

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 676 719**

51 Int. Cl.:

**G06T 13/80** (2011.01)

**H04N 5/272** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.09.2012 PCT/JP2012/072988**

87 Fecha y número de publicación internacional: **23.01.2014 WO14013627**

96 Fecha de presentación y número de la solicitud europea: **07.09.2012 E 12881432 (4)**

97 Fecha y número de publicación de la concesión europea: **02.05.2018 EP 2775452**

54 Título: **Dispositivo de procesamiento de imágenes en movimiento, método de procesamiento de imágenes en movimiento y medio de grabación de información**

30 Prioridad:

**20.07.2012 JP 2012161924**

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

**24.07.2018**

73 Titular/es:

**RAKUTEN, INC. (100.0%)  
1-14-1, Tamagawa, Setagaya-ku  
Tokyo 158-0094, JP**

72 Inventor/es:

**ONAI RIKIO;  
OKABE MAKOTO;  
KURIHARA TATSUYA;  
SANJO MASAHIRO y  
HIRANO HIROMI**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 676 719 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 de procesamiento de imágenes en movimiento, método de procesamiento de imágenes en movimiento y medio de grabación de información.

Campo técnico

5 La presente invención se refiere a un dispositivo de procesamiento de vídeo, método de procesamiento de vídeo, y medio de grabación de información no transitorio para asegurar que los objetos no satisfagan condiciones de interferencia cuando los objetos acompañados de información de profundidad y representados en múltiples vídeos se recopilan en un vídeo.

Antecedentes de la técnica

10 La tecnología para crear imágenes acompañada de información de profundidad mediante la adquisición de la distancia (profundidad o distancia z) entre un objeto fotografiado y una cámara se conoce de antes. Dichas imágenes se llaman, con frecuencia, imágenes de dimensión 2.5 (por ejemplo, es preciso ver la Bibliografía de No Patente 1).

15 Por otro lado, se ha propuesto tecnología para separar regiones donde se representan fondos virtualmente estáticos y regiones donde se representan objetos en movimiento, de vídeos que representan estados en los cuales la posición, forma, color y/o similares de un objeto cambian (por ejemplo, es preciso ver la Bibliografía de Patente 1).

20 Mediante la combinación de dichas tecnologías, es posible generar vídeos de dimensión 2.5 en los cuales la profundidad de fondo y profundidad de objeto se distinguen, después de separar el fondo y los objetos que se mueven en la parte frontal de aquellos. En fotogramas que comprenden las imágenes de dimensión 2.5, se representa la superficie del objeto (superficie frontal del objeto) vista desde la cámara, y en píxeles incluidos en la región de representación donde el objeto se representa, el color de la parte de superficie del objeto correspondiente a dichos píxeles, junto con la profundidad hasta dicha parte de superficie (profundidad de superficie anterior), se encuentran asociadas.

25 Este tipo de vídeo de dimensión 2.5 puede obtenerse mediante el uso de detección por un sensor de profundidad de manera simultánea con la filmación del mundo real por una cámara CCD, o la filmación del mundo real con una cámara estéreo para estereopsis, y descubrir el paralaje para cada píxel correspondiente y calcular la profundidad desde dicho paralaje. Aquí, en un vídeo de dimensión 2.5 realizado mediante la filmación del mundo real, el estado de la superficie del objeto (superficie posterior del objeto) que no se representa en el fotograma no se fotografía, y la información sobre la profundidad hasta la superficie posterior se desconoce.

30 Además, después de crear un vídeo que expresa un estado que tiene en cuenta y provoca el movimiento y/o similares de un objeto en un espacio virtual, mediante la adición solamente de la profundidad de superficie anterior en cada píxel del vídeo, es posible obtener un vídeo de dimensión 2.5. Que el presente tipo de transformación se logre es conveniente por motivos como, por ejemplo, que se mantenga la coherencia del formato y el vídeo de dimensión 2.5 y que la edición del vídeo sea más fácil.

35 En el campo de los gráficos tridimensionales, la tecnología de determinación de colisión se usa ampliamente con el fin de evitar que regiones ocupadas por objetos se invadan mutuamente de modo que los objetos que se mueven en el espacio virtual no interfieran, y para evitar situaciones en las cuales una penetra en la otra (por ejemplo, es preciso ver la Bibliografía de Patente 2). La colisión, según se hace referencia a ella aquí, naturalmente incluye contacto de superficie entre objetos, y también incluye casos en los cuales las regiones mutuamente ocupadas se superponen y casos en los cuales la distancia entre dos objetos es más cercana que un valor umbral dado aunque los dos estén, en realidad, separados, y, por consiguiente, tiene un significado más amplio que la definición cotidiana.

40 La Bibliografía de No Patente 2 describe un método para detectar colisiones que pueden ocurrir en un solo vídeo 2D. El método descrito comprende un algoritmo de evitación de colisión para la simulación del comportamiento de las masas. Dichos métodos no incorporan o simulan información de profundidad y son diferentes de las simulaciones previstas para la detección de colisiones en vídeos tridimensionales.

45 La Bibliografía de No Patente 3 describe un método para evitar colisiones en un solo espacio 2.5D en donde el cómputo 3D se evita mediante la representación de objetos mediante el uso de un mapa de sombras 2.5D. El método descrito se refiere al cómputo de sombras para grandes entornos incluidos miles de objetos dinámicos. Las colisiones pueden evitarse mediante la deformación de la trayectoria (y, por lo tanto, de la sombra) de cada objeto.

Lista de citas

Bibliografía de Patente

Bibliografía de Patente 1: Publicación Kokai de Solicitud de patente japonesa pendiente de examen 2009-080660.

Bibliografía de Patente 2: Publicación de Solicitud de Patente de los Estados Unidos No. US2006/149546.

Bibliografía de No Patente

5 Bibliografía de No Patente 1: Artículo sobre "2.5D" en la versión en inglés del diccionario gratuito en línea Wikipedia, en <http://en.wikipedia.org/wiki/2.5D>, adquirido el 18 de julio de 2012.

Bibliografía de No Patente 2: Simulación del Comportamiento de Evitación de Colisión de Peatones. Artículo de investigación de Franck Fuetey, adquirido el 9 de febrero de 2015.

Bibliografía de No Patente 3: Sombras en Tiempo Real para masas animadas en ciudades virtuales. Artículo de S. Kockara y otros, adquirido el 10 de julio de 2015.

10 Compendio de la invención

Problema técnico

15 Aquí, cuando múltiples objetos representados en vídeos de dimensión 2.5 que difieren se recopilan en un solo vídeo, si múltiples objetos se representan superponiéndose en la misma región, qué objeto debe tener prioridad en la representación se determina según el tamaño de la profundidad. Sin embargo, comenzar con la interferencia de los objetos entre sí como, por ejemplo, objetos que se invaden mutuamente o penetran uno en el otro, no es natural en un vídeo final.

20 Además, con el fin de que el fotomontaje de vídeos parezca más natural, es deseable que los objetos mantengan una relación posicional preestablecida entre sí incluso cuando los objetos no interfieren mutuamente. Por ejemplo, formar un fotomontaje de un vídeo en el cual se representa un objeto que es un suelo y un vídeo en el cual se representa un objeto que es un bailarín bailando, es deseable imponer condiciones de posición de modo que el objeto que es el bailarín apenas se aproxime al objeto que es el suelo con el fin de que el vídeo parezca natural.

25 Por consiguiente, cuando se recopilan múltiples vídeos, se han buscado medidas para asegurar que los objetos no satisfagan condiciones de interferencia entre sí mediante la implementación de una traslación, rotación o escalamiento o mediante la implementación de una transformación no deformante con cambio de tiempo, de modo que las trayectorias a lo largo de las cuales los objetos representados en el vídeo se mueven en el espacio tridimensional no se deforman.

30 Según lo anterior, es un objeto de la presente descripción proveer un dispositivo de procesamiento de vídeo, método de procesamiento de vídeo, y medio de grabación de información no transitorio para asegurar que los objetos no satisfagan condiciones de interferencia cuando los objetos acompañados de información de profundidad y representados en múltiples vídeos se recopilan en un vídeo.

Solución al problema

35 El dispositivo de procesamiento de vídeo de la presente descripción es un dispositivo de procesamiento de vídeo para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto, y un segundo objeto se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto, el dispositivo de procesamiento de vídeo comprendiendo:

un generador de vídeo para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y el segundo vídeo 2.5D;

40 un determinador de interferencia para determinar si el primer objeto y el segundo objeto satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto y a la información de profundidad del segundo objeto; y

45 un corrector para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación de cambio de tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.

50 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que: la longitud de tiempo del cambio de tiempo, la cantidad de traslación, cantidad de rotación, o relación de escalamiento, o una combinación de ellas, se usa como un parámetro de transformación;

el dispositivo de procesamiento de vídeo además comprende un establecedor de valor inicial para establecer un valor inicial del parámetro de transformación; y

5 el corrector aplica la transformación no deformante según el parámetro de transformación, actualiza el parámetro de transformación de modo que la condición de interferencia no se satisface, y aplica la transformación no deformante por el parámetro de transformación actualizado.

Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que: la condición de interferencia se satisface si el primer objeto y el segundo objeto colisionan.

10 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que el corrector repite la actualización del parámetro de transformación hasta la primera vez en que se determina que el primer objeto y el segundo objeto no colisionan.

Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que el corrector repite la actualización del parámetro de transformación hasta la primera vez en que "se determina que el primer objeto y el segundo objeto colisionarán".

15 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que el corrector actualiza el parámetro de transformación dentro de un rango de corrección preestablecido de modo que el tiempo transcurrido hasta que la condición de interferencia se satisfaga por primera vez aumenta.

Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que:

20 el determinador de interferencia calcula una cantidad de colisión que indica el grado de colisión entre el objeto transformado y el primer objeto; y

el corrector actualiza el parámetro de transformación dentro de un rango preestablecido de modo que la cantidad de colisión disminuye.

25 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que la cantidad de colisión es una suma estimada de los volúmenes en los cuales una forma del primer objeto estimada según la información de profundidad del primer objeto en cada fotograma y una forma del objeto transformado estimada según la información de profundidad del objeto transformado determinada según la información de profundidad del segundo objeto se superponen.

30 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que la cantidad de colisión es la suma de las áreas en las cuales un área en la cual el primer objeto debe representarse en cada fotograma y un área en la cual el objeto transformado debe representarse se superponen.

35 Además, el dispositivo de procesamiento de vídeo según la presente descripción puede estar formado de modo que el generador de vídeo obtiene una posición tridimensional donde se posiciona un punto representativo del segundo objeto, desde una posición dentro de un fotograma donde un punto representativo se representa en el fotograma y una posición de profundidad de un punto representativo determinada a partir de la información de profundidad del segundo objeto, y obtiene una posición transformada mediante la aplicación de una transformación congruente del espacio tridimensional en la posición tridimensional, y obtiene una posición de dibujo donde la posición transformada debe dibujarse dentro del fotograma, y obtiene el vídeo 2.5D transformado mediante la implementación de una transformación de traslación según una diferencia entre la posición dentro del fotograma y la posición de dibujo obtenida, y mediante la aplicación de una transformación de escalamiento según la relación de la posición de profundidad y la profundidad de la posición transformada.

45 El método de procesamiento de vídeo según la presente descripción es un método de procesamiento de vídeo para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto, y un segundo objeto se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto, el método de procesamiento de vídeo comprendiendo:

un procedimiento de generación de vídeo para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y segundo vídeo 2.5D;

50 un procedimiento de determinación de interferencia para determinar si el primer objeto y el segundo objeto satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto y a la información de profundidad del segundo objeto; y

5 un procedimiento de corrección para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación con cambio de tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.

10 El medio de grabación de información no transitoria legible por ordenador no transitorio según la presente descripción es un medio de grabación de información no transitoria legible por ordenador no transitorio en el cual se graba un programa para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto, y un segundo objeto se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto, el programa haciendo que un ordenador funcione como:

un generador de vídeo para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y el segundo vídeo 2.5D;

15 un determinador de interferencia para determinar si el primer objeto y el segundo objeto satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto y a la información de profundidad del segundo objeto; y

20 un corrector para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación con cambio de tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.

25 El programa descrito más arriba puede grabarse en un medio de grabación de información no transitorio legible por ordenador como, por ejemplo, un disco compacto, disco flexible, disco duro, disco magneto-óptico, disco de vídeo digital, cinta magnética, memoria con semiconductores y/o similares. El medio de grabación de información no transitorio puede distribuirse y venderse de forma independiente de un ordenador.

Además, el programa descrito más arriba puede distribuirse y venderse mediante un medio de transmisión como, por ejemplo, una red de comunicación de ordenador y/o similares, independiente de un ordenador en el cual se ejecuta el programa.

### 30 Efectos ventajosos de la invención

Con la presente descripción, es posible proveer un dispositivo de procesamiento de vídeo, método de procesamiento de vídeo y un medio de grabación de información no transitorio para asegurar que los objetos no satisfagan condiciones de interferencia cuando los objetos acompañados de información de profundidad y representados en múltiples vídeos se recopilan en un vídeo.

### 35 Breve descripción de los dibujos

La Figura 1 es un dibujo explicativo que muestra la relación entre un objeto representado en un vídeo de dimensión 2.5 y los elementos de aquel;

la Figura 2A es una vista en sección transversal de un objeto;

la Figura 2B es una vista en sección transversal que muestra la superficie anterior del objeto;

40 la Figura 2C es una vista en sección transversal de un objeto, la forma de cuya superficie posterior se estima haciendo que el grosor sea constante;

la Figura 2D es una vista en sección transversal de un objeto, la forma de cuya superficie posterior se estima descubriendo una profundidad de superficie anterior representativa del objeto y estableciendo el valor descubierto mediante la adición de grosor a este como la profundidad de superficie posterior;

45 la Figura 2E es una vista en sección transversal de un objeto, cuya forma de superficie posterior se estima que es una superficie esférica;

la Figura 3 es un dibujo explicativo que muestra un fotomontaje a modo de resumen de elementos que acompañan a la determinación de colisión, en un dispositivo de procesamiento de vídeo;

la Figura 4 es un diagrama de flujo que muestra el flujo del proceso de determinación de colisión;

50 la Figura 5A es un dibujo explicativo que muestra el estado de la pantalla con anterioridad al inicio del arrastre;

- la Figura 5B es un dibujo explicativo que muestra el estado de la pantalla en el punto en el tiempo en el que se inicia el arrastre;
- la Figura 5C es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo durante el arrastre;
- 5 la Figura 5D es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo en el que la imagen se ha soltado;
- la Figura 6 es un dibujo explicativo que muestra el fotomontaje a modo de resumen de la parte según un proceso de arrastrar y soltar del dispositivo de procesamiento de vídeo;
- la Figura 7 es un diagrama de flujo que muestra el flujo de control del proceso de arrastrar y soltar;
- 10 la Figura 8 es un diagrama de flujo que muestra el flujo de control de un proceso de ajuste automático;
- la Figura 9A es un dibujo explicativo que muestra el estado de la pantalla con anterioridad al inicio del arrastre;
- la Figura 9B es un dibujo explicativo que muestra el estado de la pantalla en el punto en el tiempo en el que se inicia el arrastre;
- 15 la Figura 9C es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo durante el arrastre;
- la Figura 9D es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo en el que la imagen se ha soltado;
- la Figura 9E es un dibujo explicativo que muestra el estado en el cual una corrección automática de la posición de caída se ha llevado a cabo después de la acción de soltar;
- 20 la Figura 10 es un diagrama de flujo que muestra el flujo de control de un proceso de visualización destacado en la región donde puede soltarse;
- la Figura 11A es un dibujo explicativo que muestra el estado de una trayectoria  $\text{pos}[X,i](t)$  y una trayectoria  $\text{pos}[Y,j](t)$ ;
- la Figura 11B es un dibujo explicativo que muestra el estado de la trayectoria  $\text{pos}[X,i](t)$  y la trayectoria  $\text{pos}[W,j](t)$ ;
- 25 la Figura 12A es un dibujo explicativo que muestra el estado de una trayectoria que se deforma de manera gradual a través de la repetición de un proceso;
- la Figura 12B es un dibujo explicativo que muestra el estado de una trayectoria que se deforma de manera gradual a través de la repetición de un proceso;
- 30 la Figura 12C es un dibujo explicativo que muestra el estado de una trayectoria que se deforma de manera gradual a través de la repetición de un proceso;
- la Figura 12D es un dibujo explicativo que muestra el estado de una trayectoria que se deforma de manera gradual a través de la repetición de un proceso;
- la Figura 13 es un dibujo explicativo que muestra el estado de ajuste de un vector normal primario en una parte específica de una trayectoria con el fin de evitar una colisión.
- 35 Descripción de las realizaciones
- Más abajo, se describen las realizaciones preferidas de la presente descripción. Las realizaciones preferidas son a los fines explicativos y no limitan el alcance de la presente descripción. Por consiguiente, una persona con experiencia en la técnica puede aplicar una realización en la cual varios elementos o todos los elementos en la presente memoria se reemplazan por equivalentes a aquellos, pero dichas realizaciones también se incluyen dentro
- 40 del alcance de la presente descripción.
- El dispositivo de procesamiento de vídeo según las realizaciones preferidas descritas más abajo se lleva a cabo por un ordenador que ejecuta un programa. El programa se lee en un dispositivo de memoria temporal como, por ejemplo, una RAM (memoria de acceso aleatorio) y/o similares de un dispositivo de memoria no transitorio como, por ejemplo, una ROM (memoria de solo lectura), un disco duro y/o similares. Una CPU (Unidad de Procesamiento Central) interpreta, de forma sucesiva, programas leídos en la RAM, lee y escribe datos entre varios dispositivos de
- 45 memoria y logra procesos de cálculo en los datos. Además, la CPU controla la interacción con dispositivos de

entrada y salida como, por ejemplo, teclado, ratón, panel táctil, interfaz de visualización y red y/o similares, bajo el control del programa.

5 Sin embargo, el dispositivo de procesamiento de vídeo según la presente descripción no se encuentra limitado a su realización por un ordenador que ejecuta un programa. Por ejemplo, es posible llevar a cabo las mismas funciones mediante el uso de un circuito electrónico especial para el procesamiento de imágenes, una FPGA (Matriz de Puertas Programables por Campo), DSP (Procesador de Señal Digital) y/o similares.

10 En el campo de los gráficos tridimensionales, la interferencia entre objetos se llama una colisión. En el lenguaje cotidiano, una colisión significa objetos que colisionan de forma violenta, las superficies mutuas haciendo contacto y, en algunos casos, las formas de los objetos deformándose, pero cuando se habla de interferencia de objetos o colisión de objetos, el significado deberá incluir cuando las superficies hacen contacto, casos donde regiones dentro del espacio tridimensional ocupado por un objeto se invaden y penetran, y casos donde aunque los objetos que estén separados la distancia es lo suficientemente cercana y/o similares. En las realizaciones de más abajo, la frase "colisión" se usa en lugar del término "interferencia" para facilitar la comprensión.

(Relación entre un vídeo de dimensión 2.5 y objetos)

15 Más abajo, la notación para expresar un objeto representado en un vídeo de dimensión 2.5 se organiza para facilitar la comprensión. La Figura 1 es un dibujo explicativo que muestra la relación entre un objeto representado en un vídeo de dimensión 2.5 y los elementos de aquel. La explicación de más abajo hace referencia a dicho dibujo.

En la explicación de más abajo, la frase "vídeo de dimensión 2.5" se abrevia, de forma apropiada, como "vídeo" para facilitar la comprensión.

20 Un fotograma reproducido en un tiempo transcurrido  $t$  desde el punto de inicio de reproducción (clásicamente, esto se expresa por número de fotograma) entre fotogramas incluidos en un vídeo  $X$  dado se expresa como  $X(t)$ .

