_p / 0,5$, en este ejemplo.

La Fig. 4 ilustra un diagrama de flujo 400 simplificado para la colocación de un objeto 3D, por ejemplo, mediante un dispositivo de composición de acuerdo con una realización de la invención. En la etapa 401, se recibe una imagen estereoscópica preexistente, la imagen objetivo. La posición y el alcance de los elementos de la imagen se identifican en la etapa 403, que puede ser implementada por ejemplo mediante análisis de segmentación en cada par de imágenes estereoscópicas (por ejemplo, imágenes para ojo izquierdo y para ojo derecho). El análisis de segmentación separa cada elemento (por ejemplo un objeto preexistente en la imagen) por colores y, a continuación, aplica una frontera a cada área distinta para formar un mapa de segmentación.

A continuación, en la etapa 405, se determinan los desfases espaciales para los elementos. Esto puede conseguirse comparando los mapas de segmentación de la imagen de ojo izquierdo con la imagen de ojo derecho. Los mapas de segmentación para las imágenes de ojo izquierdo y de ojo derecho son similares y, de acuerdo con esto, los elementos y las formas de los segmentos en un mapa del par se refieren naturalmente a elementos o segmentos de forma y posición similares en su complemento. El par de mapas puede compararse para determinar el desfase espacial en el eje x entre elementos complementarios del par.

Estos desfases espaciales se utilizan en la etapa 407 opcional para desarrollar un mapa legal. El mapa legal define los volúmenes 3D disponibles para la colocación de un objeto. Es decir, el mapa legal indica volúmenes no ocupados por los elementos. Este mapa legal puede ser actualizado dinámicamente mediante una duración de persistencia del objeto para ser colocado en la imagen objetivo.

En las etapas 409 y 411, se recibe un objeto para colocar y se determina su geometría para definir los límites. En los casos simples, la geometría del objeto puede definirse mediante un simple cuadro delimitador alrededor de todo el objeto con cualquier interespacio requerido (por ejemplo, distancia mínima entre el objeto y cualquier elemento de la imagen objetivo) como se explica a continuación para la Fig. 5F. En los casos complejos, la geometría del objeto puede requerir una definición más precisa que un único cuadro delimitador. Por ejemplo, la geometría del objeto puede estar caracterizada por un cuadro delimitador conformado para cada carácter tal como se explica a continuación para la Fig. 5E.

A continuación, en la etapa 413 de la Fig. 4, el objeto se sitúa en el volumen 3D obedeciendo a reglas de colocación simples, que están diseñadas preferiblemente para minimizar la fatiga en el sistema visual humano. Las reglas de colocación simples generalmente tienen uno o más de los siguientes atributos: no conflictos con elementos en la imagen original; el objeto siempre está delante (como alternativa, el objeto siempre está detrás); geometría simple del objeto; tamaño de texto o de objeto; profundidad máxima y mínima; y colocación estática.

Las reglas de colocación simples pueden ser un origen y un conjunto de opciones jerarquizadas. Por ejemplo, una regla de colocación simple para que un objeto esté centrado en la pantalla en el plano de pantalla puede ser:

Origen: 0: $x=0$ $y=0$ $z=0$
 Opciones: 1: $x=-20$
 2: $x=+20$
 3: $x<0$
 4: $x>0$
 5: $0<z\leq 50$
 5a: $x<0$
 5b: $x>0$
 6: Error

Si la posición $x=0$, $y=0$, $z=0$ no está disponible, el 20% derecho sigue preferiblemente al 20% izquierdo, en cualquier sitio a la izquierda, en cualquier sitio a la derecha, en cualquier sitio delante y arriba hasta un máximo del 50%, después hacia la izquierda, y finalmente hacia la derecha. Si ninguna de estas ubicaciones preferentes está disponible, puede devolverse un error.

En la etapa 415, se siguen opcionalmente reglas de colocación complejas en lugar de, o además de, la etapa 413. Las reglas de colocación complejas pueden implicar cambio dinámico para permitir que el objeto tenga movimiento, basándose en el movimiento de los elementos en la imagen original. La colocación compleja típicamente requiere una definición más precisa que un simple cuadro delimitador y, por ello, se utiliza a menudo geometría compleja de objeto. En realidad, en la colocación compleja, la geometría del objeto puede diferir del original debido a la forma del objeto para cumplir las preferencias complejas de colocación, como se ilustra en la Fig. 5F que se explicará a continuación.

La colocación compleja puede caracterizarse de manera general por tener uno o más de los siguientes atributos: reglas dinámicas, conflictos permitidos con objetos en la imagen original; oclusión permitida; geometría compleja de objeto; y colocación dinámica. Un ejemplo de la colocación compleja es un objeto gráfico dinámico que parece moverse desde (i)

detrás de un elemento de la imagen original hacia (ii) delante del elemento y acercándose al espectador. En este ejemplo, el objeto gráfico dinámico está sometido a un escalado, rotación y movimiento dinámicos, siendo la geometría del objeto también dinámica. Inicialmente colisiona con, o está oculta por, el elemento original de la escena, y después evita y oculta el elemento original.

Finalmente, en la etapa 417 de la Fig. 4, los valores de posición reflejan que puede obtenerse la colocación del objeto. Los valores de posición se expresan como un porcentaje del ancho de la pantalla, un porcentaje de la altura de la pantalla y un porcentaje de la profundidad percibida. El porcentaje de la profundidad percibida puede estar normalizado para un dispositivo de visualización de representación. En una realización específica, el dispositivo de visualización de representación está caracterizado por una pantalla de 1000 píxeles de ancho normalizada con una resolución de ancho de 1 milímetro por píxel.

Debe apreciarse que el diagrama de flujo 400 descrito en esta memoria es solamente con el propósito de ilustración y que pueden sugerirse varias modificaciones o cambios a la luz del mismo a los expertos en la materia. En implementaciones alternativas, las etapas indicadas en el diagrama de flujo 400 pueden realizarse en un orden distinto al indicado en la Fig. 4, pueden incluir etapas adicionales y/o pueden omitir algunas etapas a la vez. Por ejemplo, las etapas 401 y 409 pueden en realidad ser ejecutadas substancialmente a la vez o en orden inverso. Todas esas modificaciones y variaciones pretenden estar incluidas dentro del alcance de esta descripción.

Las Figs. 5A – 5F ilustran resultados de ejemplo para el diagrama de flujo 400 simplificado anterior, para la colocación de un objeto 3D de acuerdo con una realización de la invención. La Fig. 5A muestra imágenes de ojo izquierdo y de ojo derecho de la imagen estereoscópica preexistente que pueden ser recibidas en la etapa 401. Los mapas de segmentación descritos en la etapa 403 se muestran en la Fig. 5B, y se superponen en la Fig. 5C para mostrar la disparidad de separación entre los dos. Un “Objeto de texto” con un límite simple, un cuadro simple, se ilustra en la Fig. 5D correspondiente a la etapa 411 si se utiliza la geometría simple. En comparación, la Fig. 5E muestra el mismo “Objeto de texto” con geometría compleja de objeto, en la que los límites están definidos de manera precisa para cada carácter. Para la etapa 415, si se utiliza tanto la colocación compleja como la geometría compleja, el “Objeto de texto” puede ser conformado y muestra atributos de movimiento como se indica en la Fig. 5F.

Aunque no se muestra en el diagrama de flujo 400, la colocación del objeto 3D mediante, por ejemplo, un dispositivo de composición puede incluir además utilizar un mapa de profundidad de acuerdo con una realización de la invención. Los mapas de profundidad pueden refinar las determinaciones de profundidad para los elementos de la imagen preexistente, particularmente para las imágenes 3D desde múltiples puntos de vista (por ejemplo, imágenes poliscópicas), puesto que los mapas de profundidad definen de manera más precisa la forma 3D de los objetos en un análisis de desfase espacial. La Fig. 6A es un ejemplo de un mapa de profundidad 600 para el par de imágenes estereoscópicas de la Fig. 4A. El mapa de profundidad 600 es monocromo y la profundidad se representa mediante luminancia. Una luminancia mayor es z positiva, hacia el espectador y una luminancia menor es z negativa, alejándose del espectador. Los mapas de profundidad pueden incluir asimismo valores de “confianza” asociados que definen la precisión de la profundidad medida en cada porción de la imagen. El mapa de profundidad 600 solo puede ser factorizado con sus valores de confianza en las reglas de colocación de objeto para mejorar la probabilidad de evitar conflictos entre el objeto insertado y los elementos en la imagen original.