En el sistema de coordenadas de píxeles fijado para cada fotograma, en el fotograma  $X(t)$ , el píxel posicionado en el valor de coordenada  $x$  en la dirección horizontal y en el valor de coordenada  $y$  en la dirección vertical se expresa como  $X(t)[x,y]$ .

25 Que el valor de coordenada  $x$  en la dirección horizontal y el valor de coordenada  $y$  en la dirección vertical son valores eficaces en cada fotograma del vídeo  $X$  se expresa como  $[x,y] \in X$ . Es decir,  $[x,y]$  significa la posición de un píxel.

30 La Figura 1 ilustra un estado en el cual la condición de un objeto posicionado en el espacio virtual que se representa en un fotograma del vídeo se ve directamente hacia el lado, mediante el uso de tecnología de gráficos tridimensional, pero es posible concebir condiciones similares que existen incluso cuando el mundo real se fotografía con una cámara.

35 Como se muestra en el dibujo, cuando desde un punto de disparo 11 en el espacio una media línea 13 hacia el píxel  $X(t)[x,y]$  en el fotograma  $X(t)$  posicionado sobre una superficie de proyección 12 se encuentra con un punto de colisión 15 de la superficie de un objeto 14, el estado del punto de colisión 15 encontrado se representa en el píxel  $X(t)[x,y]$ . Esta es una proyección de perspectiva típica.

Según se describe más arriba, el presente dibujo es el estado de filmación visto directamente hacia el lado. Por consiguiente, en el presente dibujo, la superficie de proyección 12 se expresa por una línea recta.

40 El punto de disparo 11 corresponde a la posición de la cámara o posición de observación, y la distancia entre el punto de disparo 11 y la superficie de proyección 12 se determina por la longitud focal de las lentes y las unidades de longitud de píxel. Además, la dirección de disparo 16 es la dirección de una línea perpendicular desde el punto de disparo 11 hacia la superficie de proyección 12.

Más abajo, el color del píxel  $X(t)[x,y]$  y la profundidad de superficie anterior se expresan como  $X(t)[x,y].color$  y  $X(t)[x,y].fore$ .

45  $X(t)[x,y].color$  se determina según el color y estado de escritura en el punto de colisión 15, y la distancia entre el punto de disparo 11 y el punto de colisión 15, y/o similares.  $X(t)[x,y].color$  expresa un valor de coordenada en cada espacio de color como, por ejemplo, RGB, CYMK, HSV y/o similares, y la conversión del valor de coordenada entre espacios de color se reduce a una fórmula. Además, también estará bien añadir un valor  $\alpha$  que indica transparencia a  $X(t)[x,y].color$ .

50  $X(t)[x,y].fore$  expresa la distancia de profundidad entre la parte correspondiente al punto de colisión 15 del objeto 14 representado en el píxel  $X(t)[x,y]$ , y el punto de disparo 11. Como la distancia de profundidad, es típico utilizar una longitud 17 de un componente en la dirección de disparo 16 de un vector del punto de disparo 11 al punto de colisión

15 (correspondiente a la así llamada "distancia Z"), pero como una aproximación de aquella, estará bien utilizar la distancia del punto de disparo 11 al punto de colisión 15.

5 La distancia de profundidad puede expresarse en varios sistemas de unidad. Cuando un píxel comprende un cuadrado, si la longitud del lado del píxel se toma como la unidad, es posible normalizar la profundidad en múltiples vídeos. Más abajo, la explicación supone que la profundidad se ha normalizado.

En general, en el vídeo X un fondo se representa además de objetos en movimiento. El fondo puede pensarse como un objeto sustancialmente sin movimiento.

10 Si cada fotograma del vídeo se analiza mediante el uso de la tecnología descrita en la Bibliografía de Patente 1 y/o similares, es posible identificar si un fondo se representa o un objeto en movimiento se representa en el píxel  $X(t)[x,y]$ .

Más abajo, se considerará un caso en el cual números de identificación se han asignado a objetos identificados con el fin de facilitar la comprensión. El número de identificación del objeto que expresa el fondo es 0, y los números de identificación de otros objetos se asignan como 1, 2, 3, ....

15 Además, el número de identificación de un objeto, una parte del cual se representa en el píxel  $X(t)[x,y]$ , se expresa como  $X(t)[x,y].id$ , y el valor máximo de los números de identificación de los objetos que aparecen en el vídeo X se expresa como  $X.maxid$ .

Si los números de identificación de los objetos que aparecen en el vídeo X son 0, 1, 2, 3, ..., N, entonces  $X.maxid = N$ .

20 En el vídeo X, si solo los píxeles  $X(t)[x,y]$  que satisfacen la condición  $X(t)[x,y].id \neq 0$  se extraen, la información solamente sobre objetos (información que excluye el fondo) puede obtenerse. Además, con el fin de obtener un vídeo de solo el  $i^{\text{ésimo}}$  objeto, estará bien extraer solamente píxeles  $X(t)[x,y]$  que satisfacen la condición  $X(t)[x,y].id = i$ . Estos corresponden a vídeos para la composición de clave de croma mediante el uso de *blueback* en la tecnología de vídeo convencional.

25 Como resultado del logro de la extracción descrita más arriba, en los píxeles que han dejado de ser objetivos de procesamiento, se supondrá que un valor negativo se ha dado como un número de identificación, con el fin de facilitar la comprensión. Es decir, si  $X(t)[x,y].id < 0$ , el píxel  $X(t)[x,y]$  se interpretará como un píxel transparente. Aquí, "transparente" significa correspondiente a *blueback* en la composición de clave de croma.

30 Según la presente interpretación, es posible pensar un valor negativo establecido como  $X(t)[x,y].id$  cuando  $[x,y] \in X$  no se ha establecido. Más abajo, la explicación continuará para el caso en el cual valores negativos se establecen como  $.id$  para píxeles que no son objetivos de procesamiento y "píxeles" en posiciones fuera del fotograma del vídeo, con el fin de facilitar la comprensión.

(Varios tipos de conversiones para vídeos de dimensión 2.5)

Más abajo, se describen métodos de conversión para crear un vídeo diferente a partir de un vídeo dado. Entre dicho tipo de método de conversión se encuentran los siguientes:

- 35 (1) Traslación hacia arriba, abajo, izquierda o derecha a lo largo de un fotograma del vídeo;
- (2) Escalamiento del vídeo;
- (3) Traslación en la dirección de profundidad del vídeo;
- (4) Rotación del vídeo alrededor del eje horizontal, eje vertical o eje de profundidad;
- (5) Desplazamiento del vídeo en la dirección del tiempo;
- 40 (6) Selección de objetos específicos solamente del vídeo;
- (7) Formación de un fotomontaje mediante la superposición de un vídeo diferente sobre un vídeo dado.

Estos se explican en orden más abajo.

45 En primer lugar, es preciso considerar la traslación hacia arriba, abajo, izquierda o derecha a lo largo de un fotograma del vídeo. Si un vídeo Y se obtiene mediante la traslación del vídeo X por p en la dirección horizontal y q en la dirección vertical dentro del fotograma, entonces:

$$Y(t)[x+p,y+q] = X(t)[x,y], e$$

$$Y(t)[x,y] = X(t)[x-p, y-q].$$

Aquí, para las igualdades descritas más arriba, los valores asignados a cada píxel como, por ejemplo, .color y .fore y/o similares, tienen, todos, un significado equivalente.

5 Más abajo, el vídeo obtenido mediante la traslación del vídeo X por p en la dirección horizontal y por q en la dirección vertical dentro del fotograma se expresa como  $\text{move}(p,q,X)$ .

A continuación, es preciso considerar el escalamiento del vídeo. En una proyección transparente, representar tantas c veces de grande significa que la profundidad se ha convertido en 1/c veces y los puntos de fondo se han acercado. Por consiguiente, si el vídeo Y es la ampliación del vídeo X c veces dentro del fotograma, para otro diferente de .fore asignado a cada píxel,

$$10 \quad Y(t)[c \times x, c \times y] = X(t)[x,y],$$

es decir,

$$Y(t)[x,y] = X(t)[x/c, y/c].$$

Además, para .fore asignado a cada píxel,

$$Y(t)[c \times x, c \times y].\text{fore} = X(t)[x,y].\text{fore}/c,$$

15 es decir,

$$Y(t)[x,y].\text{fore} = X(t)[x/c, y/c].\text{fore}/c.$$

Más abajo, el vídeo obtenido mediante la ampliación del vídeo X c veces dentro del fotograma se expresa como  $\text{scale}(c,X)$ .

20 A continuación, es preciso considerar la traslación del vídeo en la dirección de profundidad. Cuando el vídeo Y se obtiene por objetos filmados en el vídeo X haciendo que estén más distantes por la dirección de profundidad r, la distancia de profundidad de  $X(t)[x,y]$  cambia de  $X(t)[x,y].\text{fore}$  a  $X(t)[x,y].\text{fore} + r$ . Es decir, la distancia de profundidad se convierte en  $(X(t)[x,y].\text{fore} + r)/X(t)[x,y].\text{fore} = k$  veces.

25 En una proyección transparente, si la distancia a un punto de disparo se convierte en k veces, el tamaño representado en la superficie de proyección se amplía o reduce a 1/k veces. Por consiguiente, para otro diferente de .fore asignado a cada píxel,

$$Y(t)[x/k, y/k] = X(t)[x,y],$$

y para .fore,

$$Y(t)[x/k, y/k].\text{fore} = X(t)[x,y].\text{fore} + r.$$

30 Más abajo, el vídeo obtenido mediante la traslación del vídeo X por r en la dirección de profundidad se expresa como  $\text{push}(r,X)$ .

Además, es preciso considerar la rotación del vídeo alrededor del eje horizontal, eje vertical o eje de profundidad. Si el vídeo Y se obtiene mediante la rotación del vídeo X por  $\theta$  alrededor del eje horizontal, los elementos diferentes de .fore se convierten en:

$$Y(t)[x,y] = X(t)[x, X(t)[x,y].\text{fore} \times \cos \theta - y \times \sin \theta],$$

35 y para .fore,

$$Y(t)[x,y].\text{fore} = X(t)[x,y].\text{fore} \times \sin \theta + y \times \cos \theta$$

se establece.

De manera similar, si el vídeo Y se obtiene mediante la rotación del vídeo X por  $\varphi$  alrededor del eje vertical, los elementos diferentes de .fore se convierten en:

$$Y(t)[x,y] = X(t)[X(t)[x,y].fore \times \cos \phi - x \times \sin \phi, y],$$

y para .fore,

$$Y(t)[x,y].fore = X(t)[x,y].fore \times \sin \phi + x \times \cos \phi$$

se establece.

- 5 Además, cuando el vídeo Y se obtiene mediante la rotación del vídeo X por  $\psi$  alrededor del eje de profundidad,

$$Y(t)[x,y] = X(t)[x \times \cos \psi - y \times \sin \psi, x \times \sin \psi + y \times \cos \psi]$$

se establece.

Más abajo, los vídeos obtenidos por las respectivas rotaciones se expresan como  $rothor(\theta, X)$ ,  $rotver(\phi, X)$  y  $rotdep(\psi, X)$ .

- 10 Cuando dichas transformaciones se logran, los valores de coordenadas de los píxeles no son valores enteros y, en ocasiones, un agujero ocurre en los valores de coordenadas. En dichos casos, los valores de .color y .fore y/o similares asignados a cada píxel se interpolan y se logra el antialiasing, y para .id, es posible adoptar un método como, por ejemplo, utilizar un valor en el punto reticular más cercano a dicha coordenada, o encontrar en qué posición el límite donde .id cambia se posiciona a través de la interpolación y determinar .id según dicho límite.

- 15 Los píxeles que se encuentran fuera del alcance del fotograma del vídeo debido a una traslación y/o similares pueden pensarse como píxeles transparentes, entonces estará bien establecer .id en un valor negativo como se describe más arriba.

A continuación, es preciso considerar las traslaciones del vídeo en la dirección del tiempo. Si el vídeo Y se descubre mediante el desplazamiento del vídeo X por d en el tiempo,

20  $Y(t)[x,y] = X(t-d)[x,y]$

se establece.

Más abajo, el vídeo en el cual el tiempo del vídeo X se desplaza por d se expresa como  $shift(d, X)$ .

Además, es preciso considerar un proceso para seleccionar solamente un objeto específico de un vídeo. Si un vídeo y representa solo objetos con el  $i^{\text{ésimo}}$  número de identificación en el vídeo X, para elementos diferentes de .id,

25  $Y(t)[x,y] = X(t)[x,y], \text{ si } X(t)[x,y].id = i$

se establece.

Además, para .id, si

$$Y(t)[x,y].id = 1, \text{ si } X(t)[x,y].id = i \text{ e } i > 0;$$

$$Y(t)[x,y].id = 0, \text{ si } X(t)[x,y].id = i \text{ e } i = 0;$$

- 30  $Y(t)[x,y].id = -1$ , de lo contrario,

entonces es posible minimizar en la medida de lo posible el valor máximo  $Y.maxid$  de los números de identificación en el vídeo Y.

Más abajo, un vídeo que representa solo el objeto que tiene el número de identificación i representado en el vídeo X se expresa como  $select(i, X)$ .

Finalmente, es preciso considerar el proceso de superponer un vídeo diferente a un vídeo dado. Si no hay colisión entre objetos representados en los vídeos X e Y, mediante la superposición de un objeto diferente del fondo del vídeo Y sobre el vídeo X (estará bien incluir el fondo, y estará bien también incluir solo objetos diferentes del fondo), un vídeo Z se obtiene en el cual los objetos representados en el vídeo X y los objetos representados en el vídeo Y se recopilan en uno.

5

En el vídeo Z, para otro diferente de .id,

$$Z(t)[x,y] = X(t)[x,y], \text{ si } Y(t)[x,y].id \leq 0;$$

$$Z(t)[x,y] = X(t)[x,y], \text{ si } X(t)[x,y].fore < Y(t)[x,y].fore;$$

$$Z(t)[x,y] = Y(t)[x,y], \text{ de lo contrario}$$

10 se establece.

Además, para .id, si

$$Z(t)[x,y].id = X(t)[x,y].id, \text{ si } Y(t)[x,y].id \leq 0;$$

$$Z(t)[x,y].id = X(t)[x,y].id, \text{ si } X(t)[x,y].fore < Y(t)[x,y].fore;$$

$$Z(t)[x,y].id = Y(t)[x,y].id + X.maxid, \text{ de lo contrario}$$

15 entonces es posible evitar números de identificación de objetos que se superponen.

Además, de los enteros 1, 2, ..., Z.maxid, si hay uno que no aparece como un valor de Z(t)[x,y].id, es posible minimizar Z.maxid mediante el desplazamiento de .id de los objetos subsiguientes.

Más abajo, el vídeo obtenido mediante la superposición de objetos diferentes del fondo del vídeo Y sobre el vídeo X se expresa como *superimpose(Y,X)*.

20 Cuando los procesos descritos más arriba *move*, *scale*, *push*, *rothor*, *rotver*, *rotdep*, *select*, *superimpose* y/o similares se logran mediante ordenador, es posible usar una función de calculadora que posee una CPU (Unidad de Procesamiento Central), y el cálculo de alta velocidad también es posible mediante el uso de un procesador para el procesamiento de gráficos, y/o similares.

25 En *scale* y *push*, el escalamiento y la traslación en la dirección de profundidad se consideran centrados en el origen, y en *rothor*, *rotver* y *rotdep*, se supone que el eje de rotación atraviesa el origen.

Por consiguiente, cuando se desea lograr el escalamiento o la traslación en la dirección de profundidad centrada en una posición arbitraria [p,q] dentro del vídeo X, estará bien primero trasladar [p,q] hacia arriba, abajo, izquierda o derecha hacia el origen, entonces lograr el escalamiento o la traslación en la dirección de profundidad y luego lograr la traslación inversa hacia arriba, abajo, izquierda o derecha. Es decir, estará bien usar:

30 `move(p,q, scale(c, move(-p, -q, X)));`

`move(p,q, push(r, move(-p, -q, X))).`

35 Cuando la posición del eje de rotación se desplaza hacia arriba, abajo, izquierda o derecha o en la dirección de profundidad, el eje de rotación se traslada primero en la dirección hacia arriba, abajo, izquierda o derecha o en la dirección de profundidad para atravesar el origen, la rotación se logra, y luego una traslación inversa se logra en la dirección hacia arriba, abajo, izquierda o derecha o en la dirección de profundidad.

Además, en una transformación como, por ejemplo, *scale* o *push* y/o similares cuando la profundidad se traslada, estará bien corregir el brillo o la intensidad y/o similares de .color según el grado de aquellos.

La normalización del vídeo también puede lograrse según las configuraciones y/o similares al momento de filmación, pero también es posible que ello se configure según los deseos del usuario o que se logre de forma automática.

5 En el caso de configuraciones por el usuario, primero la altura o ancho del vídeo se escalan simplemente según las instrucciones del usuario y/o similares, y mediante el ajuste de la resolución del tamaño de objetos representados en la imagen se convierten según se desee. Con "scale", la distancia de profundidad de los píxeles correspondientes se multiplica por un coeficiente correspondiente a la relación de ampliación, pero cuando se normaliza, la distancia de profundidad de los píxeles correspondientes se usa sin cambios.

Luego, el usuario selecciona un coeficiente y la distancia de profundidad de cada píxel se multiplica por dicho coeficiente.

10 Cuando se lleva a cabo el intento de generar un vídeo Z en el cual el objeto i y el objeto j se representan juntos en la situación en la cual un objeto i se representa en el vídeo X y un objeto j se representa en el vídeo Y es, con frecuencia, preferible que el tamaño del objeto i y el tamaño del objeto j sean sustancialmente equivalentes, y que la profundidad del objeto i y la profundidad del objeto j sean sustancialmente equivalentes. Cuando se establecen dichos prerrequisitos, la normalización automática es posible.

15 Es decir, el vídeo Y se escala simplemente de modo que el área promedio y la altura promedio y/o similares de la región en la cual el objeto j se representa en el vídeo Y concuerdan con los valores correspondientes del objeto i en el vídeo X, de modo que se obtiene el vídeo Z. Es decir, para todos los valores que incluyen .fore,

$$Z(t)[x,y] = Y(t)[x/c, y/c].$$

Además, la profundidad en el vídeo Z se multiplica por un coeficiente de modo que la profundidad promedio del objeto j en el vídeo Z y la profundidad promedio del objeto i en el vídeo X concuerdan y, de esta manera, se obtiene un vídeo W. Es decir, para otro diferente de .fore,

20  $W(t)[x,y] = Z(t)[x,y],$

y para .fore,

$$W(t)[x,y].fore = k \times Z(t)[x,y].fore.$$

25 Si c y k se establecen de forma automática de modo que las condiciones de más arriba se satisfacen, el vídeo W se obtiene mediante la normalización del vídeo Y de modo que el objeto j corresponde al objeto i. Naturalmente, según se describe más arriba, estará bien lograr la normalización mediante la configuración de c y k según los deseos del usuario.

30 Más arriba, se ha descrito la notación de varias cantidades características de los vídeos y varios tipos de transformaciones, pero dichas notaciones y transformaciones pretenden facilitar la comprensión. En particular, es posible expresar .color, .fore, .id y/o similares a través de otros formatos. Incluso cuando métodos o transformaciones equivalentes a estos se utilizan en varios formatos usados en el procesamiento de vídeo, naturalmente ello se incluye dentro del alcance técnico de la presente descripción.

Primera realización preferida

35 La presente realización preferida se usa para determinar si un objeto i que tiene un número de identificación i en el vídeo X y un objeto j que tiene un número de identificación j en el vídeo Y colisionan o no cuando el vídeo Y se superpone al vídeo X.