La Fig. 6B ilustra la utilización combinada del análisis de segmentación con el mapa de profundidad 600. Como se ha descrito anteriormente, el mapa de profundidad 600 proporciona determinaciones de formas 3D mejoradas; no obstante, los mapas de disparidad de segmentación definen de manera más precisa la profundidad perceptual de los elementos en la imagen original. De este modo, resulta ventajoso combinar la disparidad de segmentación con los mapas de profundidad para un análisis superior de la imagen 3D original. El resultado es que dentro de las fronteras de la segmentación, la forma está definida por el mapa de profundidad. Efectivamente los valores de confianza del mapa de profundidad han mejorado mucho.

En otra realización de la invención, un dispositivo de reproducción o un dispositivo de composición aplican técnicas normalizadas para ajustar una imagen estereoscópica existente para un rango de profundidad de visualización cómodo. En otras palabras, una profundidad disponible delante y detrás de un plano de pantalla se ajusta para un tamaño de pantalla, así como una preferencia individual (por ejemplo, la preferencia del artista en tres dimensiones, o el usuario final del dispositivo de reproducción). Sin perjuicio de que un rango de profundidad de visualización cómodo sea una preferencia individual, a continuación se muestra un conjunto de reglas razonable y aproximado:

(i) Pantalla muy grande (por ejemplo, Cine / I-Max). Toda la profundidad aprovechable delante, donde el infinito está en el plano de pantalla y la máxima profundidad positiva es el 60% de distancia del espectador.

(ii) Pantalla grande (por ejemplo, Cine). Más profundidad aprovechable delante de la pantalla que detrás (por ejemplo, aproximadamente 2/3 de la profundidad delante, 1/3 de la profundidad detrás).

(iii) Pantalla pequeña (por ejemplo, televisión, cine doméstico). Más profundidad aprovechable detrás del plano de pantalla que delante (por ejemplo aproximadamente 1/3 de la profundidad delante, 2/3 de la profundidad detrás).

(iv) Pantalla muy pequeña (por ejemplo, pantalla de teléfono móvil o televisión portátil). El infinito lo más lejos posible detrás del plano de pantalla, sin nada de profundidad positiva delante de la pantalla.

Suponiendo, como ejemplo para ilustrar las reglas anteriores, que la imagen estereoscópica original fue masterizada con el punto de infinito en $Z_{ni} = -100$, y que la profundidad positiva máxima es $Z_n = -40$ (siendo la imagen total de la escena $Z = |Z_{ni} - Z_n| = 60$), entonces la posición en profundidad para el infinito corregida, Z_{ci} , es como sigue:

(i) Pantalla muy grande. $Z_{ci} = Z_{ni} + 100$; infinito en el plano de pantalla, máxima salida del objeto fuera de la pantalla, 60% de la distancia del espectador.

(ii) Pantalla grande. $Z_{ci} = Z_{ni} + 80$; infinito 20% por detrás del plano de pantalla, máxima salida del objeto fuera de la pantalla, 40% de la distancia del espectador.

(iii) Pantalla pequeña. $Z_{ci} = Z_{ni} + 60$; infinito 40% por detrás del plano de pantalla, máxima salida del objeto fuera de la pantalla, 20% de la distancia del espectador.

(iv) Pantalla muy pequeña. $Z_{ci} = Z_{ni} + 40$; infinito 60% por detrás del plano de pantalla, no hay salida del objeto fuera de la pantalla.

Un dispositivo que ajusta una imagen estereoscópica puede utilizar uno o más de un valor de profundidad normalizada en metadatos, indicando una marca que la imagen ha sido ajustada, y metadatos adicionales que describen el desplazamiento del ajuste deseado para una pluralidad de pantallas de tamaño predeterminado. Para un tamaño de pantalla no incluido en los tamaños predeterminados, el dispositivo puede interpolar un valor de desviación del ajuste intermedio utilizando los tamaños predeterminados más cercanos (por ejemplo, una proporción ponderada, promediada, promediada y redondeada o una ponderación explícita especificada mediante metadatos) o simplemente utilizar un valor asociado con el tamaño predeterminado más cercano. Cuando se utiliza esta técnica de ajuste, deben proporcionarse imágenes de ojo izquierdo y de ojo derecho sobre escaneadas en su totalidad. El sobre escaneo debe ser suficiente para permitir los ajustes en los dos puntos extremos para los valores de desviación de ajuste (por ejemplo, las pantallas más pequeñas y más grandes de los dispositivos de reproducción previstos).

Adicionalmente, el anterior ajuste de una imagen estereoscópica existente puede utilizarse para corregir deficiencias en la captura de la imagen estereoscópica. Por ejemplo, cuando el disparo ha convergido, dos planos de cámara de imagen forman un ángulo entre sí y convergen en el objeto de interés. El punto de convergencia depende así de la disposición física de las cámaras de entrada, pero la distancia al objeto de reproducción desde el espectador depende todavía de la pantalla. La convergencia puede ajustarse en pos-producción desplazando las imágenes de ojo izquierdo y de ojo derecho horizontalmente entre sí.

Como otra realización de la invención, existe un medio de almacenamiento que graba un programa de instrucciones. El medio de almacenamiento puede ser al menos uno de: memoria de acceso aleatorio (RAM, random access memory), SDRAM, memoria de solo lectura (ROM, read-only memory), memoria de solo lectura programable (PROM, programmable read-only memory), memoria de solo lectura programable, borrrable eléctricamente (EEPROM, electrically erasable programmable read-only memory), memoria rápida, disco óptico (por ejemplo, CD-ROM, CD-RW, DVD, DVD-RW, disco blu-ray, disco óptico de ultra densidad (UDO ultra density optical), disco compacto, y otros), medios magnéticos (por ejemplo, cinta magnética, disco flexible, unidad de disco duro y otros), cinta de papel (por ejemplo, tarjetas perforadas), almacén holográfico y memoria molecular. El programa es ejecutable mediante un dispositivo de visualización para ejecutar un método para generar una salida visual. El método incluye recibir datos de profundidad perceptual normalizada como fracción de la distancia del espectador para al menos un objeto. Un desfase de separación de píxel se calcula para la dimensión o dimensiones del dispositivo de visualización a partir de los datos de profundidad perceptual normalizada. Las imágenes de objeto primera y segunda del al menos un objeto se insertan en las imágenes primera y segunda de la imagen estereoscópica. Las imágenes de objeto primera y segunda están desfasadas en el desfase de separación de píxel. El dispositivo de visualización visualmente produce la imagen estereoscópica junto con el objeto insertado.

La Fig. 7 ilustra un diagrama simplificado de un sistema de dos proyectores 3D 700 de acuerdo con una realización de la invención. Las técnicas para la proyección estereoscópica por parte del sistema 700 pueden incluir cualquiera de:

(i) Anaglifo - La separación de imagen del canal de ojo izquierdo y de ojo derecho se consigue mediante un filtro de dos colores, normalmente rojo para un ojo y azul cian para el otro ojo. Los elementos necesarios para los ojos incluyen un primer filtro (por ejemplo un filtro rojo) para un ojo y un segundo filtro (por ejemplo un filtro azul cian) para el otro ojo.

(ii) Polarización lineal - La separación de imagen del canal de ojo izquierdo y de ojo derecho se consigue en el proyector filtrando la imagen de un ojo a través de un polarizador lineal orientado verticalmente y filtrando la imagen del otro ojo a través de un polarizador lineal orientado horizontalmente. Los elementos necesarios para los ojos incluyen un polarizador lineal orientado verticalmente para un ojo y un polarizador orientado horizontalmente para el otro ojo.

(iii) Polarización circular - La separación de imagen del canal de ojo izquierdo y de ojo derecho se consigue en el proyector filtrando la imagen de un ojo a través de un polarizador circular a izquierdas y filtrando la imagen del otro ojo a través de un polarizador circular a derechas. Los elementos necesarios para los ojos incluyen un polarizador circular a izquierdas para un ojo y un polarizador circular a derechas para el otro ojo.

(iv) Gafas obturadoras – La separación de la imagen del canal izquierdo y derecho se consigue multiplexando imágenes en el tiempo. Los elementos necesarios para los ojos incluyen gafas obturadoras que electrónicamente obturan la lente en sincronía con la tasa de fotogramas del proyector. Típicamente se utiliza una señal inalámbrica o de infrarrojos para proporcionar una referencia de temporización a las gafas obturadoras.

(v) La separación espectral de la imagen del canal izquierdo y derecho se consigue en el proyector filtrando espectralmente. Los filtros para la imagen de ojo izquierdo y de ojo derecho dejan pasar cada uno una porción del rojo, verde y azul del espectro, proporcionando una imagen a todo color. El espectro de paso de banda del filtro del ojo izquierdo es complementario del espectro del paso de banda del filtro del ojo derecho. Los elementos necesarios para los ojos incluyen filtros con las mismas características espectrales generales que los utilizados por el proyector.

En el sistema 700, las imágenes del canal del ojo izquierdo y de ojo derecho son obtenidas, decodificadas, recuperadas o reconstruidas a partir de los datos almacenados en la unidad de disco 701 (o recibidas de una red apropiada o una transmisión recepción) por el servidor 703. De manera similar, un objeto estereoscópico para ser insertado en las imágenes de canal de ojo izquierdo y de ojo derecho, así como sus valores de posición proporcionales, pueden ser almacenados en la unidad de disco 701 (o recibidos de una red apropiada o una transmisión recepción). El servidor 703 calcula un desfase de separación de píxel para el objeto basándose en uno o más valores de posición proporcionales, tal como una profundidad normalizada recibida, e inserta los objetos en las imágenes del canal con tal desfase.

Tras la inserción, las imágenes del canal de ojo izquierdo y de ojo derecho son proyectadas desde los proyectores de canal izquierdo y derecho 705 y 707 en la pantalla 709 para ver a través de gafas 3D 711. Si el sistema de proyector 3D 700 utiliza polarización lineal o circular para la separación del canal de ojo izquierdo y de ojo derecho, la pantalla 709 mantiene la polarización. Las pantallas del tipo de mantenimiento de la polarización disponibles para su utilización como pantalla 709 se denominan comúnmente "pantallas de plata" (silver screen) debido a su color distintivo.

El sistema 700 anterior ha sido descrito como un sistema de dos proyectores (uno para el ojo izquierdo y uno para el ojo derecho); no obstante, puede utilizarse un solo proyector de D-Cine (por ejemplo, un proyector digital). En un sistema de proyector de D-Cine único alternativo, las imágenes del canal de ojo izquierdo y de ojo derecho son multiplexadas en el tiempo. Excepto para el caso de las gafas obturadoras en el que no se requieren filtros de proyección, esto significa que los filtros de proyección deben cambiar a la frecuencia de multiplexación izquierda / derecha. Esto puede realizarse con una rueda de filtros en el proyector de D-Cine sincronizado a la frecuencia de multiplexación, o con un filtro conmutado electrónicamente.

Resultará evidente que una unidad de disco 701, una red apropiada (por ejemplo Internet, WAN, LAN, WiFi, fibra, cable de video, emisión por satélite u otros), o una pluralidad de redes apropiadas pueden proporcionar cada una los mismos datos, tales como los datos de posición proporcionales, a múltiples dispositivos de reproducción, en serie, simultáneamente o bajo demanda. Las dimensiones de la pantalla de los diferentes dispositivos de reproducción pueden variar, y es probable que varíen mucho para ciertas realizaciones de la invención. Por ejemplo, un teléfono celular con una pantalla incorporada recibe la misma imagen y datos de inserción de objeto que un aparato de televisión doméstico mucho más grande, y aun así conseguir los mismos resultados de visualización mediante un ajuste apropiado del desfase de la separación de píxel a pesar de la diferencia de tamaño.

La invención puede comprender, consistir en, o consistir esencialmente en, adecuadamente, cualquier elemento (las diferentes partes o características de la invención) y sus equivalentes, tal como se han descrito en esta memoria. Además, la invención descrita ilustrativamente en esta memoria puede ser puesta en práctica en ausencia de cualquier elemento, ya esté o no específicamente descrito en esta memoria. Son posibles numerosas modificaciones y variaciones de la invención a la luz de lo descrito anteriormente. Debe entenderse, por lo tanto, que dentro del alcance de las reivindicaciones adjuntas, la invención puede ser puesta en práctica de un modo distinto de lo descrito específicamente en esta memoria.

REIVINDICACIONES

1. Metodo para renderizar al menos un objeto en una imagen estereoscópica para un dispositivo de visualización, comprendiendo el método:

recibir datos de profundidad perceptual para el al menos un objeto;

calcular un desfase de separación de píxel basado en un ancho de pantalla conocido o comunicado del citado dispositivo de visualización a partir de los datos de profundidad perceptuales;

insertar una primera imagen de objeto de al menos un objeto en una primera imagen de la imagen estereoscópica; e

insertar una segunda imagen de objeto del al menos un objeto en una segunda imagen de la imagen estereoscópica,

en el que la primera imagen del objeto y la segunda imagen del objeto están desfasadas en el desfase de separación de píxel,

caracterizado por que los datos de profundidad perceptual se reciben como fracción de una distancia del espectador prevista, independientemente de las dimensiones del dispositivo de visualización.

2. El método de la reivindicación 1, en el que el al menos un objeto es texto de subtítulos.

3. El método de la reivindicación 1, en el que la primera imagen de objeto y la primera imagen son imágenes de ojo izquierdo, la segunda imagen de objeto y la segunda imagen son imágenes de ojo derecho.

4. El método de la reivindicación 1, que comprende además tras la inserción, producir en un flujo de datos la primera imagen y la segunda imagen.

5. El método de la reivindicación 1, en el que el ancho de pantalla conocido o comunicado es el ancho de la pantalla física del dispositivo de visualización.

6. El método de la reivindicación 1, que comprende además calcular un desfase de separación de píxel normalizado para una pantalla normalizada a partir de los datos de profundidad perceptual recibidos, teniendo la pantalla normalizada un ancho de pantalla conocido y una resolución conocida, en el que el citado calcular el citado desfase de separación de píxel comprende calcular el citado desfase de separación de píxel a partir del desfase de separación normalizado y del citado ancho de pantalla conocido o comunicado del citado dispositivo de visualización.

7. El método de la reivindicación 1, en el que la citada pantalla normalizada es una pantalla de 1000 píxeles de ancho con una resolución de 1 milímetro por píxel.

8. El método de la reivindicación 1, en el que el dispositivo de visualización es al menos uno de un teléfono móvil, un reproductor manual de medios, una televisión, un monitor de ordenador y un proyector de cine.

9. El método de acuerdo con la reivindicación 1, en el que el dispositivo de visualización es un primer dispositivo de reproducción, comprendiendo además el método:

transmitir la imagen estereoscópica a un segundo dispositivo de reproducción;

transmitir los datos de profundidad perceptual como fracción de la distancia prevista del espectador, independientemente de las dimensiones del segundo dispositivo de reproducción, al segundo dispositivo de reproducción;

que el segundo dispositivo de reproducción calcula un segundo desfase de separación de píxel basado en un ancho de pantalla conocido o comunicado del segundo dispositivo de reproducción a partir de los datos de profundidad perceptual;

que el segundo dispositivo de reproducción inserta imágenes de objeto respectivas del al menos un objeto, desfasado en el segundo desfase de separación de píxel, en respectivas imágenes de la imagen estereoscópica, en el que el ancho de pantalla del primer dispositivo de reproducción es diferente del ancho de pantalla del segundo dispositivo de reproducción, de manera que el desfase de separación de píxel calculado por el primer dispositivo de reproducción es distinto del segundo desfase de separación de píxel.

10. El método de la reivindicación 9, en el que el ancho de pantalla del primer dispositivo de reproducción es al menos 5 veces menor que el ancho de pantalla del segundo dispositivo de reproducción.

11. El método de la reivindicación 10, en el que el primer dispositivo de reproducción es un dispositivo manual.

12. El método de la reivindicación 9, en el que el segundo dispositivo de reproducción comprende dos proyectores.

13. Un aparato para generar una señal de salida, que comprende:

(a) un terminal de entrada;

(b) un terminal de salida; y

(c) circuitos de procesamiento de señal acoplados al terminal de entrada y al terminal de salida, en el que los circuitos de procesamiento de señal están adaptados para:

recibir datos de profundidad perceptual para al menos un objeto;

calcular un desfase de separación de píxel basado en un ancho de pantalla conocido o comunicado de un dispositivo de visualización utilizando los datos de profundidad perceptual; e

insertar, respectivamente, una imagen de objeto de ojo izquierdo y una imagen de objeto de ojo derecho del al menos un objeto en una imagen de ojo izquierdo y en una imagen de ojo derecho de una imagen estereoscópica,

en el que la imagen de objeto de ojo izquierdo y la imagen de objeto de ojo derecho están desfasadas en el desfase de separación de píxel;

caracterizado por que los citados datos de profundidad perceptuales se reciben como fracción de una distancia de visualización prevista, independientemente de las dimensiones del citado dispositivo de visualización.