Es preciso considerar las siguientes condiciones como un caso en el cual una colisión entre el objeto i y el objeto j puede determinarse de forma relativamente fácil.

40 Primero, existe un caso en el cual hay un instante durante el cual la superficie anterior del objeto i (la superficie representada en el vídeo X) y la superficie anterior del objeto j (la superficie representada en el vídeo Y) ocupan la misma posición tridimensional. Es decir, para cualquier t, x e y, si

$$X(t)[x,y].id = i;$$

$$e Y(t)[x,y].id = j;$$

la posición [x,y] se contiene duplicada en el área donde el objeto i debe representarse y el área donde el objeto j debe representarse.

45 En la presente posición [x,y], si

$$X(t)[x,y].fore = Y(t)[x,y].fore$$

se establece, se determina que el objeto i y el objeto j colisionan.

Con el presente método, se supone que los objetos i y j son objetos tipo hoja con nada sobre la superficie posterior, sin poseer profundidad y comprendiendo solamente la forma de la superficie anterior.

5 Para que las superficies anteriores de los objetos i y j tengan la misma profundidad en una posición proyectada en un píxel dado  $X(t)[x,y]$  significa que las superficies anteriores de los objetos i y j se superponen en dicha posición. Por lo tanto, en el presente caso es posible determinar que los dos colisionan.

También es posible extender el presente estándar de determinación en la dirección del tiempo. Más abajo, la suposición es para un caso en el cual el tiempo transcurrido t se expresa por el número de fotograma, es decir que el fotograma en el tiempo t+1 sigue al fotograma en el tiempo t.

En cualquier t, x e y, si

10  $X(t)[x,y].id = X(t+1)[x,y].id = i;$

$Y(t)[x,y].id = Y(t+1)[x,y].id = j;$

$X(t)[x,y].fore < Y(t)[x,y].fore; y$

15  $X(t+1)[x,y].fore > Y(t+1)[x,y].fore$

se establecen, puede pensarse que del tiempo t al tiempo t+1, el objeto j ha llegado de la parte posterior del objeto i a la parte frontal. Además, si

$X(t)[x,y].id = X(t+1)[x,y].id = i;$

20  $Y(t)[x,y].id = Y(t+1)[x,y].id = j;$

$X(t)[x,y].fore > Y(t)[x,y].fore; y$

$X(t+1)[x,y].fore < Y(t+1)[x,y].fore$

25 se establecen, puede pensarse que del tiempo t al tiempo t+1, el objeto i ha llegado de la parte posterior del objeto j a la parte frontal.

Por lo tanto, cuando el presente tipo de "deslizamiento" ocurre, se determina que ha ocurrido una colisión de objetos.

30 Con una determinación de colisión para objetos i y j en dichas condiciones, se supone que los objetos i y j están compuestos de las formas de solamente las superficies anteriores de aquellos. Cuando los objetos son objetos con forma de placa delgada, puede pensarse que el grosor de los objetos es cero y que la superficie anterior del objeto y la superficie posterior del objeto coinciden. Por consiguiente, mediante la determinación de la colisión lograda a través del presente tipo de método, una determinación fácil y rápida es posible.

La Figura 2A es una vista en sección transversal del objeto 14 que se muestra en la Figura 1 y la Figura 2B es una vista en sección transversal que muestra la superficie anterior del objeto 14. La explicación de más abajo hace referencia a dichos dibujos.

35 Como se muestra en la Figura 2A, el objeto 14 tiene profundidad y la sección transversal de aquel tiene una extensión.

40 Sin embargo, cuando el objeto 14 se expresa a través de la información contenida en un vídeo de dimensión 2.5, como se muestra en la Figura 2B, la forma de aquel es un objeto planar que comprende solo la superficie anterior del objeto 14. Dado que el presente dibujo es una vista en sección transversal, un objeto planar se expresa por una línea curva.

Por consiguiente, con el fin de lograr una determinación de colisión entre objetos de forma más exacta, es necesario establecer o calcular la forma de la superficie posterior del objeto 14 a partir de la información sobre el objeto planar como se muestra en la Figura 2B.

45 Aquí, la profundidad hasta la superficie anterior del objeto 14 se obtiene de la información inherente al vídeo de dimensión 2.5, entonces con el fin de aprender la profundidad del objeto 14, estará bien obtener la profundidad hasta la superficie posterior del objeto 14.

Más abajo, la profundidad hasta la superficie posterior de un objeto representado en el píxel  $X(t)[x,y]$  en el tiempo  $t$  en el vídeo  $X$  se expresa como  $X(t)[x,y].back$ , y se describe un método para obtener dicha profundidad de superficie posterior.

5 En un primer método, el usuario establece con antelación la distancia entre la superficie anterior y la superficie posterior para cada objeto.

Es decir, si la distancia (grosor) entre la superficie anterior y la superficie posterior se establece como una constante  $thick(i)$  para un objeto  $i$  en el vídeo  $X$ , entonces el resultado es:

$$X(t)[x,y].back = X(t)[x,y].fore + thick(X(t)[x,y].id).$$

10 La Figura 2C es una vista en sección transversal de un objeto para el cual la forma la superficie posterior se estima haciendo que el grosor sea una constante. Como se muestra en el dibujo, la superficie posterior del objeto 14 es la forma de la superficie anterior del objeto 14 que ha experimentado una traslación en la dirección de disparo 16.

Además, existe un método para determinar una profundidad de superficie anterior representativa para el objeto  $i$  y para tomar como la profundidad de superficie posterior un valor encontrado mediante la adición a ello del valor de  $thick(i)$ .

15 Lo siguiente puede concebirse como una profundidad de superficie anterior representativa  $repfore(X,t,i)$  del objeto  $i$  en un tiempo dato  $t$ :

(1) Utilizar el valor máximo de la profundidad de superficie anterior del objeto  $i$ :

$$repfore(X,t,i) - \max_{x,y;X(t)[x,y].id=i} X(t)[x,y].fore$$

(2) Utilizar el valor promedio de la profundidad de superficie anterior del objeto  $i$ :

20  $repfore(X,t,i) - avg_{x,y;X(t)[x,y].id=i} X(t)[x,y].fore$

Aquí,  $max$  y  $avg$  significan el valor máximo y el valor promedio de los valores de la parte principal cuando, en la parte de subíndice, se hace que las variables dispuestas en frente del punto y coma cambien dentro del rango que satisface las condiciones colocadas después del punto y coma. Cuando "que se satisfacen constantemente" se utiliza como la condición, solo las variables se expresan como el subíndice.

25 Es decir, ello significa enumerar combinaciones de  $x,y$  que satisfacen  $X(t)[x,y].id = i$  bajo un  $X$ ,  $t$  e  $i$  dados en (1) y (2) descritos más arriba, encontrar el valor de  $X(t)[x,y].fore$  para cada una de dichas combinaciones y encontrar el valor máximo o valor promedio de aquellas.

Si  $repfore(X,t,i)$  por consiguiente establecido se usa, la profundidad de superficie posterior del píxel  $X(t)[x,y]$  cuando el objeto  $i$  se representa en el píxel  $X(t)[x,y]$  ( $i = X(t)[x,y].id$ ) puede establecerse de modo que:

30  $X(t)[x,y].back = repfore(X,t,i) + thick(i).$

o,

$$X(t)[x,y].back = \max[repfore(X,t,i) + thick(i), X(t)[x,y].fore].$$

Aquí,  $\max[...]$  significa el valor máximo de los valores dispuestos dentro de los corchetes.

35 La Figura 2D es una vista en sección transversal de un objeto para el cual la forma de la superficie posterior del objeto se estima descubriendo una profundidad de superficie anterior representativa del objeto y añadiendo grosor a aquella para descubrir la profundidad de superficie posterior. Como se muestra en el dibujo, la superficie posterior del objeto 14 es un plano perpendicular a la dirección de disparo 16, y el objeto 14 parece una forma de columna que se extiende en la dirección de disparo 16.

40 Además, existe un método de aproximación a la superficie posterior mediante el establecimiento de un punto representativo del objeto  $i$  en el tiempo  $t$ , y la creación de una superficie esférica centrada en dicho punto representativo.

Primero, el número de píxel  $area(X,t,i)$  de la región donde el objeto  $i$  se representa en el tiempo  $t$  en el vídeo  $X$  se descubre de la siguiente manera:

$$area(X,t,i) = \sum_{x,y;X(t)[x,y].id=i} 1$$

Aquí, el subíndice de  $\Sigma$  tiene el mismo significado que con *max* y *avg*.

Una coordenada horizontal  $x_c(X,t,i)$  y una coordenada vertical  $y_c(X,t,i)$  del punto representativo del objeto *i* en el tiempo *t* se establecen de la siguiente manera:

$$x_c(X,t,i) = \Sigma_{x,y;X(t)[x,y].id=i} x / \text{area}(X,t,i);$$

5  $y_c(X,t,i) = \Sigma_{x,y;X(t)[x,y].id=i} y / \text{area}(X,t,i).$

Además, un ancho  $w(X,t,i)$  y una altura  $h(X,t,i)$  de la región en la cual el objeto *i* se representa en el tiempo *t* en el vídeo *X* se establecen de la siguiente manera:

$$w(X,t,i) = \max_{x,y;X(t)[x,y].id=i} x - \min_{x,y;X(t)[x,y].id=i} x;$$

$$h(X,t,i) = \max_{x,y;X(t)[x,y].id=i} y - \min_{x,y;X(t)[x,y].id=i} y.$$

10 Existen varios métodos para establecer el diámetro  $D(X,t,i)$  de la superficie esférica como, por ejemplo, aquellos descritos más abajo:

$$D(X,t,i) = \max [w(X,t,i), h(X,t,i)];$$

15  $D(X,t,i) = (w(X,t,i)^2 + h(X,t,i)^2)^{1/2};$

$$D(X,t,i) = \text{area}(X,t,i)^{1/2};$$

$$D(X,t,i) = \max_{x,y;X(t)[x,y].id=i} ((x-x_c(X,t,i))^2 + (y-y_c(X,t,i))^2)^{1/2};$$

20  $D(X,t,i) = \text{avg}_i \text{area}(X,t,i)^{3/2} / \text{area}(X,t,i)$

Aquí,  $\text{avg}_i \text{area}(X,t,i)$  significa el promedio de tiempo para la región en la cual el objeto *i* se representa en el vídeo *X*. Por consiguiente,  $\text{avg}_i \text{area}(X,t,i)^{3/2}$  corresponde al valor aproximado del volumen que el objeto *i* ocupa, y cuando este se divide por  $\text{area}(X,t,i)$ , se obtiene un valor aproximado de la longitud de profundidad.

25 Ahora, una vez que se establece el diámetro  $D(X,t,i)$ , la coordenada de profundidad  $z_c(X,t,i)$  del punto representativo puede establecerse de la siguiente manera:

$$z_c(X,t,i) = X(t)[x_c(X,t,i), y_c(X,t,i)].\text{fore} + D(X,t,i)/2$$

Además, estará bien establecer simplemente esto de la siguiente manera:

$$z_c(X,t,i) = X(t)[x_c(X,t,i), y_c(X,t,i)].\text{fore}.$$

30 Por lo tanto, si la profundidad de superficie posterior del píxel  $X(t)[x,y]$  cuando el objeto *i* se representa en el píxel  $X(t)[x,y]$  ( $i = X(t)[x,y].id$ ) se establece por:

$$X(t)[x,y].\text{back} = \max [z_c(X,t,i) + [\max [(D(X,t,i)^2) - (x-x_c(X,t,i))^2 - (y-y_c(X,t,i))^2], 0]^{1/2}, X(t)[x,y].\text{fore}],$$

es posible aproximarse mejor a la superficie esférica, y si

$$X(t)[x,y].\text{back} = \max [z_c(X,t,i) + D, X(t)[x,y].\text{fore}]$$

se establece, es posible aproximarse mejor a un cilindro.

35 La Figura 2E es un dibujo en sección transversal de un objeto, cuya forma de superficie posterior se supone que es una superficie esférica. En el ejemplo que se muestra en el presente dibujo, la superficie posterior del objeto 14 es una superficie esférica centrada en un punto representativo 21, pero de la superficie anterior del objeto 14 a la superficie esférica la forma se parece más a la forma de un cilindro que se extiende en la dirección de disparo 16.

40 Hasta ahora, se han descrito métodos para establecer y calcular la profundidad de superficie posterior de un objeto en movimiento, pero estará bien pensar la profundidad de superficie posterior como  $\infty$  para objetos correspondientes al fondo. Es decir, cuando

$X(t)[x,y].id = 0$

se establece, entonces

$X(t)[x,y].back = \infty$ .

5 La zona de ocupación en la dirección de profundidad ocupada en el píxel  $X(t)[x,y]$  por el objeto  $X(t)[x,y].id$  representado en el píxel  $X(t)[x,y]$  en el vídeo X en el tiempo t se extiende de la profundidad de superficie anterior  $X(t)[x,y].fore$  a la profundidad de superficie posterior  $X(t)[x,y].back$ .

Si existe una superposición en la zona de ocupación de la profundidad de superficie anterior a la profundidad de superficie posterior de dos objetos representados en la misma posición de píxel, dichos dos objetos colisionan.

Por consiguiente, para cualquier t, x, y:

10  $X(t)[x,y].id = i;$

$Y(t)[x,y].id = j,$

y si cualquiera de las cuatro relaciones:

15 (1)  $X(t)[x,y].fore \leq Y(t)[x,y].fore \leq X(t)[x,y].back;$

(2)  $X(t)[x,y].fore \leq Y(t)[x,y].back \leq X(t)[x,y].back;$

(3)  $Y(t)[x,y].fore \leq X(t)[x,y].fore \leq Y(t)[x,y].back;$

20 (4)  $Y(t)[x,y].fore \leq X(t)[x,y].back \leq Y(t)[x,y].back.$

se establece, cuando el vídeo X y el vídeo Y se superponen, es posible determinar que el objeto i representado en el vídeo X y el objeto j representado en el vídeo Y colisionan.

25 Más abajo, la cantidad de superposición en las dos zonas de ocupación se expresa como  $overlap(X,Y,t,x,y)$ . Si ninguno de (1) a (4) de más arriba se satisface,  $overlap(X,Y,t,x,y) = 0$ , y si  $overlap(X,Y,t,x,y) > 0$ , los objetos i y j colisionan.

De manera específica, en el caso de (1) de más arriba,

$overlap(X,Y,t,x,y) = \min [X(t)[x,y].back, Y(t)[x,y].back] - Y(t)[x,y].fore,$

en el caso de (2) de más arriba

$overlap(X,Y,t,x,y) = Y(t)[x,y].back - \max [X(t)[x,y].fore, Y(t)[x,y].fore],$

30 en el caso de (3) de más arriba,

$overlap(X,Y,t,x,y) = \min [X(t)[x,y].back, Y(t)[x,y].back] - X(t)[x,y].fore,$

y en el caso de (4) más arriba,

$overlap(X,Y,t,x,y) = X(t)[x,y].back - \max [X(t)[x,y].fore, Y(t)[x,y].fore].$

Aquí,  $\min [...]$  es el opuesto de  $\max [...]$  y devuelve el valor mínimo de los valores en los corchetes.

35 Cuando cualquiera de (1) a (4) de más arriba se establece y  $overlap(X,Y,t,x,y) = 0$ , es posible pensar los objetos i y j como objetos no separados pero mutuamente en contacto. Por consiguiente, es posible distinguir entre una colisión en la cual los dos mutuamente se invaden entre sí y una colisión en la cual las superficies se tocan.

40 Además, estará bien en la determinación de colisión no usar la propia zona de ocupación descrita más arriba pero descubrir si existe o no una superposición mediante una ligera ampliación de la zona de ocupación mediante la adición de un margen al límite superior y límite inferior de aquella. En el presente caso, haciendo que un margen se añada a la determinación de colisión de los objetos i y j, se determina que los objetos i y j colisionan (o están en peligro de colisionar) si no se separan por cierta cantidad de distancia.

Por otro lado, si el margen descrito más arriba es negativo, es posible permitir colisiones con una ligera invasión. Ello es apropiado cuando el objeto es elástico y se suponen indentaciones, y/o similares.

La Figura 3 es un dibujo explicativo que muestra un fotomontaje de elementos a modo de resumen para lograr una determinación de colisión, en el dispositivo de procesamiento de vídeo según la presente realización preferida.

5 Como se muestra en el presente dibujo, los elementos para lograr una determinación de colisión dentro del dispositivo de procesamiento de vídeo 101 comprenden un adquirente de profundidad de superficie posterior 102 y un determinador de colisión 103.

El dispositivo de procesamiento de vídeo 101 toma el primer vídeo X y el segundo vídeo Y como objetivos de procesamiento. Por ejemplo,

(a) como el primer vídeo X, se utiliza una imagen de dimensión 2.5 que filma a un bailarín bailando en el mundo real,

(b1) como el segundo vídeo Y, se utiliza un vídeo que expresa a un personaje bailando sin fondo, y

10 (b2) como el segundo vídeo Y, se utiliza una filmación de vídeo para expresar a otro usuario bailando y luego a otro usuario, que ha logrado la edición y ha retirado toda la información de fondo de modo que el otro usuario está bailando sin fondo.

El dispositivo de procesamiento de vídeo 101 en la presente realización preferida compone, finalmente, un tercer vídeo en el cual un bailarín está bailando junto con un personaje u otro usuario.

15 Como resultado, primero el dispositivo de procesamiento de vídeo 101 determina una colisión entre el primer objeto i representado en el primer vídeo X y el segundo objeto j representado en el segundo vídeo Y que debe superponerse al primer vídeo X. El dispositivo de procesamiento de vídeo 101 se lleva a cabo normalmente mediante la ejecución de un programa en un ordenador.

20 Aquí, el primer vídeo X se acompaña de una primera profundidad de superficie anterior del primer objeto i en el lado representado en el primer vídeo X, y el segundo vídeo Y se acompaña de una segunda profundidad de superficie anterior del segundo objeto j en el lado representado por el segundo vídeo Y.

Mediante el uso de la notación de más arriba, si el primer objeto i se representa en un píxel en la posición  $[x,y]$  en el fotograma  $X(t)$  en el tiempo t contenido en el primer vídeo X, entonces  $X(t)[x,y].id = i$  y la primera profundidad de superficie anterior es  $X(t)[x,y].fore$ .

25 El primer vídeo X puede comprender un objeto en movimiento solo o puede incluir fondo. En el primer caso, un objeto con  $i > 0$  se utiliza como el primer objeto, y en el último caso, un objeto con  $i \geq 0$  se utiliza como el primer objeto.

30 Además, si el segundo objeto j se representa en un píxel en la posición  $[x,y]$  en un fotograma  $Y(t)$  en el tiempo t contenido en el segundo vídeo Y, entonces  $Y(t)[x,y].id = j$  y la segunda profundidad de superficie anterior es  $Y(t)[x,y].fore$ .

Los objetos que son el objetivo de la determinación de colisión en el segundo vídeo Y no incluyen el fondo. Es decir, como el segundo objeto, se utilizan objetos con  $j > 0$ .

35 En primer lugar, aquí, el adquirente de profundidad de superficie posterior 102 adquiere la primera profundidad de superficie posterior del primer objeto i en el lado no representado en el primer vídeo X, y la segunda profundidad de superficie posterior del segundo objeto j en el lado no representado en el segundo vídeo Y.

40 Si se usa la notación descrita más arriba, el adquirente de superficie posterior 102 descubre la primera profundidad de superficie posterior  $X(t)[x,y].back$  a través del método descrito más arriba para el tiempo t y la posición  $[x,y]$  que satisface  $X(t)[x,y].id = i$ . Además, el adquirente de superficie posterior 102 descubre la segunda profundidad de superficie posterior  $Y(t)[x,y].back$  a través del método descrito más arriba para el tiempo t y la posición  $[x,y]$  que satisface  $Y(t)[x,y].id = j$ .