14. Un medio de almacenamiento que graba un programa de instrucciones que es ejecutable por un dispositivo de visualización para llevar a cabo un método de acuerdo con cualquiera de las reivindicaciones 1 – 12.

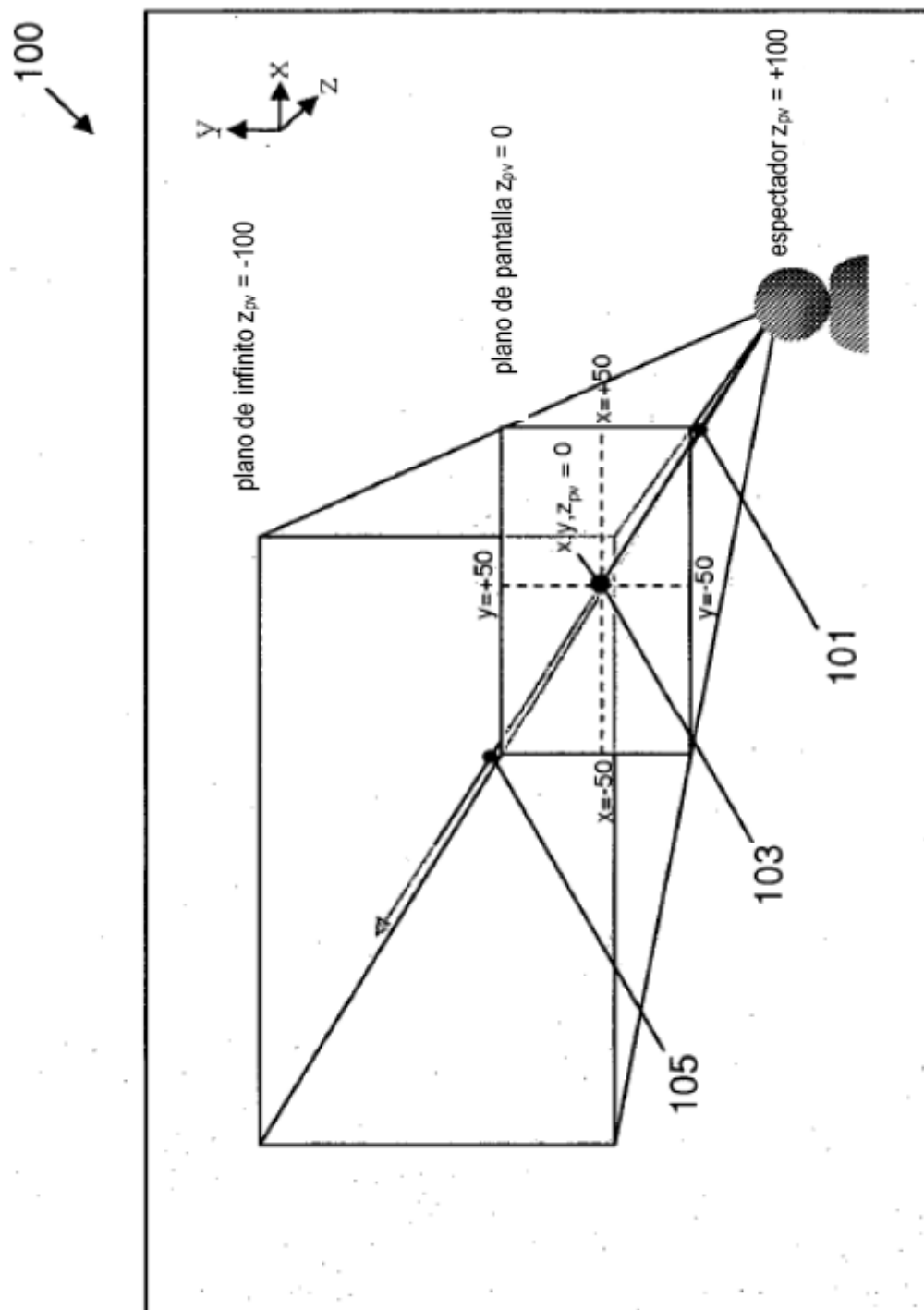


FIG. 1

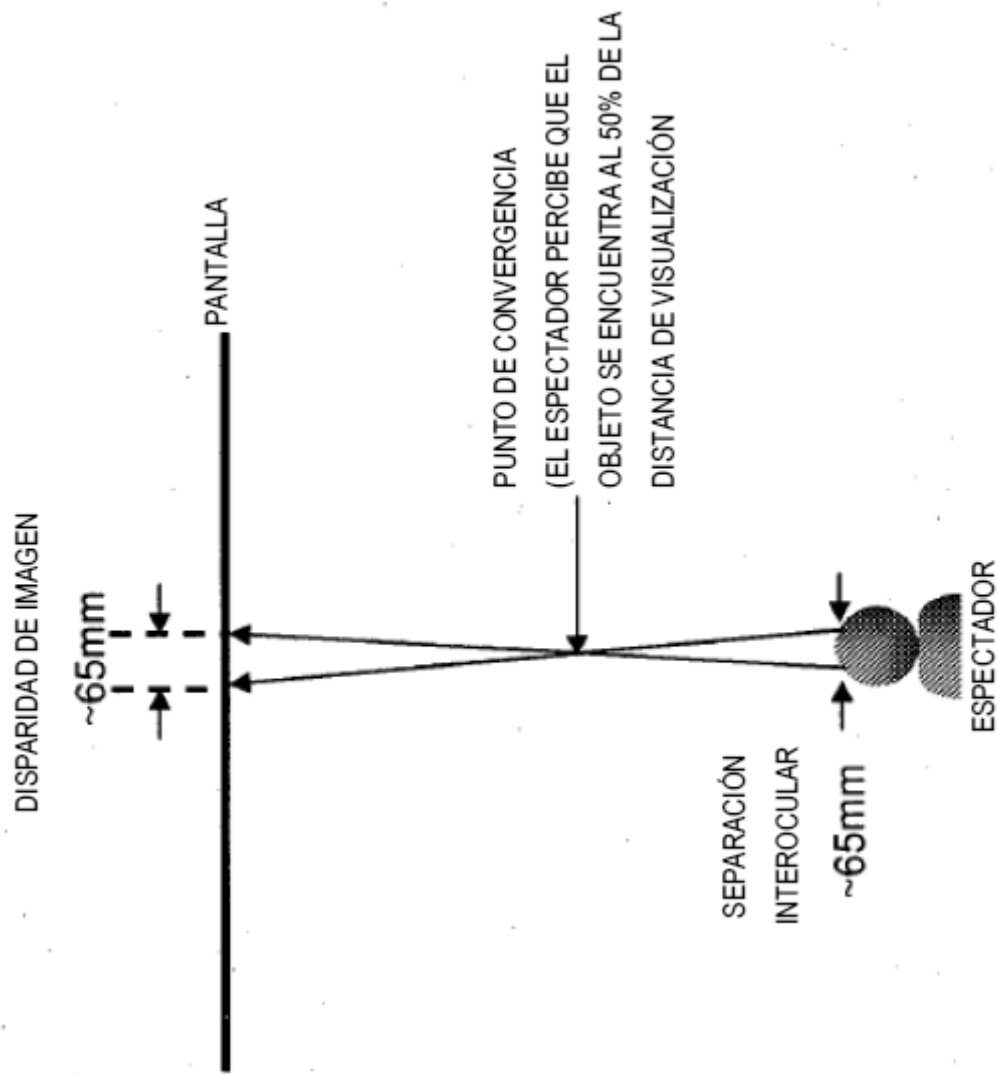


FIG. 2A

Colocación de objeto como porcentaje de la distancia del espectador desde la pantalla, calculada para una pantalla de 7 metros de ancho con una resolución horizontal de 1998 píxeles

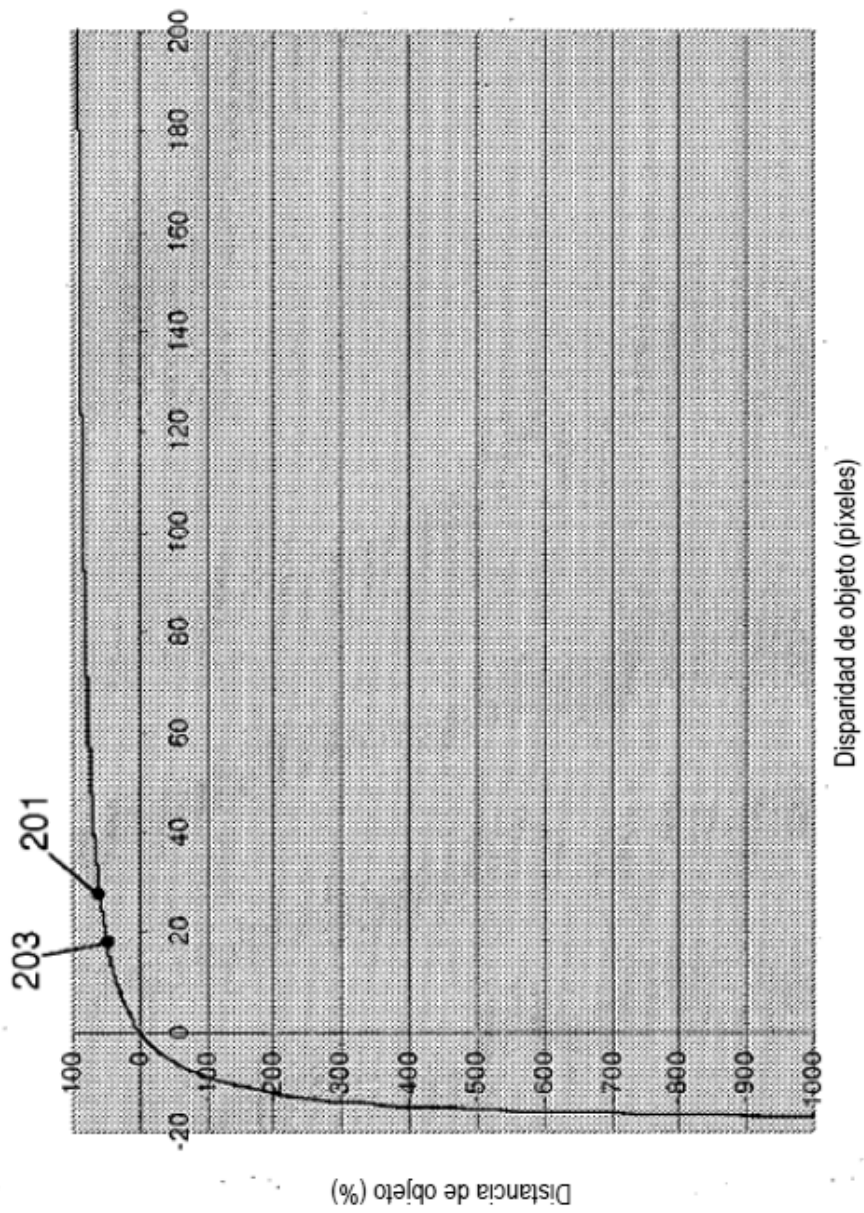


FIG. 2B

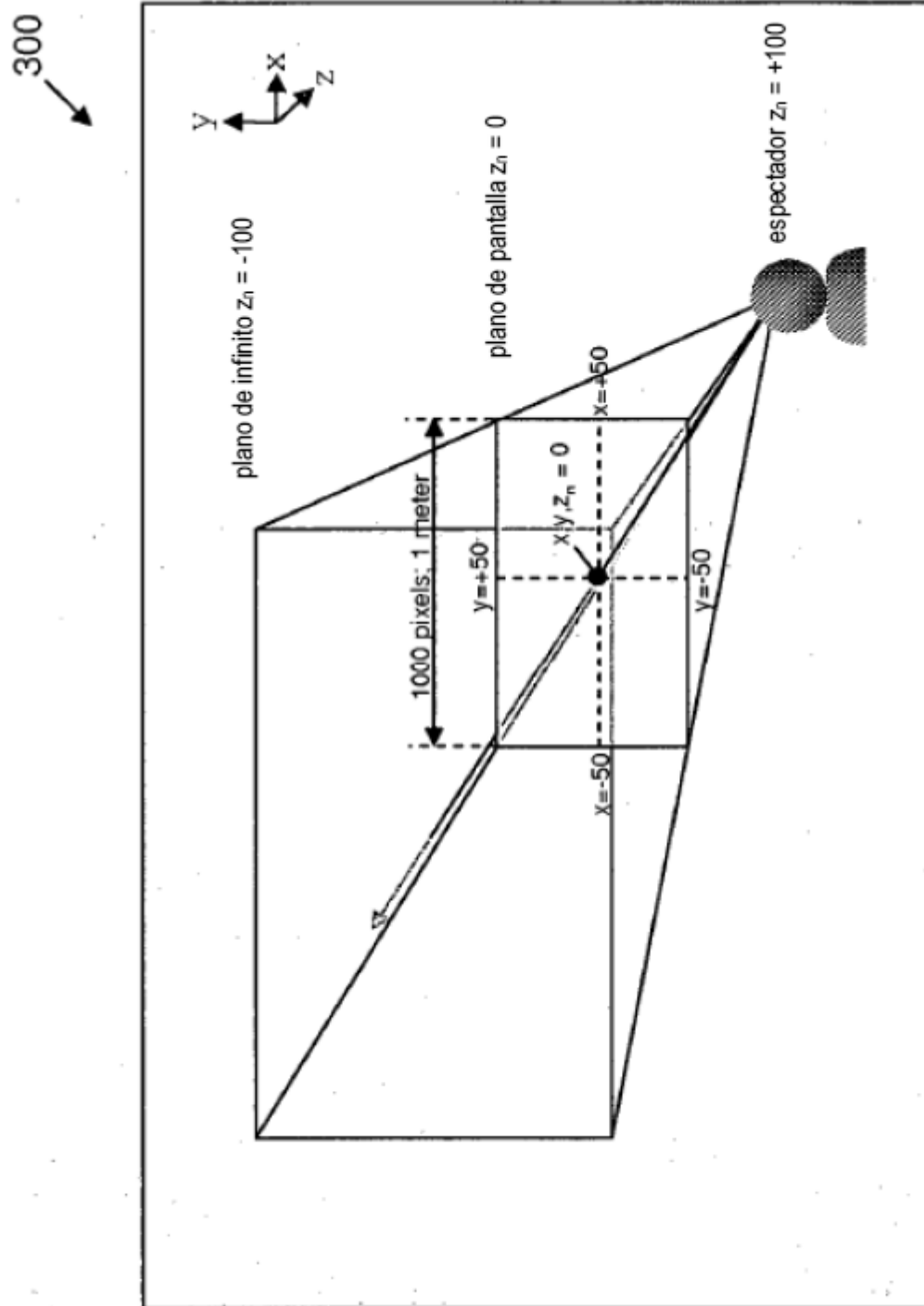


FIG. 3

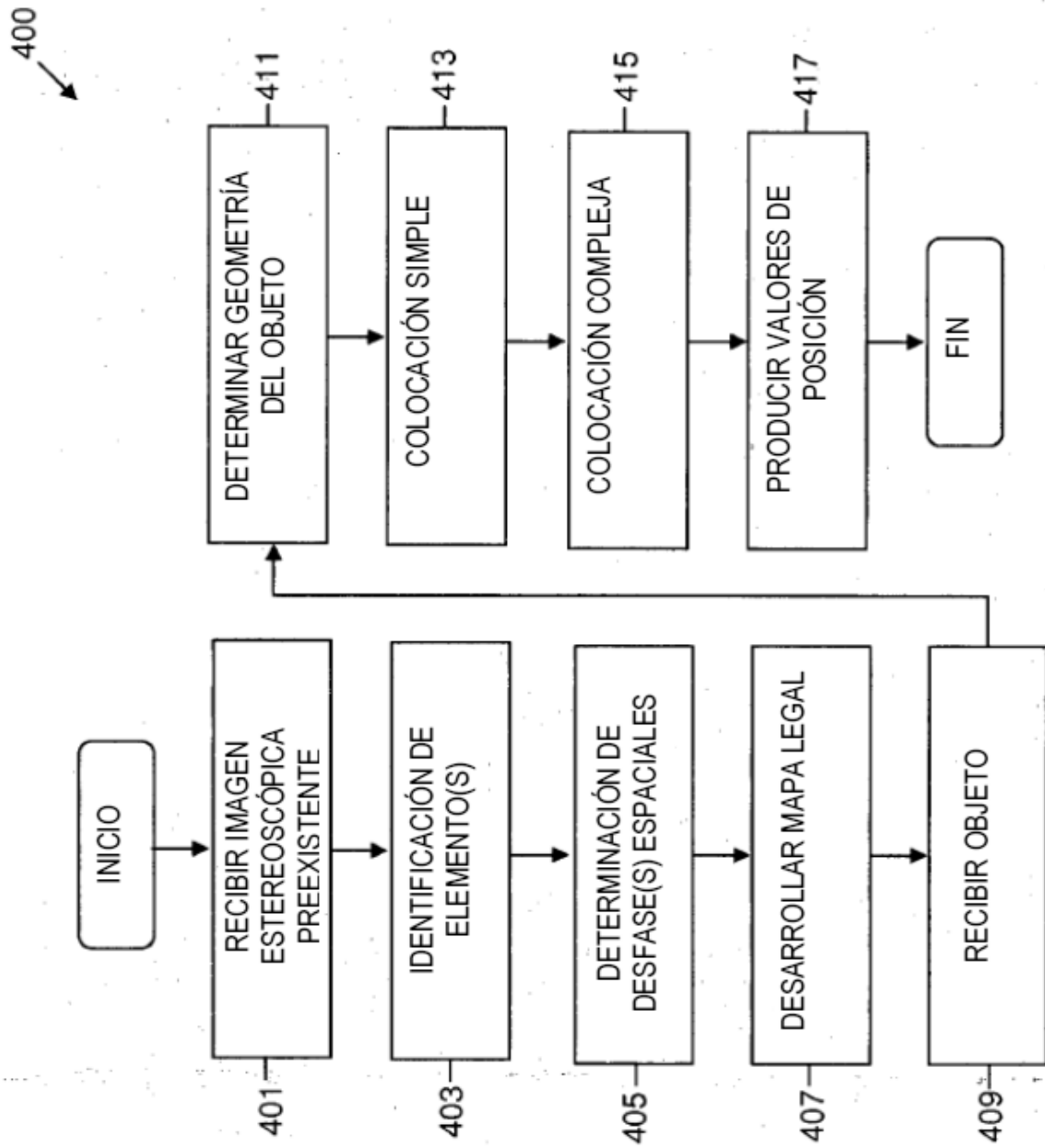


FIG. 4

IMAGEN DE OJO DERECHO ORIGINAL

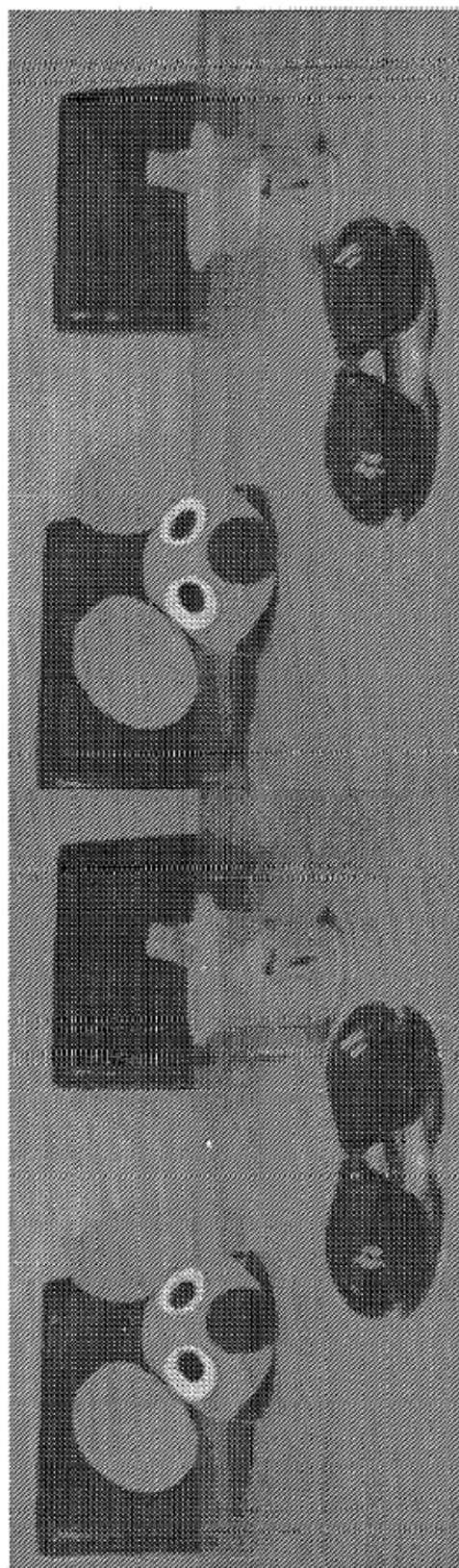


IMAGEN DE OJO IZQUIERDO ORIGINAL