45 Por otro lado, el determinador de colisión 103 funciona como un determinador de interferencia para determinar la interferencia entre objetos. Además, el determinador de colisión 103 determina que el primer objeto y el segundo objeto colisionan en la primera área de dibujo donde el primer objeto i debe representarse y la segunda área de dibujo donde el segundo objeto j debe representarse se superponen como resultado de la superposición del primer vídeo X y el segundo vídeo Y, y la primera zona de ocupación de la primera profundidad de superficie anterior a la primera profundidad de superficie posterior del primer objeto i y la segunda zona de ocupación de la segunda profundidad de superficie anterior a la segunda profundidad de superficie posterior del segundo objeto j se superponen en un área de superposición donde la primera área de dibujo y la segunda área de dibujo se superponen.

50 Si se usa la notación descrita más arriba, la primera área de dibujo en un tiempo t dado es el conjunto de posiciones  $[x,y]$  que satisfacen  $X(t)[x,y].id = i$ , y la segunda área de dibujo en un tiempo t dado es el conjunto de posiciones  $[x,y]$

que satisfacen  $Y(t)[x,y].id = j$ . Si la parte común de dichos conjuntos no es el conjunto vacío, entonces la primera área de dibujo y la segunda área de dibujo se superponen en el tiempo  $t$ , y la parte común de dichos conjuntos corresponde al área de superposición en el tiempo  $t$ .

Para una posición  $[x,y]$  que es un elemento del área de superposición en el tiempo  $t$ , claramente:

- 5  $X(t)[x,y].id = i;$   
 $Y(t)[x,y].id = j$

se establecen.

- 10 En la presente posición  $[x,y]$ , el primer objeto  $i$  ocupa la primera zona de ocupación de  $X(t)[x,y].fore$  a  $X(t)[x,y].back$ , como la profundidad. Además, el primer objeto  $j$  ocupa la segunda zona de ocupación de  $Y(t)[x,y].fore$  a  $Y(t)[x,y].back$ .

- 15 Cuando la primera zona de ocupación y la segunda zona de ocupación se superponen, es decir, cuando la segunda profundidad de superficie anterior o la segunda profundidad de superficie posterior se incluyen en la primera zona de ocupación o la primera profundidad de superficie anterior o la primera profundidad de superficie posterior se incluyen en la segunda zona de ocupación, se determina que el primer objeto  $i$  y el segundo objeto  $j$  colisionan en el tiempo  $t$  y en la posición  $[x,y]$ .

En la explicación descrita más arriba, si la primera área de dibujo en la cual el primer objeto  $i$  se representa y la segunda área de dibujo en la cual el segundo objeto  $j$  se representa se superponen en el fotograma en el tiempo  $t$ , una determinación de colisión para el primer objeto  $i$  y segundo objeto  $j$  en el fotograma en el tiempo  $t$  se logra mediante el uso de información sobre la profundidad de superficie posterior.

- 20 Por consiguiente, con el fin de confirmar que una colisión no existe en un vídeo en su conjunto, estará bien lograr la determinación de colisión descrita más arriba mediante la configuración de tiempos sucesivos para todos los fotogramas en el primer vídeo  $X$  y segundo vídeo  $Y$ .

- 25 La Figura 4 es un diagrama de flujo que muestra el flujo del proceso de determinación de colisión según la presente realización preferida. Más abajo, el proceso de determinación de colisión ejecutado por el dispositivo de procesamiento de vídeo 101 según la presente realización preferida se describe con referencia al presente dibujo.

Primero, el dispositivo de procesamiento de vídeo 101 recibe el primer vídeo  $X$  y el segundo vídeo  $Y$  que son objetivos de determinación de colisión (etapa E151). Los objetos se representan en el primer vídeo  $X$  y segundo vídeo  $Y$ , y la información de profundidad del objeto se anexa.

- 30 A continuación, un proceso de exploración de fotogramas del primer vídeo  $X$  y segundo vídeo  $Y$  en orden desde el inicio se repite (etapa E152).

Es decir, en el fotograma actualmente explorado, se realiza una determinación sobre si un área en la cual un objeto se representa en el primer vídeo  $X$  y un área en la cual un objeto se representa en el segundo vídeo  $Y$  se superponen o no (etapa E153). Si no hay superposición (etapa E153: No), el proceso procede a la etapa E157 y el proceso se repite.

- 35 Si hay superposición (etapa E153: Sí), la información de profundidad sobre el objeto en el primer vídeo  $X$  para el cual hay una superposición en dicha región de dibujo (correspondiente al "primer objeto" descrito más arriba), y la información de profundidad sobre el objeto en el segundo vídeo  $Y$  (correspondiente al "segundo objeto" descrito más arriba) se adquieren (etapa E154).

- 40 Además, se lleva cabo una determinación sobre si existe o no una superposición en las zonas de ocupación de los objetos obtenidos de la información de profundidad (etapa E155). Si no hay superposición, el proceso procede a la etapa E157 y el proceso se repite.

Si hay una superposición (etapa E155: Sí), un resultado de determinación a los fines de que el objeto representado en el primer vídeo  $X$  y el objeto representado en el segundo vídeo  $Y$  colisionan se produce (etapa E156), y el proceso luego concluye.

- 45 Cuando la repetición de las etapas E152 a E157 concluye, un resultado de determinación a los fines de que el objeto representado en el primer vídeo  $X$  y el objeto representado en el segundo vídeo  $Y$  no colisionan se produce (etapa E158), y el proceso entonces concluye.

- 50 En la descripción de más arriba, para facilitar la comprensión se supone que solo un primer objeto se representa en el primer vídeo  $X$  y solo un segundo objeto se representa en el segundo vídeo  $Y$ , pero cuando hay múltiples objetos, estará bien repetir el mismo proceso para cada objeto.

Además, en la descripción de más arriba, una determinación de superposición de zonas de ocupación se ha logrado según la información de profundidad y superposición de áreas de dibujo para cada fotograma, pero cuando hay numerosos fotogramas, es posible lograr un muestreo apropiado.

5 Por ejemplo, si se utiliza un método en el cual inicialmente es posible determinar si la primera área de dibujo y la segunda área de dibujo se superponen o no mediante la selección de fotogramas mientras se hace un salto con igual espaciado, y cuando hay superposición, detener el muestreo de los fotogramas antes y después, es posible ahorrar tiempo de cálculo que se necesita al llevar a cabo la determinación de colisión para los vídeos en su conjunto.

10 En general, cuando no se determina que "en cualquier tiempo y en cualquier posición, cualquiera de los objetos representados en el primer vídeo X y cualquiera de los objetos representados en el segundo vídeo Y colisionan", es decir, cuando no se genera ninguna colisión, no surge contradicción alguna en la relación frontal y posterior de los objetos incluso si el segundo vídeo Y se superpone al primer vídeo X. Por consiguiente, es posible descubrir el tercer vídeo superimpose(Y,X) con el segundo vídeo Y superpuesto al primer vídeo X.

15 Por otro lado, cuando se determina que "en cualquier tiempo y en cualquier posición, cualquiera de los objetos representados en el primer vídeo X y cualquiera de los objetos representados en el segundo vídeo Y colisionan", no es posible superponer el primer vídeo X y el segundo vídeo Y. Por consiguiente, un esquema como, por ejemplo, aquel descrito más abajo es necesario.

#### Segunda realización preferida

20 Con la presente realización preferida, cuando un objeto del primer vídeo X y un objeto del segundo vídeo Y colisionan debido a la cantidad de traslación del segundo vídeo Y especificada por el usuario, dicha colisión se evita y el ajuste automático se logra de modo que el objeto del segundo vídeo Y se posiciona para tocar el suelo o techo del primer vídeo X.

25 Con dicho ajuste automático, una prueba de si una colisión ocurre cuando se traslada el segundo vídeo Y hacia arriba, abajo, izquierda o derecha se repite al menos una vez, y bajo muchas condiciones repetidamente ejecutadas múltiples veces. El presente proceso se logra por el corrector 106 del dispositivo de procesamiento de vídeo 101. La Figura 8 es un diagrama de flujo que muestra el flujo de control del proceso de ajuste automático según la presente realización preferida. La explicación de más abajo hace referencia al presente dibujo.

30 La cantidad de traslación hacia arriba, abajo, izquierda o derecha en una prueba se expresa como  $\Delta p$  en la dirección horizontal y  $\Delta q$  en la dirección vertical. Cuando se usan constantes para  $\Delta p$  y  $\Delta q$ , el segundo vídeo Y se ajusta mientras se desplaza en una dirección preestablecida.

Con el fin de facilitar la comprensión, de manera similar a las Figuras 5A a 5D, se considerará el caso en el cual un usuario hace clic en una posición [s,t] dentro del segundo vídeo Y, arrastra hacia una posición [u,v] dentro del primer vídeo X mientras mantiene el clic y luego detiene el clic y suelta.

35 En el presente caso, el dispositivo de procesamiento de vídeo 101 de la realización preferida logra el proceso de más abajo.

40 Primero, el establecedor de cantidad de desplazamiento 104 del dispositivo de procesamiento de vídeo 101 respectivamente establece el valor inicial de la variable p como u y el valor inicial de la variable q como v (etapa E191). A través de ello, el valor inicial de la cantidad de desplazamiento en la dirección horizontal se establece como p-s = u-s y el valor inicial de la cantidad de desplazamiento en la dirección vertical se establece como q-t = v-t, y el establecedor de cantidad de desplazamiento 104 funciona como un establecedor de valor inicial.

A continuación, el determinador de colisión 103 del dispositivo de procesamiento de vídeo 101 determina si la condición de posición "una colisión de objetos no ocurre entre el primer vídeo X y el vídeo intermedio move(p-s, q-t, Y) que es el segundo vídeo Y desplazado por una cantidad de desplazamiento establecida" se establece o no (etapa E192).

45 Si ello se establece (etapa E192: Sí), el generador de vídeo 105 produce el tercer vídeo superimpose(move(p-s, q-t, Y),X) a través del fotomontaje (etapa E913), y el presente proceso concluye.

Por otro lado, si la presente condición no se establece, el corrector 106 aumenta el valor de la variable p en  $\Delta p$  y aumenta el valor de la variable q en  $\Delta q$  (etapa E194). A través de ello, la cantidad de desplazamiento se actualiza ligeramente.

50 Además, cuando se satisface una condición de abandono como, por ejemplo, la cantidad de traslación (p-s, q-t) del vídeo Y se convierte en demasiado grande, o cuando el número de repeticiones supera un valor umbral preestablecido, y/o similares (etapa E195: Sí), se determina que una colisión no puede evitarse, y se emite una alerta (etapa E196) y el presente proceso concluye.

Por otro lado, cuando la condición de abandono no se satisface y la repetición continúa (etapa E195: No), el control del dispositivo de procesamiento de vídeo 101 regresa a la etapa E192.

5 En los sistemas de coordenadas de numerosos vídeos, la dirección del aumento del valor de coordenada en la dirección vertical corresponde a la dirección hacia abajo en la pantalla. En el posicionamiento mediante el movimiento del objeto representado en el segundo vídeo Y en la dirección hacia abajo de la pantalla, estará bien establecer  $(\Delta p, \Delta q) = (0, 1)$ .

Además, como resultado de  $(\Delta p, \Delta q) = (0, 1)$  estableciéndose inicialmente y el proceso de ajuste automático descrito más arriba ejecutándose, cuando la determinación es que una colisión es inevitable, estará bien reiniciar  $(\Delta p, \Delta q) = (0, -1)$  y ejecutar nuevamente el mismo proceso de ajuste automático.

10 Ello corresponde, primero, a una búsqueda de una cantidad de traslación en la cual una colisión puede evitarse mientras se mueve de arriba abajo, y si esta no puede encontrarse, entonces una búsqueda de una cantidad de traslación en la cual una colisión puede evitarse mientras se mueve de abajo arriba.

15 En la explicación descrita más arriba, el determinador de colisión 103 se ha utilizado simplemente como el determinador de interferencia. Ello corresponde a imponer la condición de posición de colisiones de objetos que no ocurren para la relación de posición entre objetos. Sin embargo, si se usan los resultados de determinación del determinador de colisión 103, es posible utilizar otras condiciones como condiciones de posición en el determinador de interferencia.

20 Por ejemplo, es posible cambiar la condición de posición a "no ocurre ninguna colisión de objetos entre el primer vídeo X y el vídeo  $\text{move}(p-s, q-t, Y)$  y 'una colisión de objetos ocurre entre el primer vídeo X y el vídeo  $\text{move}(p-s, q+1-t, Y)$ ". Mediante dicho cambio, el objeto representado en el segundo vídeo Y se posiciona para tocar virtualmente el suelo cuando superpone el objeto representado en el segundo vídeo Y al primer vídeo X.

Además, también es posible cambiar la condición de posición a "no ocurre ninguna colisión de objetos entre el primer vídeo X y el vídeo  $\text{move}(p-s, q-t, Y)$  y 'una colisión de objetos ocurre entre el primer vídeo X y el vídeo  $\text{move}(p+\Delta p-s, q+\Delta q-t, Y)$ ".

25 En el presente caso, cuando se descubre una cantidad de traslación en una dirección dada, la configuración se lleva a cabo para proceder hacia adelante en la medida de lo posible y la prueba se repite. Por ejemplo, cuando se mueve de arriba abajo, el objeto del segundo vídeo Y se mueve tan lejos como para tocar el suelo del primer vídeo X, y cuando se mueve de abajo arriba, el objeto del segundo vídeo Y se mueve tan lejos como para tocar el techo del primer vídeo X.

30 En casos en los cuales el suelo se representa en el primer vídeo X, un objeto que se mueve en un plano horizontal se representa en el segundo vídeo Y y la orientación de la cámara en el primer vídeo X con respecto al suelo y la orientación de la cámara en el segundo vídeo Y con respecto a la orientación horizontal sustancialmente concuerdan y/o similares, si el proceso descrito más arriba se logra, cuando el objeto representado en el segundo vídeo Y se superpone al primer vídeo X, el objeto representado en el segundo vídeo Y sustancialmente toca el suelo o se posiciona para flotar separado del suelo.

Hay casos en los cuales un objeto representado en el primer vídeo X permanece quieto como, por ejemplo, el suelo, techo, pared y/o similares. Por ejemplo, hay casos en los que un usuario ha especificado un objeto en un fotograma dado con antelación, casos en los que el historial de edición del pasado por el usuario se comprende, y casos en los que cada fotograma del primer vídeo X se comprende a través del contraste.

40 En el presente caso, establecer las condiciones de posición según se describe más arriba significa que el objeto compuesto establece el rango movable que puede ocuparse en el espacio tridimensional. Es decir, es posible lograr el fotomontaje de vídeo en la cual las condiciones para el rango movable del objeto en el espacio tridimensional se imponen como, por ejemplo, "el objeto puede flotar en el espacio si no hay colisión alguna", "el objeto toca el suelo o la tierra", "el objeto toca el techo" y/o similares.

45 Según se describe más arriba,  $(\Delta p, \Delta q)$  es un vector que expresa el desplazamiento de la posición de dibujo del objeto representado en el segundo vídeo Y, y mediante la configuración de la orientación de aquel, es posible cambiar de forma apropiada la dirección de movimiento y la cantidad de movimiento.

50 En la situación descrita más arriba, mediante el movimiento del segundo vídeo Y a lo largo de una dirección establecida con antelación y mediante la superposición de la imagen al primer vídeo X, un nuevo vídeo se genera de modo que los objetos representados en el primer vídeo X y segundo vídeo Y no colisionan, y el movimiento paralelo solamente en la dirección hacia arriba o hacia abajo del vídeo Y se supone.

Sin embargo, después de que el movimiento descrito más arriba se ha logrado, estará bien corregir la cantidad de desplazamiento para minimizar la diferencia entre la cantidad de desplazamiento y el valor inicial establecido por el

usuario, mientras también se satisfacen las condiciones de posición. En el presente caso, se usa la cantidad de desplazamiento más cercana a la designación del usuario.

Además, como la dirección de  $(\Delta p, \Delta q)$ , es posible utilizar la diferencia de vector entre un punto representativo del primer objeto y un punto representativo del segundo objeto. En el presente caso, se lleva a cabo una traslación de modo que el segundo objeto se mueve en una dirección que se aproxima al primer objeto o en una dirección que se aleja del primer objeto.

Continuando con dichos métodos, estará bien utilizar un método que establece, de forma apropiada, una función de evaluación y evitar colisiones de objetos mediante la minimización de la presente función. Ello puede pensarse como minimización de la "cantidad de colisión" cuando los objetos colisionan.

10 Cuando se considera la cantidad de colisión del vídeo X y vídeo Y como la suma de las áreas de zonas de colisión donde los objetos que colisionan se representan en el vídeo X y vídeo Y, la cantidad de colisión puede calcularse a través de:

$$\text{collision}(Y, X) = \sum_{t, x, y; X(t)[x, y].id \geq 0, Y(t)[x, y].id > 0, \text{overlap}(X, Y, t, x, y) > 0} 1$$

Además, cuando la cantidad estimada de la suma de volúmenes donde los objetos que colisionan se superponen en el espacio tridimensional se toma como la cantidad de colisión, es posible utilizar:

$$\text{collision}(Y, X) = \sum_{t, x, y; X(t)[x, y].id \geq 0, Y(t)[x, y].id > 0, \text{overlap}(X, Y, t, x, y) > 0} \text{overlap}(X, Y, t, x, y).$$

La cantidad de colisión calculada de la presente manera se toma como la función de evaluación.

20 Con el fin de minimizar la cantidad de colisión, estará bien actualizar, de forma repetida, el valor de la variable que indica el grado de desplazamiento del vídeo Y como, por ejemplo, p y q y/o similares mediante el establecimiento de la dirección de actualización  $(\Delta p, \Delta q)$  de la variable no activada a través de un recorrido aleatorio, método del descenso más pronunciado o método del gradiente conjugado, preparar  $(1, 0)$ ,  $(0, 10)$ ,  $(-1, 0)$ ,  $(0, -1)$  y/o similares con antelación como candidatos para  $(\Delta p, \Delta q)$ , encontrar el cambio en la cantidad de colisión cuando se utilizan estos y seleccionar aquel para el cual la cantidad de colisión más se reduce.

25 Por ejemplo, cuando se logra la traslación, es posible encontrar las cantidades de traslación p-s y q-t para evitar colisiones entre objetos si s, t, u y v se establecen según se describe más arriba, los valores iniciales de las variables p y q se toman para que sean u y v, se descubren p y q que minimizan  $\text{collision}(\text{move}(p-s, q-t, Y), X)$  y  $\text{collision}(\text{move}(p-s, q-t, Y), X) = 0$  se establece en el punto final al que se llega.

Además, aparte de las traslaciones hacia arriba, abajo, izquierda y derecha, estará bien combinar, de forma arbitraria, la traslación en la dirección de profundidad, rotación, escalamiento y desplazamiento en el tiempo.

30 Por ejemplo, cuando se combinan las traslaciones hacia arriba, abajo, izquierda o derecha y en la dirección de profundidad, s, t, u y v se establecen según se describe más arriba, u, v y 0 se toman como valores iniciales de p, q y r, y se descubre que p, q y r minimizan  $\text{collision}(\text{push}(r, \text{move}(p-s, q-t, Y)), X)$ .

35 Cuando se combinan rotaciones alrededor del eje horizontal y eje vertical y traslaciones, s, t, u y v se establecen según se describe más arriba, u, v, 0 y 0 se toman como los valores iniciales de p, q,  $\theta$  y  $\varphi$ , y se descubre que p, q,  $\theta$  y  $\varphi$  minimizan  $\text{collision}(\text{rotver}(\varphi, \text{rothor}(\theta, \text{move}(p-s, q-t, Y))), X)$ .