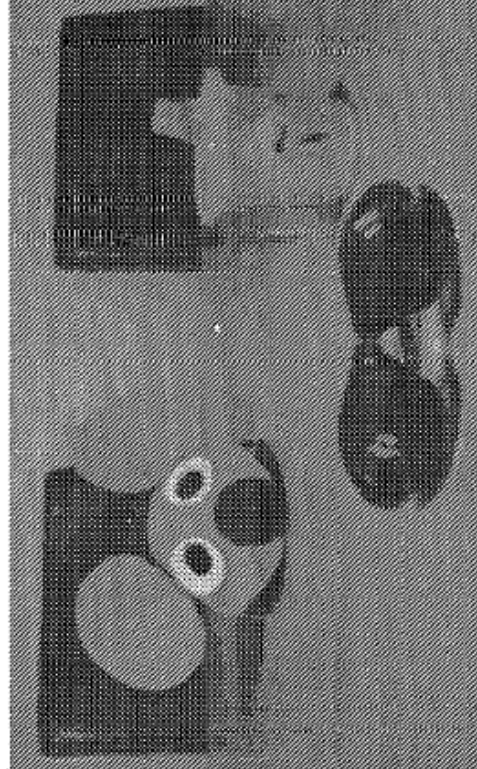


FIG. 5A

IMAGEN DE OJO IZQUIERDO SEGMENTADA IMAGEN DE OJO DERECHO SEGMENTADA

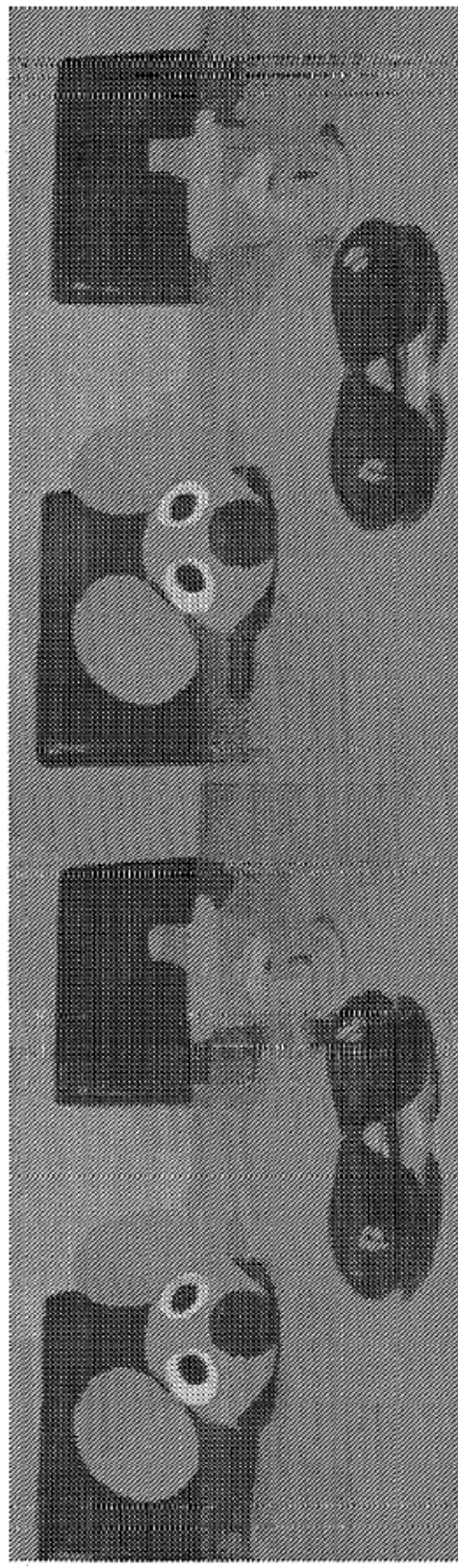


FIG. 5B

MAPA DE DISPARIDAD DE SEGMENTACIÓN SUPERPUESTA CON MEDIDAS DE SEPARACIÓN



FIG. 5C

LÍMITES SIMPLES PARA EL OBJETO

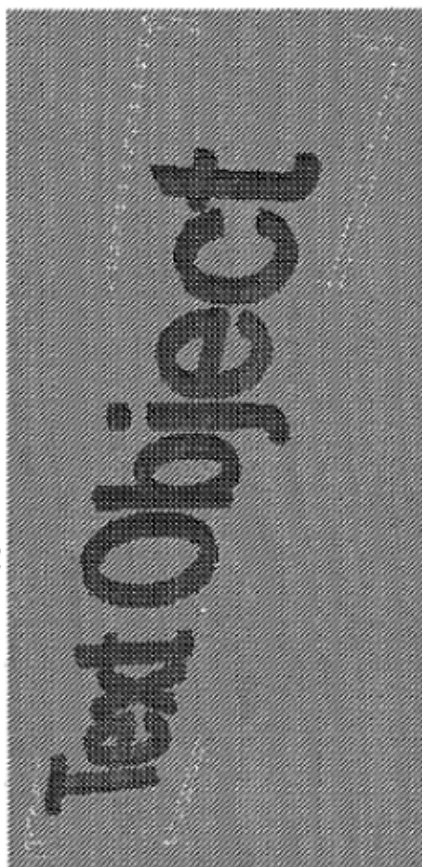


FIG. 5D

GEOMETRÍA DE OBJETO COMPLEJA

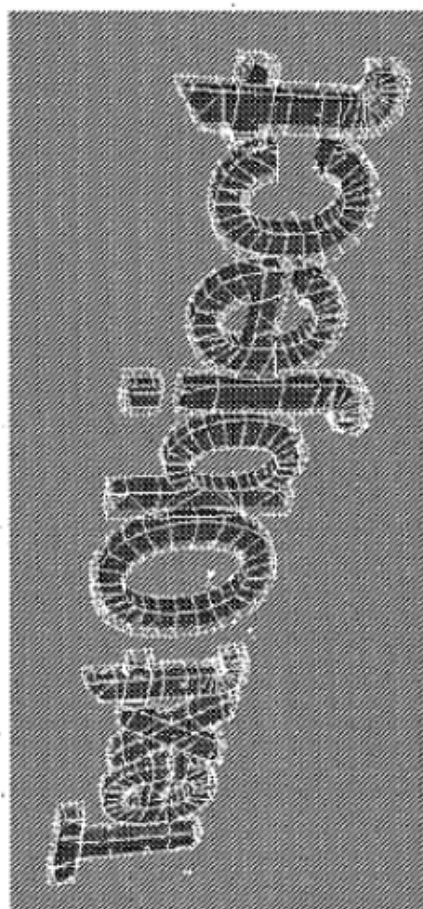


FIG. 5E

GEOMETRÍA DE OBJETO COMPLEJA
CON
GEOMETRÍA DE COLOCACIÓN COMPLEJA