40 Cuando ocurren traslaciones o rotaciones extremas, la representación de forma convencional de la superficie posterior del objeto se vuelve necesaria. Sin embargo, si el vídeo Y es un vídeo de dimensión 2.5, no hay información sobre la superficie posterior del objeto. Por lo tanto, los límites superiores y límites inferiores se establecen con antelación para las cantidades de traslación p-s, q-t y r y las cantidades de rotación  $\theta$ ,  $\varphi$ ,  $\psi$  y/o similares, y a través de ello es posible evitar que el movimiento de un objeto en el vídeo Y se vea de forma no natural.

45 Además, con el fin de reducir la cantidad de desplazamiento en la medida de lo posible, también existe un método para tomar como la función de evaluación el resultado de añadir la cantidad de colisión descrita más arriba a un valor encontrado mediante la multiplicación de una constante de entero preestablecida por el valor al cuadrado o valor absoluto de un parámetro de transformación de las cantidades de traslación p-s, q-t o r o las cantidades de rotación  $\theta$ ,  $\varphi$ ,  $\psi$  y/o similares y continuando con la actualización del parámetro de transformación. Cuando el valor inicial del parámetro de transformación se establece en un valor especificado por el usuario, si el desplazamiento en el parámetro de transformación se combina con la función de evaluación, es posible encontrar la cantidad de desplazamiento más cercana al valor especificado por el usuario.

50 Además, más abajo se considera una situación en la cual la cantidad de traslación en la dirección de profundidad se ajusta por un comando del usuario.

Primero, con la cantidad de traslación en la dirección de profundidad tomada como 0, en el método descrito más arriba p y q se seleccionan de modo que el objeto toca el suelo. Como resultado, un vídeo con el vídeo  $\text{move}(p-s, q-t, Y)$  superpuesto al primer vídeo X se ha obtenido.

5 Luego, el usuario especifica una cantidad de traslación r en la dirección de profundidad a través del funcionamiento de la rueda del ratón y/o similares.

Luego, la actualización de la variable (p,q) se repite mediante el uso  $(\Delta p, \Delta q)$  de aquella como en la situación descrita más arriba hasta que la condición "los objetos no colisionan entre el primer vídeo X y el vídeo  $\text{push}(r, \text{move}(p-s, q-t, Y))$ " y "los objetos colisionan entre el primer vídeo X y el vídeo  $\text{push}(r, \text{move}(p-s, q+1-t, Y))$ " se satisfaga.

10 Mediante la utilización de dichas situaciones, es posible generar un vídeo en el cual el movimiento de profundidad solo sea la cantidad de traslación especificada por el usuario y el objeto continúe tocando el suelo.

15 Cuando se ajusta de forma automática, es posible primero utilizar una situación en la cual el tercer vídeo Z se genera mediante la formación de un fotomontaje del primer vídeo X y el segundo vídeo Y (o un vídeo en el cual el segundo vídeo Y se desplaza según los comandos de usuario). En la presente situación, cuando ocurre el fotomontaje, si cada píxel en el tercer vídeo Z se deriva de uno de los objetos y un objeto dado se representa como escondiendo un objeto diferente, el valor de píxel en la zona escondida del objeto escondido y la información de profundidad se almacenan de forma separada.

20 El tercer vídeo Z se corrige mientras se hace referencia a la presente información almacenada de modo que no ocurren colisiones de objetos. Al llevar a cabo dicha corrección, hay ocasiones en las que la dirección de profundidad del objeto cambia. En el presente caso, estará bien hacer referencia a la información de píxel en el segundo vídeo Y original para preservar en la medida de lo posible la resolución de la apariencia exterior del objeto dado que el escalamiento se lleva a cabo cuando el objeto se representa dentro del fotograma.

Dichas situaciones, cuando se comparan con los métodos descritos más arriba, simplemente sustituyen el orden del proceso de cálculo como una situación en la cual el método para almacenar información difiere, y el principio de acción de aquel puede verse como sustancialmente el mismo.

25 Además, se considera que la cantidad de colisión muestra la cantidad de penetración entre objetos, entonces es posible lograr el fotomontaje de un vídeo mediante la imposición de la condición de que "cierto grado de penetración entre objetos está bien".

Incluso si se utilizan los varios métodos descritos más arriba, hay ocasiones en las que una colisión de objetos es inevitable. En dichos casos, esquemas como, por ejemplo, aquellos explicados más abajo son necesarios.

30 Tercera realización preferida

Con la segunda realización preferida descrita más arriba, cuando el usuario ha especificado la cantidad de traslación hacia arriba, abajo, izquierda o derecha para un objeto en el vídeo Y y ha intentado superponer esta al vídeo X, si surge una colisión de objetos, la explicación es para una situación para ajustar, de forma automática, la cantidad de traslación.

35 La presente realización preferida puede aplicarse tanto a la realización preferida descrita más arriba como al ejemplo dado en los pasajes comenzando en el párrafo [0328], y muestra un área donde el usuario puede soltar un objeto, es decir, un área que comprende una posición de caída donde un vídeo en el cual los objetos no colisionan se genera a través de una función de soltar por el usuario, en una manera fácil de comprender anterior a la función de soltar por el usuario.

40 La Figura 9A es un dibujo explicativo que muestra la condición de la pantalla con anterioridad al inicio de una caída, la Figura 9B es un dibujo explicativo que muestra la condición de la pantalla en el punto en el tiempo en el que la caída comienza, la Figura 9C es un dibujo explicativo que muestra la condición de la pantalla en un punto en el tiempo durante una caída, y la Figura 9D es un dibujo explicativo que muestra la condición de la pantalla en el punto en el tiempo en el que se logra la caída comienza. La explicación de más abajo hace referencia a dichos dibujos.

45 La Figura 10 es un diagrama de flujo que muestra el flujo de control de un proceso de visualización destacado para un área donde el objeto puede soltarse. La explicación de más abajo hace referencia a dicho dibujo.

50 Primero, una disposición F que tiene como con índice candidatos para cantidades de desplazamiento arbitrario utilizables cuando desplaza el segundo vídeo Y en la dirección horizontal y dirección vertical se prepara en la RAM (etapa E501). Más abajo, se supone que [ps, qt] se prepara como el índice, y que ps y qt son valores negativos. Es decir, ps puede ser un entero entre el valor encontrado mediante la negación del signo del ancho del segundo vídeo Y y la suma del ancho del primer vídeo X y el ancho del segundo vídeo Y, y qt puede ser un entero entre el valor encontrado mediante la negación del signo de la altura del segundo vídeo Y y la suma de la altura del primer vídeo X y la altura del segundo vídeo Y.

Además, para los candidatos [ps, qt] para la cantidad de desplazamiento utilizable para el segundo vídeo Y (etapa E501), una determinación de colisión para un objeto se logra entre el primer vídeo X y el vídeo intermedio move(ps, qt, Y) que es el segundo vídeo Y desplazado (etapa E502), y un proceso para sustituir resultados de la determinación de colisión en los elementos F[ps,qt] de la matriz F (etapa E503) se repite (etapa E504).

- 5 Luego, de manera similar a la realización preferida descrita más arriba, la ventana 411 del primer vídeo X y la ventana 412 del segundo vídeo Y se muestran en la pantalla (etapa E506) y el inicio de un arrastre se recibe (etapa E507).

Aquí, el arrastre comienza cuando el usuario hace clic en una posición [s,t] dentro del segundo vídeo Y mediante el uso de un ratón.

- 10 Suponiendo que la posición [s,t] en el segundo vídeo Y se suelta en una posición [p,q] dentro del primer vídeo X, el vídeo intermedio encontrado mediante el desplazamiento del segundo vídeo Y se convierte en move(p-s, q-t, Y).

Por consiguiente, estará bien remitirse al elemento F[p-s, q-t] de la matriz para determinar si los objetos colisionan o no entre el primer vídeo X y el vídeo intermedio move[p-s, q-t, Y].

- 15 El dispositivo de procesamiento de vídeo 101, para cada posición [p,q] dentro del primer vídeo X (etapa E508), repite (etapa E510) un proceso para hacer que el color (cualquiera de la saturación de color, brillo o tonalidad, o una combinación de ellos) de los píxeles en cada posición [p,q] dentro del primer vídeo X cambie (etapa E509) según el valor de F[p-s, q-t].

A través de ello, el usuario puede distinguir entre áreas donde las caídas son posibles y áreas donde las caídas no son posibles, y una de las áreas se muestra resaltada.

- 20 Los métodos para lograr una visualización resaltada incluyen "hacer que el color de los píxeles que se resaltarán cambie y hacer que el color de otros píxeles no cambie" o "no hacer que el color de píxeles que se resaltarán cambie y hacer que el color de otros píxeles cambie".

Luego, en las realizaciones preferidas descritas más arriba, estará bien proceder a la etapa E183.

- 25 La explicación de más abajo cita un ejemplo. Con anterioridad al inicio del arrastre, el primer vídeo X se representa sin cambios en la ventana 411, como se muestra en la Figura 9A.

En el instante en el que comienza un arrastre, el dispositivo de procesamiento de vídeo calcula el área en la que puede soltarse. Además, el área en la que puede soltarse dentro de la ventana 411 se muestra resaltada, como se muestra en la Figura 9B. En el presente dibujo, el resaltado se ilustra mediante un rayado.

- 30 Si el usuario arrastra el cursor 413 hacia dentro del área en la que puede soltarse, como se muestra en la Figura 9C, y luego logra un proceso de caída como se muestra en la Figura 9D, se obtiene un tercer vídeo en el cual los objetos no colisionan.

Con el método descrito más arriba, es posible hacer que la velocidad de experiencia de la visualización resaltada después de un clic mejore, y es posible lograr un redibujo de alta velocidad cuando se vuelve a hacer clic (no se muestra).

- 35 Sin embargo, cuando puede llevarse a cabo una velocidad de cálculo suficiente, estará bien, cuando el usuario inicia una función de arrastre, lograr una determinación de colisión y resaltado de efecto para una posición arbitraria [p,q] dentro del primer vídeo X después de que [s,t] se haya establecido, según la cantidad de desplazamiento [p-s, q-t].

De esta manera, con el presente método cuando el usuario comienza a arrastrar el segundo vídeo Y, es fácil comprender dónde en el primer vídeo X una función de caída puede llevarse a cabo para obtener un nuevo vídeo.

- 40 El ajuste automático de la cantidad de traslación en la segunda realización preferida corresponde a corregir la posición de caída a una dentro del área donde puede soltarse cuando una caída se lleva a cabo fuera del área donde puede soltarse.

- 45 En la presente realización preferida, cuando la presente realización preferida se utiliza después de determinar si las caídas son posibles o no para todas las posiciones dentro del fotograma del primer vídeo X, cuando una caída se lleva a cabo fuera del área donde puede soltarse, mediante la simple corrección de la posición de caída por "un punto más cercano a dicha posición de caída dentro del área donde puede soltarse" o "el punto central del área donde puede soltarse" y/o similares, es posible lograr el ajuste automático de la cantidad de traslación hacia arriba abajo, izquierda o derecha.

La corrección de la posición de caída por el límite que delinea el área donde puede soltarse y diferente del área donde puede soltarse significa que hay un instante en el cual un objeto en el primer vídeo y un objeto en el segundo vídeo se tocan mutuamente.

5 Por consiguiente, es concebible que con el fin de hacer que el objeto toque el suelo en la medida de lo posible, estará bien corregir la posición de caída por el límite inferior del área donde puede soltarse.

La Figura 9E es un dibujo explicativo que muestra la condición en la cual una corrección automática de la posición de caída se logra después de una caída. Según se muestra en el presente dibujo, la posición  $[s,t]$  dentro de la ventana 412 se corrige para superponerse al límite inferior del área donde puede soltarse, debajo de la posición  $[u,v]$  dentro de la ventana 411.

10 Con la presente realización preferida, es fácil para el usuario comprender posiciones donde puede soltarse, e incluso cuando una caída se lleva a cabo en una posición donde no puede soltarse, la posición de caída puede corregirse fácilmente mediante el uso de información ya calculada.

Cuarta realización preferida

15 Las coordenadas de un punto representativo en el espacio tridimensional para el objeto  $i$  represento en el vídeo  $X$  pueden expresarse como  $(x_c(X,t,i), y_c(X,t,i), z_c(X,t,i))$ , según se describe más arriba. Ello puede verse como un vector de posición tridimensional  $pos[X,i](t)$  que cambia con el tiempo.

20 En las realizaciones preferidas descritas más arriba, cuando el objeto  $j$  representado en el vídeo  $Y$  se superpone al vídeo  $X$ , cuando el objeto  $j$  colisiona con un objeto representado en el vídeo  $X$ , un vídeo  $W$  se obtiene mediante la traslación del vídeo  $Y$  en las direcciones arriba, abajo, izquierda o derecha o en la dirección de profundidad, o mediante la rotación del vídeo  $Y$ , o mediante un desplazamiento en el tiempo, y el presente vídeo  $W$  se superpone entonces al vídeo  $X$ .

25 Dado que la traslación y rotación y/o similares son transformaciones congruentes, cuando solo dichas transformaciones se usan el vector de posición  $pos[Y,j](t)$  y el vector de posición  $pos[W,j](t)$  que son trayectorias del objeto  $j$  pueden tener posiciones y orientaciones que difieren pero la forma es la misma. Es decir, se aplica una transformación no deformante que no deforma la trayectoria, entonces la trayectoria  $pos[Y,j](t)$  y la trayectoria  $pos[W,j](t)$  son congruentes.

La Figura 11A es un dibujo explicativo que muestra la condición de la trayectoria  $pos[X,i](t)$  y la trayectoria  $pos[Y,j](t)$ .

La Figura 11B es un dibujo explicativo que muestra la condición de la trayectoria  $pos[X,i](t)$  y la trayectoria  $pos[W,j](t)$ .

30 En dichos dibujos, la posición de cada trayectoria en la posición horizontal se muestra en el eje horizontal y el paso del tiempo se muestra en el eje vertical. Además, a lo largo de cada trayectoria, se forman rectángulos estrechos que indican la extensión de dicho objeto en la dirección horizontal.

La trayectoria  $pos[W,j](t)$  en la Figura 11B es la trayectoria  $pos[Y,j](t)$  en la Figura 11A movida en la dirección horizontal, y las formas de las dos son congruentes.

35 Si el objeto  $i$  en el vídeo  $X$  y el objeto  $j$  en el vídeo  $Y$  colisionan en el tiempo  $T$ , como se muestra en la Figura 11A, la trayectoria  $pos[X,i](t)$  y la trayectoria  $pos[Y,j](t)$  se tocan en el tiempo  $T$ , y los rectángulos que representan la extensión de los objetos en la dirección horizontal se superponen.

40 Por otro lado, no ocurre colisión alguna entre el objeto  $i$  del vídeo  $X$  y el objeto  $j$  del vídeo  $W$ . Por consiguiente, como se muestra en la Figura 11B, los rectángulos que representan la extensión de los objetos en la dirección horizontal no se superponen.

45 Sin embargo, con el presente tipo de transformación congruente, hay casos en los cuales las colisiones son inevitables. Aquí, se provee un ejemplo en el cual el corrector 106 evita colisiones entre objetos mediante la aplicación en el vídeo de una transformación deformante que deforma la trayectoria del objeto. En la medida en que el presente ejemplo implica el uso de transformaciones deformantes, este no cae dentro del alcance de la invención reivindicada, y se provee a los fines ilustrativos solamente.

Más abajo, se explica un método en el cual la trayectoria  $pos[X,i](t)$  del objeto  $i$  representado en el vídeo  $X$  que tiene un fondo no se deforma, pero la trayectoria  $pos[Y,j](t)$  del objeto  $j$  representado en el vídeo  $Y$  superpuesto al vídeo  $X$  se deforma.

50 Primero, el vector de posición  $pos[Y,j](t)$  indica que la trayectoria se deforma para convertirse en el vector de posición  $morph(t)$ .

Más abajo, los diferentes valores de coordenadas del vector de posición se expresan anexando .x, .y y .z.

A través de la deformación de la trayectoria, la profundidad se convierte en  $\text{morph}(t).z/\text{pos}[Y,j](t).z$  veces. Ello significa que en el tiempo  $t$ , el objeto  $j$  se amplía a  $\text{pos}[Y,j](t).z/\text{morph}(t).z$  veces dentro del fotograma del vídeo  $Y$ .

5 Además, en el tiempo  $t$  el objeto  $j$  se mueve en la dirección horizontal y dirección vertical. Las cantidades de dichos movimientos son respectivamente  $(\text{morph}(t).x - \text{pos}[Y,j](t).x) \times \text{pos}[Y,j](t).z/\text{morph}(t).z$ ,  $(\text{morph}(t).y - \text{pos}[Y,j](t).y) \times \text{pos}[Y,j](t).z/\text{morph}(t).z$ .

Por consiguiente, cuando la trayectoria se deforma, mediante la combinación de la selección del objeto, traslación y escalamiento, es posible obtener un vídeo en el cual la trayectoria de un objeto dado se deforma.

10 Si la deformación de una trayectoria se asume para evitar una colisión entre objetos, mediante la superposición del vídeo con la trayectoria deformada, es posible generar un vídeo en el cual los objetos representados en múltiples vídeos se representan en un estado que se mueve sin colisionar.

Más abajo, se describen varios métodos para deformar trayectorias.

15 En un primer método, en un tiempo  $t$  en el cual  $\text{overlap}(X,Y,t,x,y) > 0$  se establece, un proceso de deformación de la trayectoria del objeto  $j$  de modo que la posición  $\text{pos}[Y,j](t)$  de un punto representativo del objeto  $j = Y(t)[x,y].id$  contenido en el vídeo  $Y$  se traslada en una dirección lejos de la posición  $\text{pos}[X,i](t)$  de un punto representativo del objeto  $i = X(t)[x,y].id$  contenido en el vídeo  $X$ , o en una dirección del vector normal primario hacia la trayectoria  $\text{pos}[Y,j](t)$ , se repite de modo que el cambio en el grado de deformación antes y después del tiempo se convierte en más pequeño, y de modo que  $\text{overlap}(X,Y,t,x,y) = 0$  se establece en todo momento.

20 Las Figuras 12A a 12D son dibujos explicativos que muestran el estado en el cual la trayectoria gradualmente se deforma mediante la repetición del proceso. La explicación de más abajo hace referencia a dichos dibujos.

En la Figura 12A, similar a la Figura 11A, la trayectoria  $\text{pos}[X,i](t)$  y la trayectoria  $\text{pos}[Y,j](t)$  se tocan en el tiempo  $T$ .

En la Figura 12B, la posición del objeto  $j$  en el tiempo  $T$  se corrige de modo que no ocurre ninguna colisión.

25 Dado que dicha cantidad de corrección es grande, en la Figura 12C las posiciones del objeto  $j$  en los tiempos  $T-1$  y  $T+1$  antes y después del tiempo  $T$  se corrigen por una cantidad más pequeña que la cantidad de corrección en la Figura 12B.

Dado que dicha cantidad de corrección es aún grande, en la Figura 12D las posiciones del objeto  $j$  en los tiempos  $T-2$  y  $T+2$  antes y después de la corrección previa se corrigen más por una cantidad más pequeña que la cantidad de corrección en la Figura 12C.

30 Con ello, la corrección es suficiente y la trayectoria que se muestra en la Figura 12D corresponde a la trayectoria  $\text{morph}(t)$  finalmente obtenida.