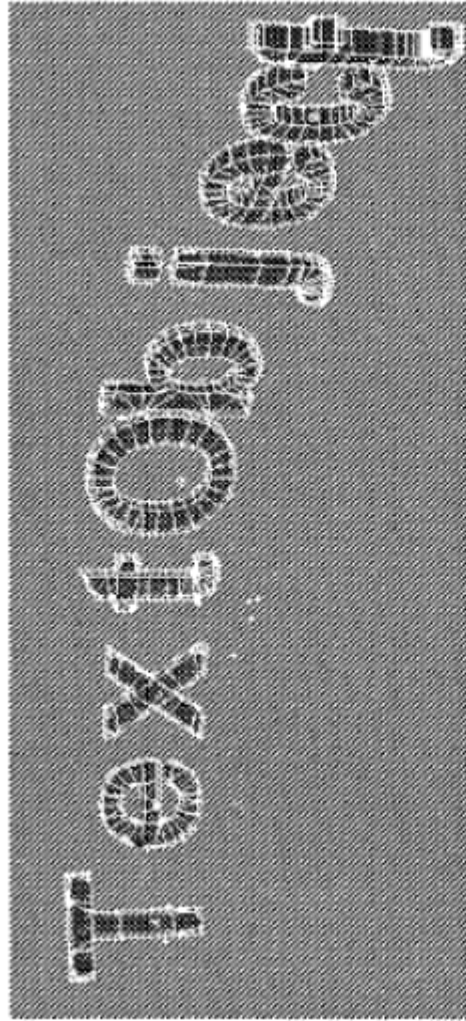


FIG. 5F

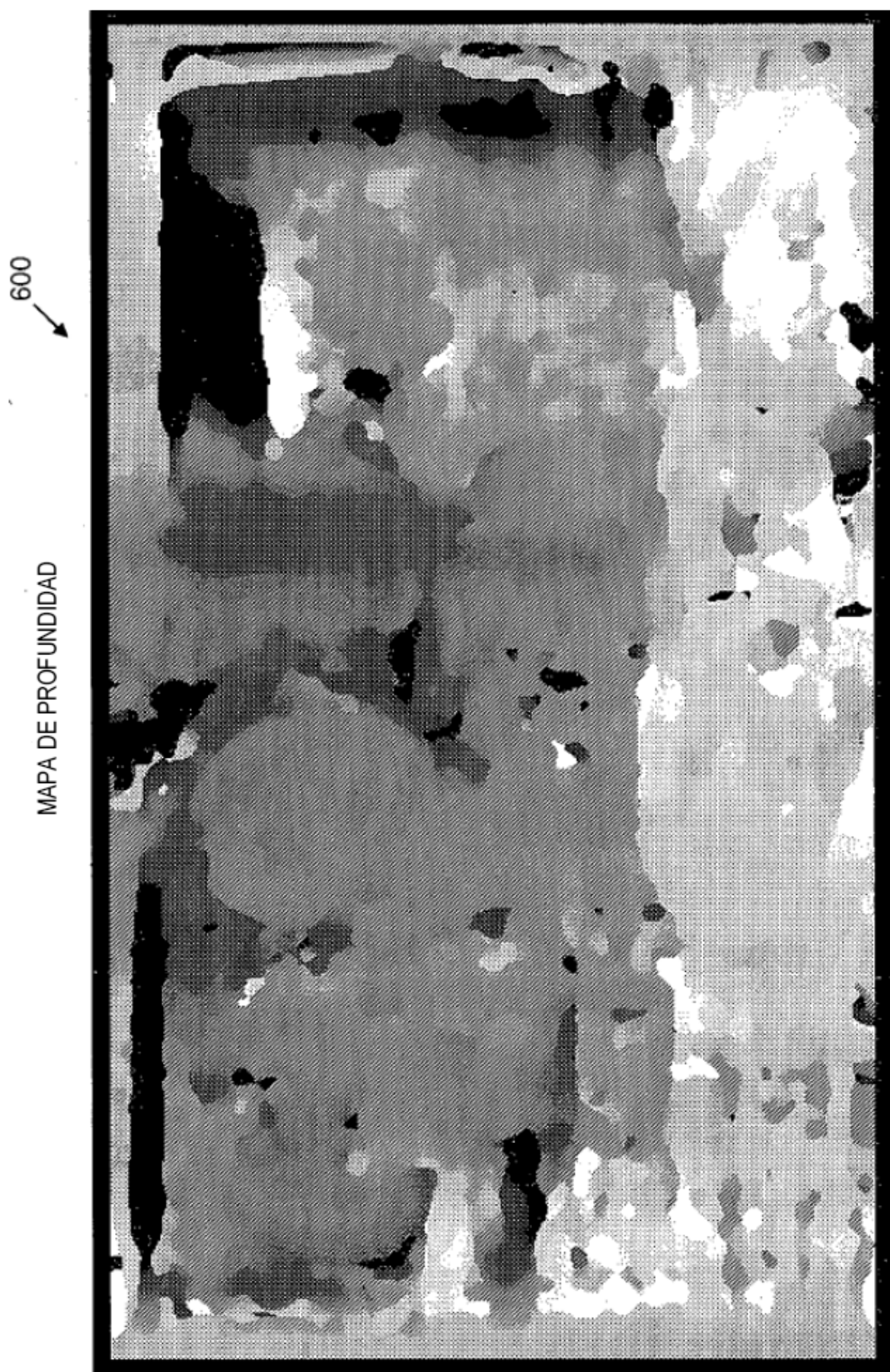


FIG. 6A

MAPA DE SEGMENTACIÓN Y DE DISPARIDAD DE PROFUNDIDAD COMBINADAS

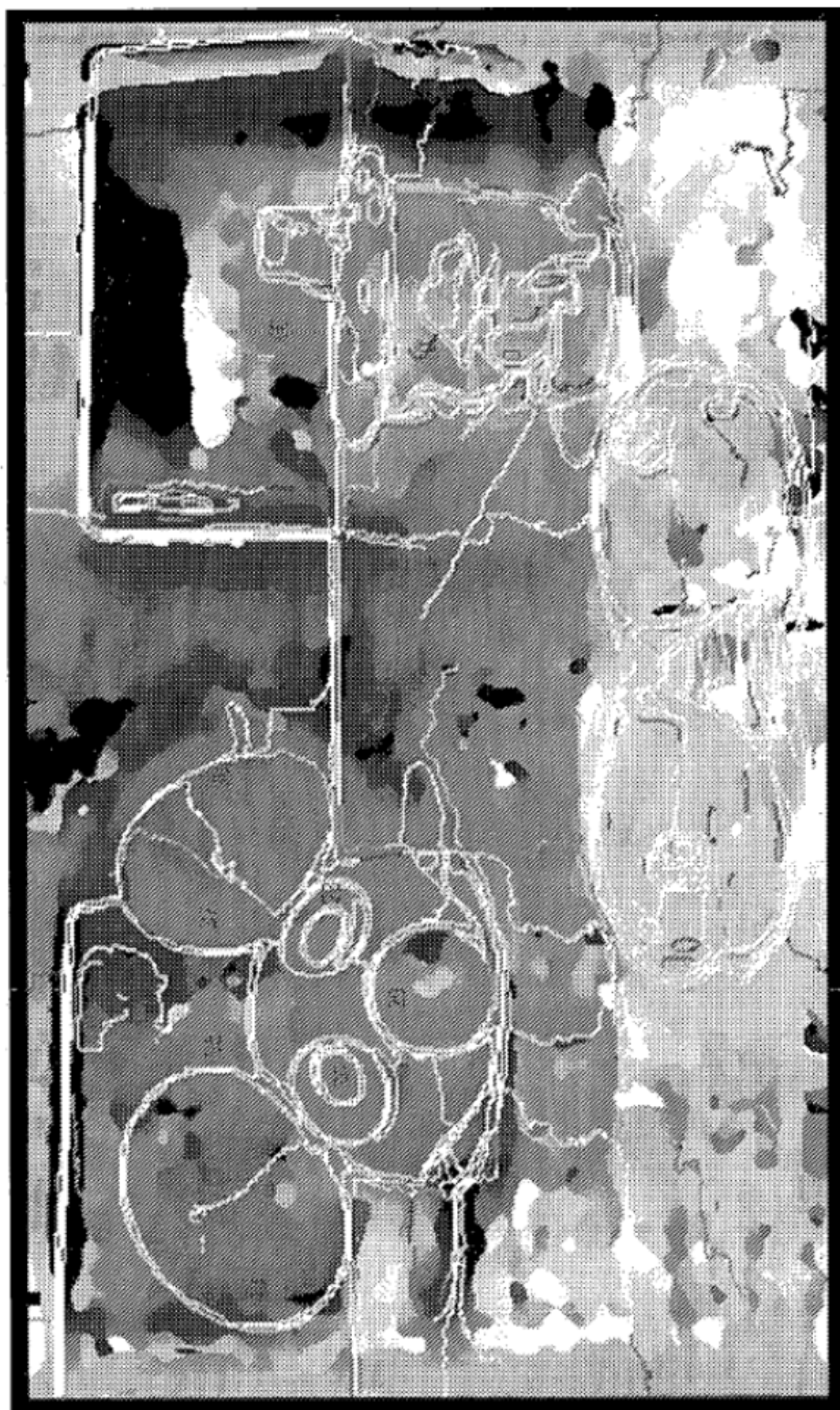


FIG. 6B

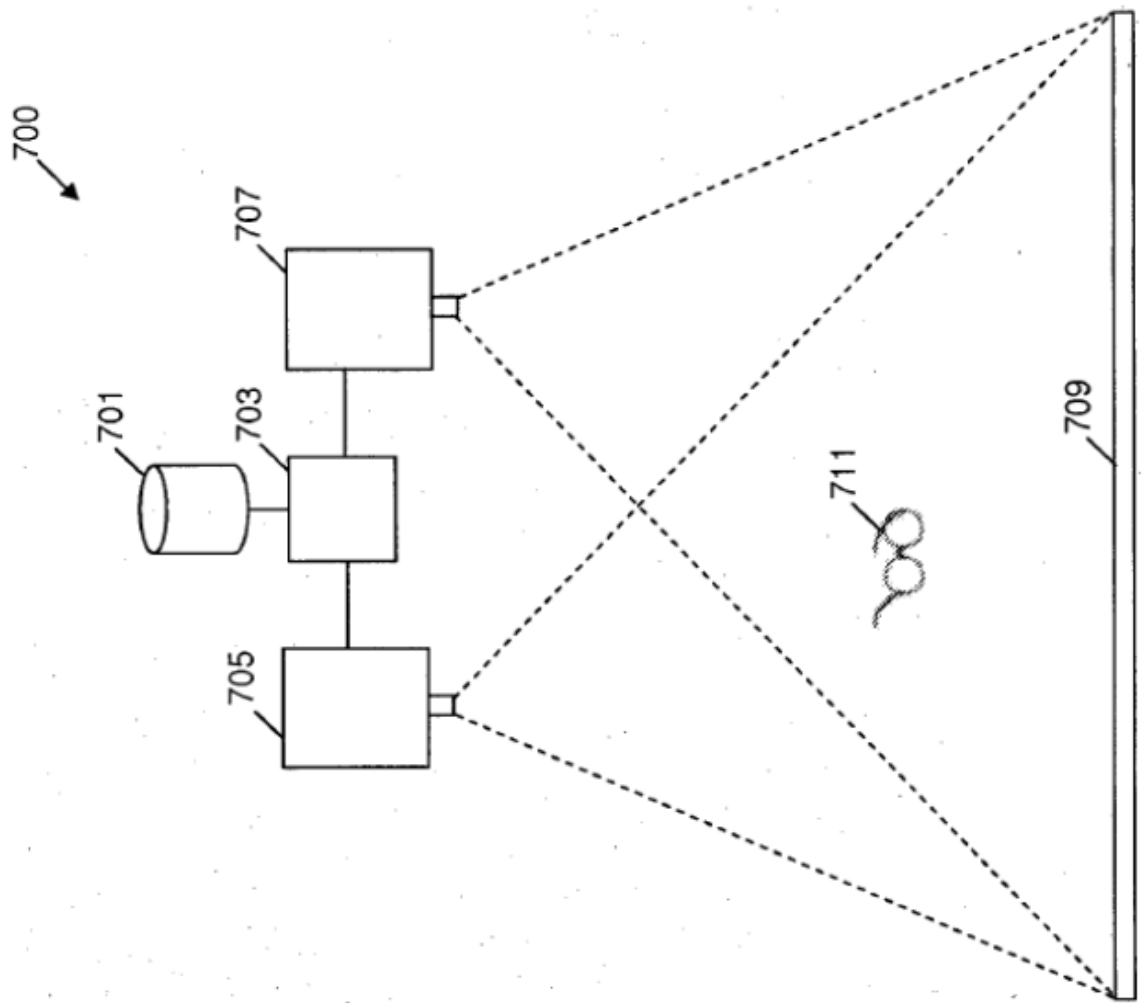


FIG. 7