La cantidad de corrección propagada a fotogramas adyacentes se reduce mediante la elevación de la cantidad de corrección para el propio fotograma a una potencia constante no menor que 0 pero menor que 1, y si la cantidad de corrección es más pequeña que un valor umbral preestablecido, estará bien detener la propagación.

35 Cuando se deforma la trayectoria, estará bien no propagar las correcciones antes y después del tiempo según se describe más arriba pero lograr la corrección suave mediante la utilización de la interpolación mediante *splines*.

Además, con el presente método, si aún ocurre una colisión después de lograr el tipo de corrección y propagación de cantidad de corrección descrito más arriba en el punto en el tiempo en el que la cantidad de colisión es mayor, es posible hacer que la cantidad de corrección y frecuencia de corrección sean tan pequeñas como sea posible mediante la repetición del mismo proceso.

40 Un segundo método es un método que, cuando deforma la trayectoria  $\text{pos}[Y,j](t)$  en el tiempo  $t$ , provoca solo el tamaño del vector normal primario (un vector ortogonal al vector de velocidad y correspondiente a una dirección que dobla la orientación del vector de velocidad) sin cambiar el tamaño del vector tangente para la trayectoria (correspondiente al vector de velocidad) y minimiza a 0  $\text{collision}(Y,X)$  mientras también minimiza la suma de cambios en el tamaño del vector normal primario (normalmente, la suma de cuadrados).

45 La Figura 13 es un dibujo explicativo que muestra el estado en el cual el vector normal primario se ajusta en partes específicas de la trayectoria. El presente dibujo muestra las formas de la trayectoria sin cambios.

En el presente dibujo, la trayectoria  $\text{pos}[X,i](t)$  colisiona con la trayectoria  $\text{pos}[Y,j](t)$  en el tiempo  $T$ . Por lo tanto, en la zona ajustada  $T-a$  a  $T+a$  antes y después del tiempo  $T$ , la curva de la trayectoria  $\text{pos}[Y,j](t)$  se ajusta de modo que no ocurre una colisión, y se obtiene una trayectoria  $\text{poscorrección morph}(t)$

Las formas de la trayectoria  $\text{pos}[Y,j](t)$  y trayectoria  $\text{morph}(t)$  después del tiempo  $T+a$  son congruentes.

En el primer método, el cálculo es simple pero la divergencia entre la trayectoria y el movimiento del objeto  $j$  expresada por el vídeo se vuelve grande y, en este momento, hay ocasiones en las que parece que el objeto  $j$  está haciendo repentinamente una "caminata lunar".

5 Si el objeto  $j$  lleva a cabo o no movimientos no naturales puede determinarse a través de si el tamaño y la orientación del vector tangente a la trayectoria del objeto  $j$ , el tamaño del vector normal primario (este es un valor correspondiente a la curvatura de la trayectoria) y el cambio en la orientación superan o no los valores umbrales preestablecidos.

10 Cuando se determina que el movimiento del objeto no es natural debido a la deformación de la trayectoria, si se produce una alerta a tal efecto, la conveniencia del usuario es alta.

Al repetir el proceso de deformación de la trayectoria, estará bien establecer con antelación un límite superior para la cantidad de deformación de la trayectoria. En el presente caso, el movimiento del objeto no se convierte en no natural, pero existe una posibilidad de que pueda no ser posible evitar colisiones.

15 Con el segundo método, los cálculos son más complejos pero la dirección de movimiento del objeto  $j$  cambia solo ligeramente y el tamaño de la velocidad de movimiento no cambia, de modo que no ocurre ninguna "caminata lunar".

Cuando un fondo se representa en el vídeo  $X$  junto con el objeto  $i$ , cuando se hace que la trayectoria del objeto  $i$  cambie, hay casos en los cuales espacios sin información de píxel ocurren entre el fondo y el objeto  $i$ . Por lo tanto, con el presente ejemplo, solo la trayectoria del objeto  $j$  representado en el vídeo  $Y$  se deforma.

20 Sin embargo, cuando el fondo no se contiene dentro del vídeo  $X$ , cuando el objeto  $i$  y objeto  $j$  colisionan, estará bien deformar ambas trayectorias. Con el primer método, estará bien hacer que el objeto  $i$  y objeto  $j$  se muevan en direcciones que se distancian mutuamente más, y con el segundo método, estará bien minimizar tanto la cantidad de cambio en el vector normal primario de la trayectoria del objeto  $i$  como la cantidad de cambio en el vector normal primario de la trayectoria del objeto  $j$ .

25 En dichos métodos, es posible prevenir el movimiento no natural del objeto y aumentar la conveniencia del usuario mediante el establecimiento de un límite superior sobre la cantidad de cambio en el vector normal primario o producir una alerta cuando la cantidad de cambio en el vector normal primario supera un valor umbral preestablecido.

En la presente configuración, incluso cuando una colisión es inevitable con una simple traslación, rotación, escalamiento y/o similares de los objetos representados en los vídeos, es posible prevenir una colisión mediante la deformación de la trayectoria.

30 En el presente ejemplo y en los ejemplos descritos más arriba, se describen métodos para evitar colisiones de objetos. Sin embargo, también es posible imponer la condición restrictiva de que el contacto entre las superficies de los objetos ocurra en un punto deseado en el tiempo.

35 Es decir, mediante la imposición de la restricción de que en el punto deseado en el tiempo hay una superposición en la zona de ocupación de la profundidad de superficie anterior a la profundidad de superficie posterior de los dos objetos en una posición donde los dos objetos se representan superpuestos, y la superposición es 0, el movimiento del objeto y la deformación de trayectorias se logran de modo que la restricción se satisface.

40 Además, estará bien provocar la deformación de ambas trayectorias de los objetos  $i$  y  $j$  representados en ambos vídeos  $X$  e  $Y$ . En el presente caso, estará bien lograr la deformación para minimizar la cantidad de colisión mediante la compilación y tratamiento de los parámetros de deformación para el objeto  $i$  y los parámetros de deformación para el objeto  $j$ .

Quinta realización preferida

Cuando el objeto  $j$  representado en el vídeo  $Y$  es un objeto que baila sobre una superficie plana y un suelo se incluye en el fondo en el vídeo  $X$ , es deseable que el objeto  $j$  se superponga para bailar sobre el suelo del vídeo  $X$ . La presente realización preferida lleva esto a cabo.

45 Con la realización preferida descrita más arriba, la explicación era para la condición en la cual la inclinación de una cámara con respecto al suelo en el vídeo  $X$  y la inclinación de la cámara con respecto a un plano en el vídeo  $Y$  concuerdan, pero en la presente realización preferida este tipo de concordancia no se requiere necesariamente.

50 Primero, como un punto representativo del objeto  $j$  representado en el vídeo  $Y$ , se selecciona un punto representado lo más lejos posible hacia abajo en la pantalla. Cuando el objeto  $j$  es un personaje que baila, el punto representativo seleccionado de esta manera corresponde a los pies del personaje.

La posición de píxel [xb(Y,t,j), yb(Y,t,j)] del punto representativo puede calcularse, por ejemplo, de la siguiente manera:

$$yb(Y,t,j) = \min_{x,y;Y(t)[x,y].id=j} Y$$

$$xb(Y,t,j) = \min_{x,y;Y(t)[x,yb(Y,t,j)].id=j} X.$$

5 Aquí, si el punto más bajo del objeto j hace contacto con el suelo del vídeo X, entonces:

$$X(t)[xb(Y,t,j), xb(Y,t,j)].id = 0;$$

$$Y(t)[xb(Y,t,j), xb(Y,t,j)].id = j;$$

10  $X(t)[xb(Y,t,j), xb(Y,t,j)].fore = Y(t)[xb(Y,t,j), xb(Y,t,j)].fore$

deben establecerse.

Por consiguiente, si la cantidad de traslación en las direcciones hacia arriba, abajo, izquierda y derecha y también la dirección de profundidad del vídeo Y y la cantidad de rotación alrededor de los diferentes ejes se determina de modo que además de minimizar collision(Y,X) a 0, la suma de las diferencias (normalmente, suma de cuadrados) de  $X(t)[xb(Y,t,j), xb(Y,t,j)].fore$  y  $Y(t)[xb(Y,t,j), xb(Y,t,j)].fore$  también se minimiza, el objeto j representado en el vídeo Y toca la superficie del suelo representada en el vídeo X.

15 Cuando el objeto j salta desde el suelo en el medio de un baile, es imposible que  $X(t)[xb(Y,t,j), xb(Y,t,j)].fore$  y  $Y(t)[xb(Y,t,j), xb(Y,t,j)].fore$  concuerden perfectamente, pero mediante la minimización, es posible prevenir la condición de separación del suelo en instantes en los que no se salta.

20 Además, a través de la minimización, es posible hacer que el comportamiento del objeto j superpuesto al vídeo X sea natural en la medida de lo posible haciendo que el suelo sobre el cual se encuentra el objeto j en el vídeo Y y el suelo en el vídeo X concuerden en la medida de lo posible.

Sexta realización preferida

25 Con la realización preferida descrita más arriba, en las diferentes transformaciones del vídeo de dimensión 2.5, un proceso de cálculo independiente se ha logrado para cada posición e información de profundidad dentro de un solo objeto, y se han realizado cálculos para reflejar ello en la medida de lo posible. En la presente realización preferida, la carga de cálculo se reduce más mediante el logro de aproximaciones mediante el uso de puntos representativos de objetos.

30 Es decir, en la presente realización preferida, las coordenadas (xc(X,t,i), yc(X,t,i)) en el vídeo X del punto representativo del objeto i en el tiempo t en el vídeo X y la coordenada de profundidad zc(X,t,i) del punto representativo se calculan.

35 El objeto i se mueve como un grupo en su totalidad, entonces las coordenadas de profundidad de las diferentes partes del objeto i pueden aproximarse por zc(X,t,i). Por consiguiente, cuando la profundidad del punto representativo cambia de zc(X,t,i) a zc(X,t,i)+r, la profundidad en cada posición se aproxima como [zc(X,t,i)+r / zc(X,t,i)] veces. El presente tipo de aproximación es eficaz para traslaciones en la dirección de profundidad, y rotaciones alrededor del eje horizontal o eje vertical, del vídeo.

40 De manera específica, el destino de movimiento de solamente el punto representativo del objeto i que es el objetivo de la traslación o rotación alrededor del eje horizontal o eje vertical se descubre a través de los mismo cálculos que en las realizaciones preferidas descritas más arriba. Además, mediante el escalamiento y posicionamiento del vídeo según cambios en la profundidad del punto representativo centrado en dicho destino de movimiento, el cálculo descrito más arriba se simplifica.

45 Con la transformación en la presente realización preferida, se supone que la forma del objeto i es un objeto en el cual una imagen se representa sobre la superficie de una hoja perpendicular a la dirección de profundidad, e incluso cuando una traslación en la dirección de profundidad o rotación alrededor del eje horizontal o eje vertical se lleva a cabo, ello corresponde al movimiento de la hoja de forma perpendicular a la dirección de profundidad.

50 Con una traslación en la dirección de profundidad, la hoja permanece perpendicular a la dirección de profundidad. En la presente realización preferida, mediante el establecimiento de un límite superior para el ángulo de rotación alrededor del eje horizontal o eje vertical, se hace que la posición de la hoja cambie a través de la rotación con un ángulo o rotación hasta un número de grados, pero se piensa la orientación de la hoja como una que permanece perpendicular a la dirección de profundidad.

Más abajo, se describe el cálculo de transformación específico.

El vídeo obtenido mediante el movimiento del objeto  $i$  representado en el vídeo  $X$  por  $r$  en la dirección de profundidad se obtiene mediante el logro de los siguientes procesos.

- (a) Solo el objeto  $i$  se selecciona del vídeo  $X$ ,
- 5 (b) una traslación se lleva a cabo a lo largo del fotograma con  $(xc(X,t,i), yc(X,t,i))$  como el origen,
- (c) el vídeo se escala a  $zc(X,t,i) / [zc(X,t,i)+r]$  veces, y
- (d) una traslación se lleva a cabo a lo largo del fotograma de modo que el origen se convierte en  $(xc(X,t,i), yc(X,t,i))$ .

Por consiguiente, cuando la transformación del movimiento del objeto  $i$  representado en el vídeo  $X$  por  $r$  en la dirección de profundidad se expresa como  $push'(i,r,X)$ , el resultado es:

$$push'(i,r,X) = move(xc(X,t,i), yc(X,t,i), scale(zc(X,t,i) / [zc(X,t,i)+r],$$

10  $move(-xc(X,t,i), -yc(X,t,i), select(i,X)))$ .

Además, cuando el objeto  $i$  representado en el vídeo  $X$  se rota en  $\theta$  alrededor del eje horizontal, las coordenadas en el vídeo del punto representativo del objeto se mueven de  $(xc(X,t,i), yc(X,t,i))$  a  $(xc(X,t,i), zc(X,t,i) \times \cos \theta - yc(X,t,i) \times \sin \theta)$ , y la profundidad se mueve de  $zc(X,t,i)$  a  $zc(X,t,i) \times \sin \theta + yc(X,t,i) \times \cos \theta$ . Es decir, estará bien lograr los procesos de más abajo.

- 15 (a) Seleccionar solo el objeto  $i$  del vídeo  $X$ ,
- (b)  $move(xc(X,t,i), yc(X,t,i))$  a  $(xc(X,t,i), xc(X,t,i) \times \cos \theta - yc(X,t,i) \times \sin \theta)$  a lo largo del fotograma, y
- (c) escalar el vídeo a  $zc(X,t,i) / [zc(X,t,i) \times \sin \theta + yc(X,t,i) \times \cos \theta]$  veces.

Por consiguiente, cuando la transformación de la rotación del objeto  $i$  representado en el vídeo  $X$  en  $\theta$  alrededor del eje horizontal se expresa como  $rothor'(i, \theta, X)$ , entonces:

$$rothor'(i, \theta, X) = scale(zc(X,t,i) / [zc(X,t,i) \times \sin \theta + yc(X,t,i) \times \cos \theta], move(0,$$

20  $zc(X,t,i) \times \cos \theta - yc(X,t,i) \times \sin \theta - yc(X,t,i), select(i,X)))$ .

Una transformación  $rotver'(i, \theta, X)$  para rotar el objeto  $i$  representado en el vídeo  $X$  en  $\phi$  alrededor del eje vertical se define de manera similar. Es decir,

$$rothor'(i, \phi, X) = scale(zc(X,t,i) / [zc(X,t,i) \times \sin \theta + xc(X,t,i) \times \cos \theta], move$$

( $zc(X,t,i) \times \cos \theta - xc(X,t,i) \times \sin \theta - xc(X,t,i), 0, select(i,X)))$ .

- 25 Con la presente realización preferida, una transformación de un solo objeto representado en el vídeo se expresa a través de una combinación de los cálculos *move*, *select* y *scale* que tienen una ligera carga de cálculo, de modo que es posible acelerar el procesamiento.

Séptima realización preferida

- 30 Con la realización preferida descrita más arriba, el tiempo del vídeo  $X$  ha experimentado la transformación de desplazamiento  $shift(d,X)$  por  $d$ , pero es posible concebir una transformación en la cual el vídeo  $X$  se reproduce lentamente o se adelanta rápidamente.

Si el vídeo  $Y$  es el vídeo  $X$  avanzado rápidamente por  $a$  veces, entonces:

$$Y(t)[x,y] = X(t/a)[x,y]$$

se establece. Más abajo, la presente transformación se expresa como  $fastforward(a,X)$ .

- 35 Dado que el objeto  $i$  representado en el vídeo  $X$  y el objeto  $j$  representado en el vídeo  $Y$  repiten, ambos, una acción dada, cuando el deseo es componer un vídeo  $Z$  de modo que el objeto  $i$  y objeto  $j$  aparecen, de forma alternativa, en una ubicación designada dentro del vídeo, es posible usar una combinación de desplazamiento y avance rápido.

5 Por ejemplo, es preciso considerar el caso en el cual el deseo es formar un fotomontaje del vídeo X, en el cual una persona (objeto i) que está realizando la acción de moler una torta de arroz en un mortero con una mano de mortero de madera, y el vídeo Y, en el cual una persona (objeto j) está realizando la acción de amasar con la mano la torta de arroz en el mismo mortero, y generar un vídeo que expresa el estado de la persona A y la persona B haciendo conjuntamente arroz.

Cuando la aparición en y alrededor del mortero, y la posición y orientación de la cámara que hizo la filmación, son comunes en el vídeo X y vídeo Y, estará bien encontrar parámetros para desplazar y avanzar rápidamente para minimizar la función de evaluación que expresa la cantidad de colisión.

10 Además, cuando el deseo es determinar a partir del vídeo si es posible o no posicionar muebles y productos electrónicos dentro de una casa, es posible aplicar el presente tipo de transformación.

Por ejemplo, cuando el deseo es determinar a partir del vídeo si es posible o no posicionar una máquina lavadora cuya puerta se abre y cierra dentro de una sala de lavado del hogar, estará bien lograr el siguiente tipo de proceso.

15 Es decir, un fabricante de productos eléctricos o un vendedor de productos eléctricos proveen en la Web un vídeo X en el cual la situación de la puerta de la máquina lavadora que se abre y cierra de forma periódica se filma desde la parte frontal.

Por otro lado, un usuario que considera la compra de la máquina lavadora prepara un vídeo Y en el cual la situación de una puerta con respecto a la sala en la cual la máquina lavadora se instalará abriéndose y cerrándose se filma desde la parte frontal.

20 Después de lograr la normalización de modo que la distancia de la cámara a la máquina lavadora en el vídeo X y la distancia de la cámara a la ubicación candidata para la instalación de máquina lavadora en el vídeo Y sustancialmente concuerdan, el usuario arrastra y suelta el objeto de la máquina lavadora en el vídeo X en la ubicación candidata de instalación en el vídeo Y.

25 Además, se descubre la cantidad de colisión de los objetos en los vídeos X e Y. Si no hay colisión alguna, es posible calcular que la máquina lavadora puede instalarse en la ubicación deseada. Además, incluso cuando hay una colisión, es posible descubrir parámetros para el desplazamiento y avance rápido para minimizar la cantidad de colisión y llevar esto a cabo para que no ocurran colisiones, naturalmente es posible calcular que la máquina lavadora puede instalarse en la ubicación deseada.

30 Cuando el usuario filma el interior de la sala, la dirección de filmación es, a veces, limitada. En dicho caso, estará bien proveer vídeos en los cuales la máquina lavadora cuya puerta se abre y cierra periódicamente se filma desde una variedad de direcciones como, por ejemplo, situaciones en las cuales la filmación es desde la parte frontal y también situaciones en las cuales la filmación es desde el lado izquierdo y situaciones en las cuales la filmación es desde el lado derecho, y para permitir al usuario seleccionar la filmación del vídeo desde la misma dirección que la dirección en la cual el usuario filma la sala.

Octava realización preferida

35 En las realizaciones preferidas descritas más arriba, es posible componer un vídeo de modo que las colisiones en las cuales los objetos se penetran mutuamente no ocurren, y es también posible formar un fotomontaje de vídeos de modo que las superficies de objetos se tocan en un instante deseado o la parte inferior de un objeto toca el suelo. Más abajo, se describen aplicaciones específicas de las realizaciones preferidas descritas más arriba.

40 El primer ejemplo de aplicación es para permitir a los expositores preparar fácilmente vídeos que muestran la situación de productos en mercados electrónicos como, por ejemplo, subastas en Internet y/o similares.

45 Es decir, el operador de un mercado electrónico provee materiales de vídeo de modo que los usuarios pueden usarlos libremente, además de proveer herramientas de edición para varios tipos de vídeos. Dicho material de vídeo puede prepararse por expositores u oferentes en el mercado electrónico, o un sistema de distribución de vídeo pueden prepararse de forma separada de modo que los vídeos públicamente utilizables se cargan y pueden leerse detenidamente por otros usuarios, y los vídeos acumulados en el presente sistema de distribución de vídeo pueden volver a usarse.

50 Estará bien que los expositores filmen con antelación vídeos que expresan fondos como, por ejemplo, en salas de la casa de un expositor y/o similares, y cuando se hace esto, preparar una mesa decorada o pantalla plegable y/o similares de modo que el producto puede presentarse de manera agradable, y también estará bien que los productos en la sala se mantengan prolijos y en orden, y limpios, de modo que la información privada no pueda entenderse. Además, el vídeo de fondo puede seleccionarse del material de vídeo.

Luego, cuando el expositor quiere exhibir un nuevo producto, es posible lograr la filmación en una ubicación preferida dentro de la sala sin tener que ordenar y organizar productos o hacer una limpieza.

5 Luego, el expositor arrastra y suelta el vídeo del nuevo producto que se filmó en el vídeo de fondo filmado con antelación. Luego, es posible componer un vídeo que muestra una situación en la cual el producto no colisiona con productos (la pantalla plegable y/o similares) de la imagen de fondo y toca el suelo (la superficie de la mesa decorada y/o similares) de la imagen de fondo, a través de la evitación de colisión descrita más arriba y el ajuste automático de posición.

El fotomontaje del vídeo de producto en el vídeo de fondo puede llevarse a cabo por oferentes así como por expositores. Por ejemplo, si se exhibe una vestido de fiesta, un vídeo de producto puede proveerse, el cual muestra una situación en la cual la modelo del expositor y/o similares está vistiendo dicho vestido, caminando y dándose vuelta.

10 El oferente puede filmar él mismo un vídeo de fondo filmando el interior de un salón de actos donde una fiesta a la que quieren asistir vistiendo el vestido se está celebrando, o puede seleccionar de entre los materiales de vídeo. Además, el oferente forma un fotomontaje de un vídeo de producto del vestido que usa la modelo, y el vídeo de fondo del vestíbulo de la fiesta.

15 Mediante la aplicación de la realización preferida descrita más arriba, la modelo se mueve sobre la parte superior del suelo, pero el vídeo se compone de modo que no hay colisiones con otras personas, muebles o instalaciones dentro del vestíbulo.

Mediante la lectura detenida del presente vídeo, el oferente puede comprobar con antelación si el vestido exhibido como un producto concuerda o no con la atmósfera del vestíbulo de la fiesta, con anterioridad a las ofertas.

Un segundo ejemplo de aplicación además utiliza los materiales de vídeo.

20 Por ejemplo, si un vídeo de fondo que muestra una situación en la cual un artista famoso está bailando música se provee como material de vídeo, un usuario puede componer un vídeo que muestra una situación en la cual el usuario está bailando junto con el cantante famoso, mediante la filmación y el fotomontaje de un vídeo del usuario bailando.

25 Además, cuando un vídeo de fondo que muestra una situación en la cual un carrusel se está moviendo con asientos vacíos se provee como material de vídeo, es posible componer un vídeo que muestra una situación en la cual el usuario está montado en un carrusel, mediante la filmación de un vídeo del usuario montando una bicicleta, un carro y/o similares u otro vehículo moviéndose a lo largo de sustancialmente la misma trayectoria que el carrusel y logrando el fotomontaje del vídeo mientras impone la condición restrictiva de que "el usuario está tocando continuamente el carrusel".

30 Además, cuando un usuario forma un fotomontaje de un vídeo moviendo su brazo o una raqueta forzosamente hacia abajo con un vídeo de fondo en el cual una bola ferozmente golpea el suelo, si se impone la condición restrictiva de que la mano del usuario o raqueta y la bola hagan contacto en el instante en el que el usuario mueve su brazo hacia abajo, es posible generar un vídeo que muestra al usuario rematando un balón de vóleybol o una pelota de tenis.

35 Además, múltiples materiales de vídeo pueden componerse filmando a una sola persona practicando esgrima o kendo, y si se impone la condición que los objetos tengan contacto con sincronización en el instante en el que el jugador que se ha filmado lleva a cabo un movimiento decisivo, un vídeo puede componerse que parece como si un partido estuviera en progreso.

Descripción del ejemplo

40 Aquí se provee un método a modo de ejemplo. En la medida en que el presente ejemplo no implica el uso de un corrector para corregir el segundo vídeo, este no cae dentro del alcance de la invención reivindicada, y se provee a los fines ilustrativos solamente.

El presente ejemplo concibe una interfaz de usuario cuando superpone mediante desplazamiento el segundo vídeo Y al primer vídeo X a través de una función de arrastrar y soltar y/o similares mediante el uso de un ratón.

45 Con el presente ejemplo, con el fin de facilitar la comprensión, se supondrá un caso en el cual a través de la función de arrastrar y soltar del usuario, una cantidad de traslación hacia arriba, abajo, izquierda o derecha se especifica con el fin de desplazar y superponer el segundo vídeo Y al primer vídeo X. Sin embargo, el sistema de funcionamiento del presente ejemplo no se encuentra necesariamente limitado a arrastrar y soltar, dado que es también posible reemplazar el arrastrar y soltar a través de la utilización de varias funciones de entrada especificadas como, por ejemplo, una acción de deslizamiento en una pantalla táctil, o repetir dos veces una función de clic, y/o similares.

50 La Figura 5A es un dibujo explicativo que muestra el estado de la pantalla antes del inicio del arrastre, la Figura 5B es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo en el que el arrastre se inicia, la Figura 5C es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo durante el

arrastre y la Figura 5D es un dibujo explicativo que muestra el estado de la pantalla en un punto en el tiempo en el que la acción de soltar se lleva a cabo.

5 Además, la Figura 6 es un dibujo explicativo que muestra el fotomontaje a modo de resumen de partes del dispositivo de procesamiento de vídeo 101 relacionado con el proceso de arrastrar y soltar. La Figura 7 es un diagrama de flujo que muestra el flujo de control del proceso de arrastrar y soltar ejecutado por el dispositivo de procesamiento de vídeo 101 según el presente ejemplo. La explicación de más abajo hace referencia a dichos dibujos.

10 En el proceso de arrastrar y soltar según el presente ejemplo, en el dispositivo de procesamiento de vídeo 101, una configuración de cantidad de deslizamiento 104 y un generador de vídeo 105 actúan según la función de arrastrar y soltar del usuario.

El dispositivo de procesamiento de vídeo 101, cuando se desliza y superpone el segundo vídeo Y al primer vídeo X, se provee con un determinador de interferencia para determinar si las condiciones de posición se satisfacen o no de modo que la relación posicional de los objetos satisface las condiciones de posición preestablecidas.

15 Sin embargo, en el presente ejemplo, con el fin de utilizar una colisión de objetos que no ocurre como una condición de posición, el determinador de colisión 103 descrito más arriba se utiliza como el determinador de interferencia, y el resultado de determinación de aquel se usa. En el presente dibujo, el adquirente de profundidad de superficie posterior 102 en frente del determinador de colisión 103 se omite del dibujo.

20 Según se describe más abajo, como una condición de posición, además de los objetos que no interfieren, es decir, las condiciones de interferencia no se satisfacen, es posible establecer otras condiciones, por ejemplo, que los objetos preestablecidos sean inmediatamente adyacentes, y/o similares. Para los resultados de determinación del determinador de colisión 103 descrito más arriba, es posible también usar una determinación de condiciones de posición con referencia a las relaciones de posición diferentes de la interferencia de los objetos.

Además, un corrector 106 que se muestra en el dibujo es un elemento que puede omitirse, y el siguiente ejemplo explica las funciones de aquel.

25 Primero, el dispositivo de procesamiento de vídeo 101 muestra en la pantalla un fotograma representativo del primer vídeo X y un fotograma representativo del segundo vídeo Y (etapa E181). Por ejemplo, como se muestra en la Figura 5A, una ventana 411 del primer vídeo X, una ventana 412 del segundo vídeo Y, y un cursor 413 se muestran en una pantalla 401.

30 En el presente dibujo, el primer vídeo X es un vídeo con un fondo y el borde de la pantalla 411 se representa por una línea continua. Por otro lado, el segundo vídeo Y es un vídeo sin fondo. La periferia de la ventana 412 se representa con un color transparente y el borde de aquella se representa por una línea punteada. El borde de la ventana 412 no necesita mostrarse necesariamente en la pantalla 401. En el presente caso, el usuario ve una visualización en la pantalla 401 como si el objeto solo sin fondo se representara en el segundo vídeo Y.

35 A continuación, el dispositivo de procesamiento de vídeo 101 recibe una acción de clic (inicio del arrastre) en la ventana 412 del segundo vídeo Y del usuario (etapa E182). Aquí, como se muestra en la Figura 5B, el usuario ha usado el ratón para hacer que el cursor 413 se mueva y ha hecho clic en una posición [s,t] dentro del segundo vídeo Y.

40 Además, el dispositivo de procesamiento de vídeo mueve la posición de visualización de la ventana 412 del segundo vídeo Y según esto (etapa E184) mientras el arrastre por el usuario continúa (etapa E183: Sí). Por ejemplo, cuando el usuario hace que el cursor 413 se mueva mientras mantiene el clic, como se muestra en la Figura 5C, hace que la ventana 412 en su conjunto se mueva de modo que el cursor 413 indica la posición [s,t] dentro del segundo vídeo Y.

Cuando el usuario ha hecho que el cursor 413 se mueva tan lejos como una posición [u,v] dentro del primer vídeo X, como se muestra en la Figura 5D, y detiene el clic, la ventana 412 se suelta en la ventana 411.

45 Luego, la posición [u,v] dentro del primer vídeo X y la posición [s,t] dentro del segundo vídeo Y se superponen. Por consiguiente, puede pensarse que el usuario ha superpuesto el segundo vídeo Y al primer vídeo X mediante una transformación haciendo que el segundo vídeo Y se mueva por u-s en la dirección horizontal y por v-t en la dirección vertical.

50 De esta manera, cuando el arrastre por el usuario finaliza (etapa E183: No), el establecedor de cantidad de desplazamiento 104 del dispositivo de procesamiento de vídeo 101 establece la dirección horizontal u-s y la dirección vertical v-t como la cantidad de desplazamiento (etapa E185). Es decir, el dispositivo de procesamiento de vídeo 101 interpreta el cambio como uno en el que el usuario ha superpuesto el vídeo move(u-s, v-t, Y) obtenido moviendo el segundo vídeo hacia arriba, abajo, izquierda o derecha en paralelo con el fotograma, al primer vídeo X.

Por lo tanto, el dispositivo de procesamiento de vídeo 101, mediante el uso del determinador de colisión 103, determina si el objeto representado en el primer vídeo X y el objeto representado en el vídeo intermedio  $\text{move}(u-s, v-t, Y)$  realizado mediante el desplazamiento del segundo vídeo Y colisionan en cualquier tiempo del fotograma o cualquier posición en el fotograma, o si no ocurre colisión alguna (etapa E186).

- 5 Cuando el resultado de la función de soltar que se muestra en la Figura 5C es que no ocurre colisión alguna (etapa E186: No), el generador de vídeo 105 del dispositivo de procesamiento de vídeo 101 produce la superposición del tercer vídeo ( $\text{move}(u-s, v-t, Y), X$ ) a través del fotomontaje, y el proceso concluye.

A través del presente tipo de fotomontaje, es posible que el usuario obtenga un tercer vídeo en el cual los objetos representados en el vídeo Y se posicionan para no colisionar con objetos en el vídeo X. En el tercer vídeo, dado que no ocurren colisiones con objetos representados en el vídeo Y, es posible prevenir representaciones no naturales como, por ejemplo, objetos que se comen entre sí.

Por otro lado, cuando ocurre una colisión incluso cuando la función de soltar que se muestra en la Figura 5C se lleva a cabo (etapa E186: Sí), el dispositivo de procesamiento de vídeo 101 produce una alerta como, por ejemplo, audio, vibración, una visualización de pantalla y/o similares (etapa E188), el proceso concluye y no se produce un nuevo vídeo. Es decir, si los comandos del usuario son apropiados, el tercer vídeo se genera, pero si los comandos son inapropiados, solo se genera una alerta y el tercer vídeo no se genera.

Las varias transformaciones descritas más arriba se logran en unidades de fotogramas. Por consiguiente, cuando una transformación se asume en un vídeo dado para obtener un vídeo diferente, después de los procesos de transformación logrados como, por ejemplo, mover y/o similares en cada fotograma, estará bien ejecutar un proceso de cálculo diferente como, por ejemplo, una determinación de colisión y/o similares, y estará bien utilizar una evaluación de extensión dirigida por demanda para lograr los procesos de transformación.

Si los procesos de transformación se logran como dirigidos por demanda, incluidos los ejemplos de aquí en adelante, es necesario lograr una determinación de superposición y/o similares para un fotograma dado, y la transformación, inicialmente, se logra para dicho fotograma. Por consiguiente, con el fin de obtener una conclusión de que no ocurren colisiones, una transformación para todos los fotogramas es necesaria, pero cuando una colisión se identifica en un fotograma dentro del vídeo, el proceso de transformación no se ejecuta para fotogramas subsiguientes.

Además, con el método para especificar la cantidad de desplazamiento a través de arrastrar y soltar, la cantidad de desplazamiento se expresa por cuánto respecto a la posición de la ventana 411 del primer vídeo X la posición de la ventana 412 que se ha soltado del segundo vídeo Y se ha desplazado. Sin embargo, dado que la cantidad de desplazamiento es una cantidad aplicada al segundo vídeo Y, cuando se ingresa la cantidad de desplazamiento, una comparación con el primer vídeo X no se requiere. Por consiguiente, estará bien que el usuario use el teclado y directamente ingrese un valor numérico, o que use un método para ingresar la cantidad de desplazamiento visualmente mediante el uso de un ratón y/o similares.

35 Más abajo, se describe un método para lograr el ajuste automático en lugar de simplemente emitir una alerta cuando las instrucciones de usuario son inapropiadas.

#### Aplicabilidad industrial

Con la presente descripción, es posible proveer un dispositivo de procesamiento de vídeo, método de procesamiento de vídeo y un medio de grabación de información no transitorio para asegurar que los objetos no satisfacen condiciones de interferencia cuando los objetos acompañados de información de profundidad y representados en múltiples vídeos se recopilan en un vídeo.

#### Lista de signos de referencia

- 11 Punto de disparo
- 12 Superficie de proyección
- 45 13 Media línea
- 14 Objeto
- 15 Punto de colisión
- 17 Distancia z
- 21 Punto representativo
- 50 101 Dispositivo de procesamiento de vídeo

102 Adquirente de profundidad de superficie posterior

103 Determinador de colisión

104 Establecedor de cantidad de desplazamiento

105 Generador de vídeo

5 106 Corrector

401 Pantalla

411 Ventana

412 Ventana

413 Cursor

10

**REIVINDICACIONES**

1. Un dispositivo de procesamiento de vídeo (101) para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto (14) se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto (14), y un segundo objeto (14) se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto (14), el dispositivo de procesamiento de vídeo comprendiendo:
- 5 un generador de vídeo (105) para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y el segundo vídeo 2.5D;
- 10 un determinador de interferencia (103) para determinar si el primer objeto (14) y el segundo objeto (14) satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto (14) y a la información de profundidad del segundo objeto (14); y
- 15 un corrector (106) para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación con desplazamiento en el tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.
2. El dispositivo de procesamiento de vídeo (101) según la reivindicación 1, en donde:
- la longitud de tiempo del desplazamiento en el tiempo, cantidad de traslación, cantidad de rotación, o relación de escalamiento, o una combinación de estas, se usa como un parámetro de transformación;
- 20 el dispositivo de procesamiento de vídeo (101) además comprende un establecedor de valor inicial (104) para establecer un valor inicial del parámetro de transformación; y
- el corrector (106) aplica la transformación no deformante según el parámetro de transformación, actualiza el parámetro de transformación de modo que la condición de interferencia no se satisface, y aplica la transformación no deformante según el parámetro de transformación actualizado.
- 25 3. El dispositivo de procesamiento de vídeo (101) según la reivindicación 1 o reivindicación 2, en donde la condición de interferencia se satisface si el primer objeto (14) y el segundo objeto (14) colisionan.
4. El dispositivo de procesamiento de vídeo (101) según la reivindicación 2, en donde el corrector (106) repite la actualización del parámetro de transformación hasta la primera vez en que se determina que el primer objeto (14) y el segundo objeto (14) no colisionan.
- 30 5. El dispositivo de procesamiento de vídeo (101) según la reivindicación 2, en donde el corrector (106) repite la actualización del parámetro de transformación hasta la primera vez en que se determina que el primer objeto (14) y el segundo objeto (14) no colisionan y se determina que si la siguiente actualización se lleva a cabo, el primer objeto (14) y el segundo objeto (14) colisionarán.
- 35 6. El dispositivo de procesamiento de vídeo (101) según la reivindicación 4 o reivindicación 5, en donde el corrector (106) actualiza el parámetro de transformación dentro de un rango de corrección preestablecido de modo que el tiempo transcurrido hasta que la condición de interferencia se satisfaga por primera vez aumenta.
7. El dispositivo de procesamiento de vídeo (101) según la reivindicación 4 o reivindicación 5, en donde:
- el determinador de interferencia (103) calcula una cantidad de colisión que indica el grado de colisión entre el objeto transformado (14) y el primer objeto (14); y
- 40 el corrector (106) actualiza el parámetro de transformación dentro de un rango preestablecido de modo que la cantidad de colisión disminuye.
8. El dispositivo de procesamiento de vídeo (101) según la reivindicación 7, en donde la cantidad de colisión es una suma estimada de los volúmenes en los cuales una forma del primer objeto (14) estimada según la información de profundidad del primer objeto (14) en cada fotograma y una forma del objeto transformado (14) estimada según la información de profundidad del objeto transformado (14) determinado según la información del segundo objeto (14) se superponen.
- 45 9. El dispositivo de procesamiento de vídeo (101) según la reivindicación 7, en donde la cantidad de colisión es la suma de las áreas en las cuales un área en la cual el primer objeto (14) debe representarse en cada fotograma y un área en la cual el objeto transformado (14) debe representarse se superponen.

10. El dispositivo de procesamiento de vídeo (101) según cualquiera de las reivindicaciones 1 a 9, en donde el generador de vídeo (105) obtiene una posición tridimensional donde un punto representativo del segundo objeto se posiciona, desde una posición dentro de un fotograma donde un punto representativo se representa en el fotograma y una posición de profundidad de un punto representativo determinado a partir de la información de profundidad del segundo objeto, y obtiene una posición transformada mediante la aplicación de una transformación congruente de espacio tridimensional en la posición tridimensional, y obtiene una posición de dibujo donde la posición transformada debe dibujarse dentro del fotograma, y obtiene el vídeo 2.5D transformado mediante la implementación de una transformación de traslación según una diferencia entre la posición dentro del fotograma y la posición de dibujo obtenida, y la aplicación de una transformación de escalamiento según una relación de la posición de profundidad y la profundidad de la posición transformada.

11. Un método de procesamiento de vídeo para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto (14) se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto (14), y un segundo objeto se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto (14), el método de procesamiento de vídeo comprendiendo:

un procedimiento de generación de vídeo para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y segundo vídeo 2.5D;

un procedimiento de determinación de interferencia para determinar si el primer objeto y el segundo objeto satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto (14) e información de profundidad del segundo objeto (14); y

un procedimiento de corrección para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación con desplazamiento en el tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.

12. Un programa para formar un fotomontaje de un primer vídeo 2.5D que comprende múltiples fotogramas de imágenes y un segundo vídeo 2.5D que comprende múltiples fotogramas de imágenes, en donde un primer objeto (14) se representa en el primer vídeo 2.5D y se acompaña de información de profundidad del primer objeto (14), y un segundo objeto (14) se representa en el segundo vídeo 2.5D y se acompaña de información de profundidad del segundo objeto (14), el programa haciendo que un ordenador funcione como:

un generador de vídeo (105) para generar un tercer vídeo 2.5D que comprende múltiples fotogramas de imágenes mediante la formación de un fotomontaje del primer vídeo 2.5D y el segundo vídeo 2.5D;

un determinador de interferencia (103) para determinar si el primer objeto (14) y el segundo objeto (14) satisfacen o no una condición de interferencia en el tercer vídeo 2.5D, mediante referencia a la información de profundidad del primer objeto (14) y a la información de profundidad del segundo objeto (14); y

un corrector (106) para corregir el segundo vídeo 2.5D mediante la aplicación de una transformación no deformante que mantiene la forma de una trayectoria que expresa movimiento con el paso del tiempo, de modo que la condición de interferencia no se satisface, cuando se determina que la condición de interferencia se satisface, en donde la transformación no deformante es una o una combinación de: una transformación con desplazamiento en el tiempo, una transformación de traslación, una transformación de rotación y una transformación de escalamiento.

FIG. 1

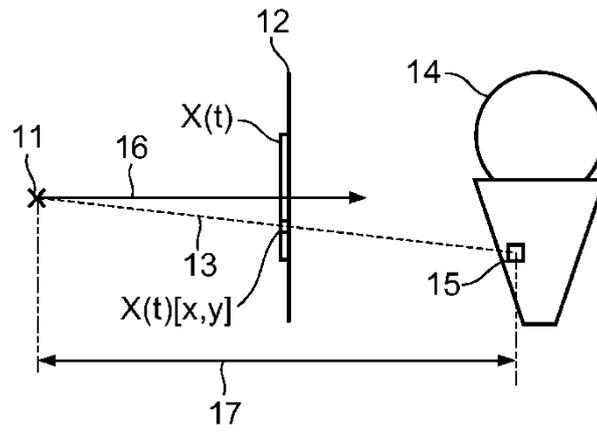


FIG. 2A

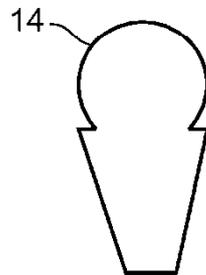


FIG. 2B

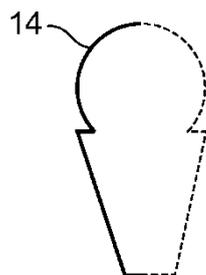


FIG. 2C

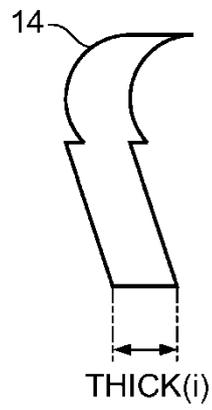


FIG. 2D

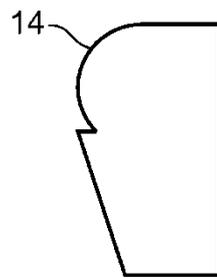


FIG. 2E

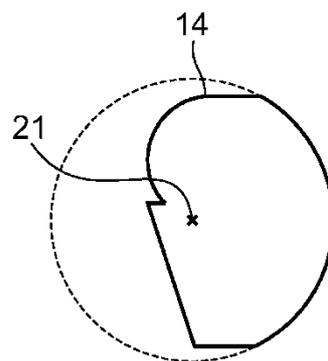


FIG. 3

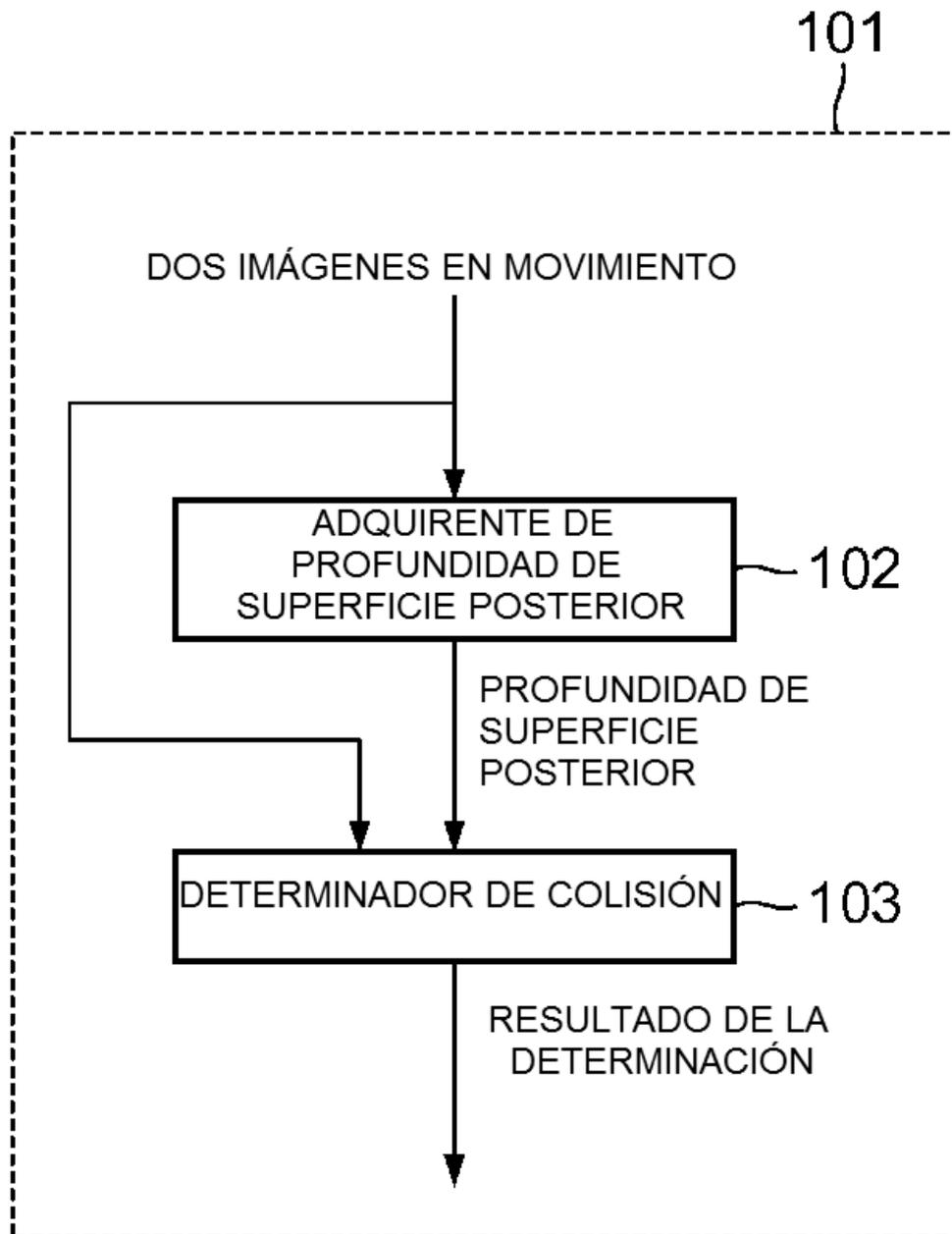


FIG. 4

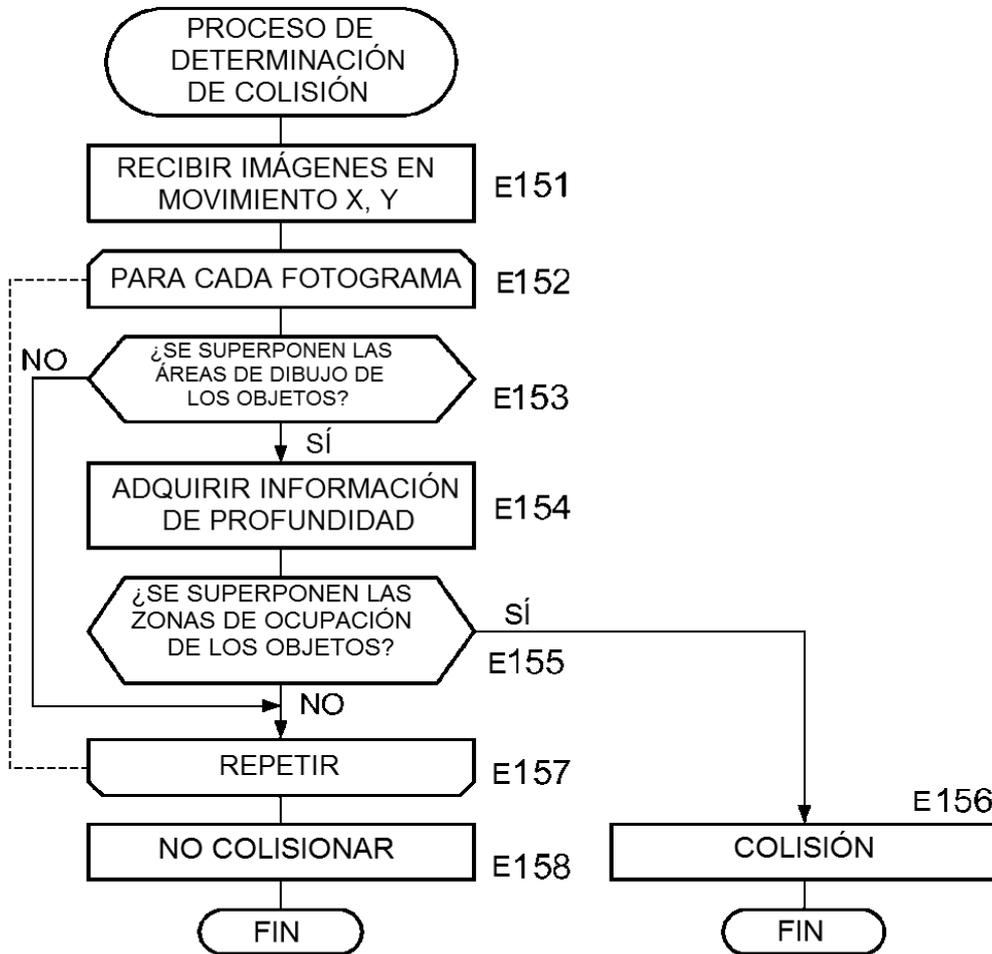


FIG. 5A

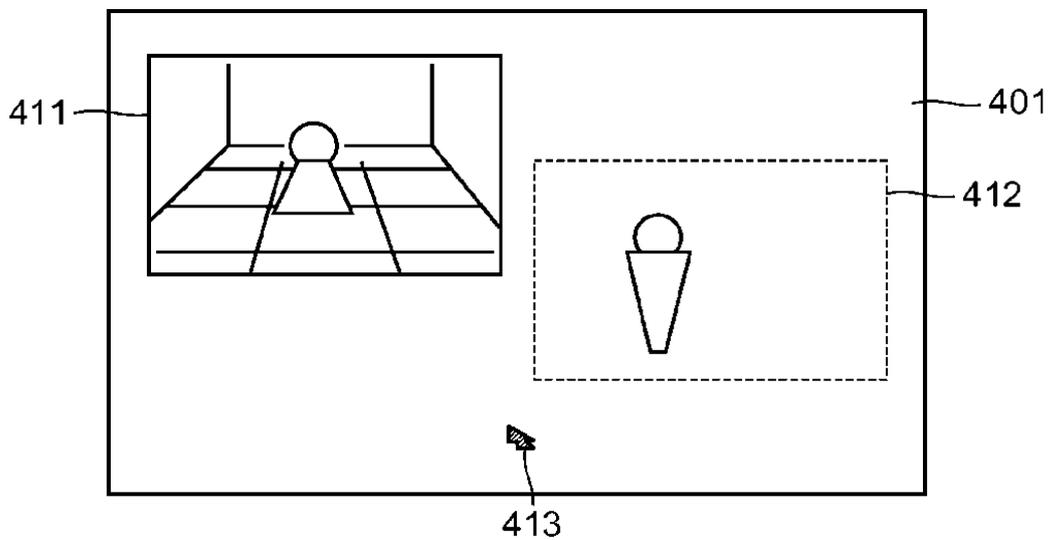


FIG. 5B

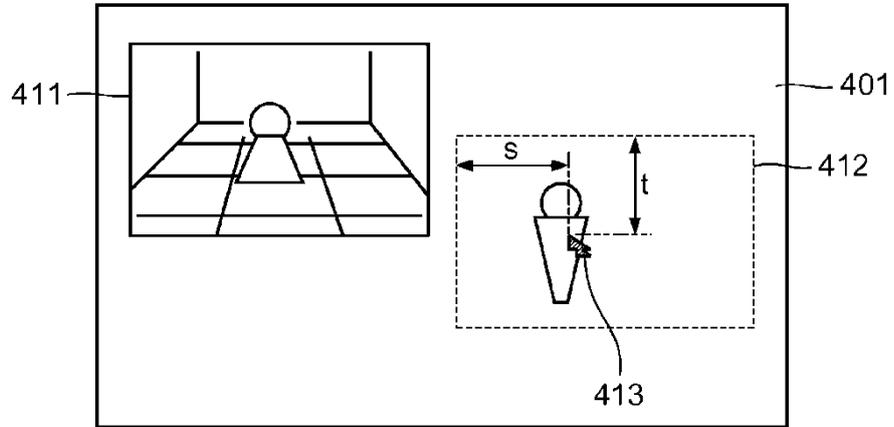


FIG. 5C

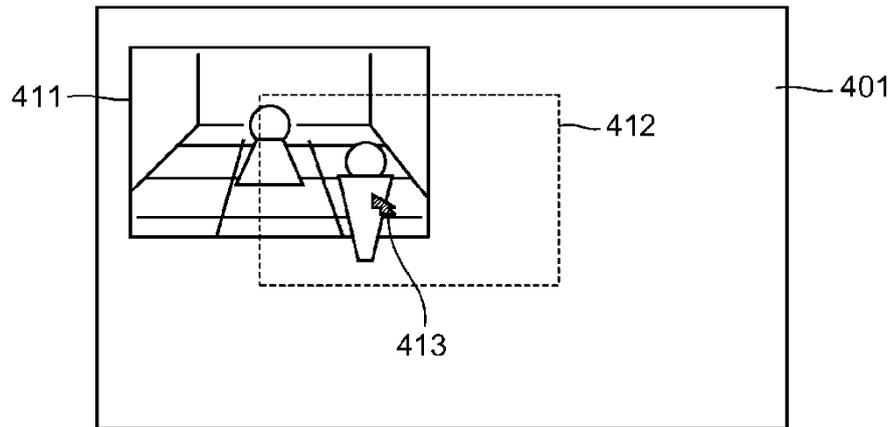


FIG. 5D

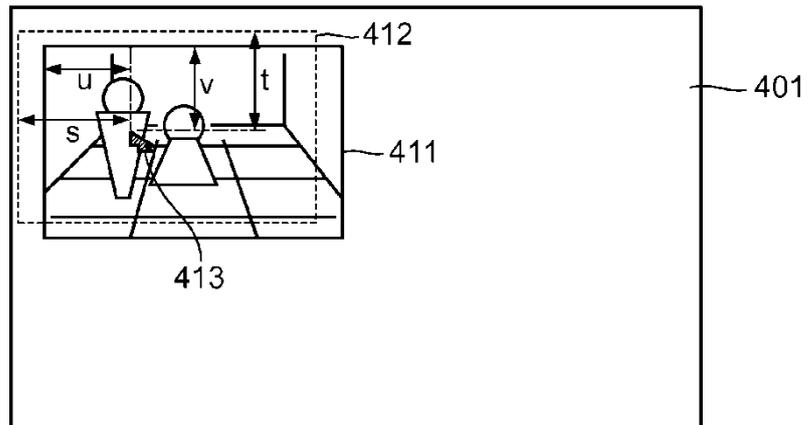




FIG. 7

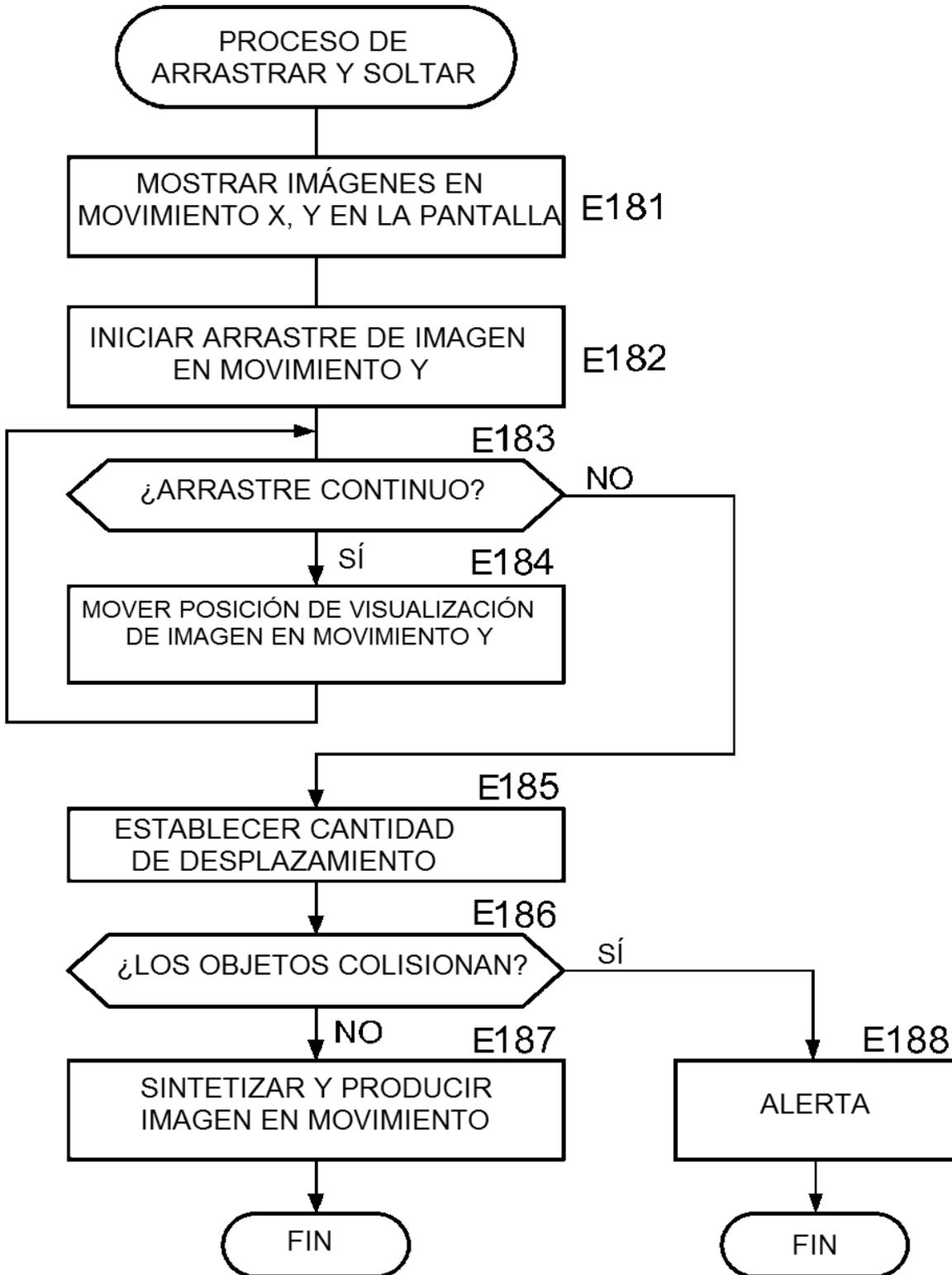


FIG. 8

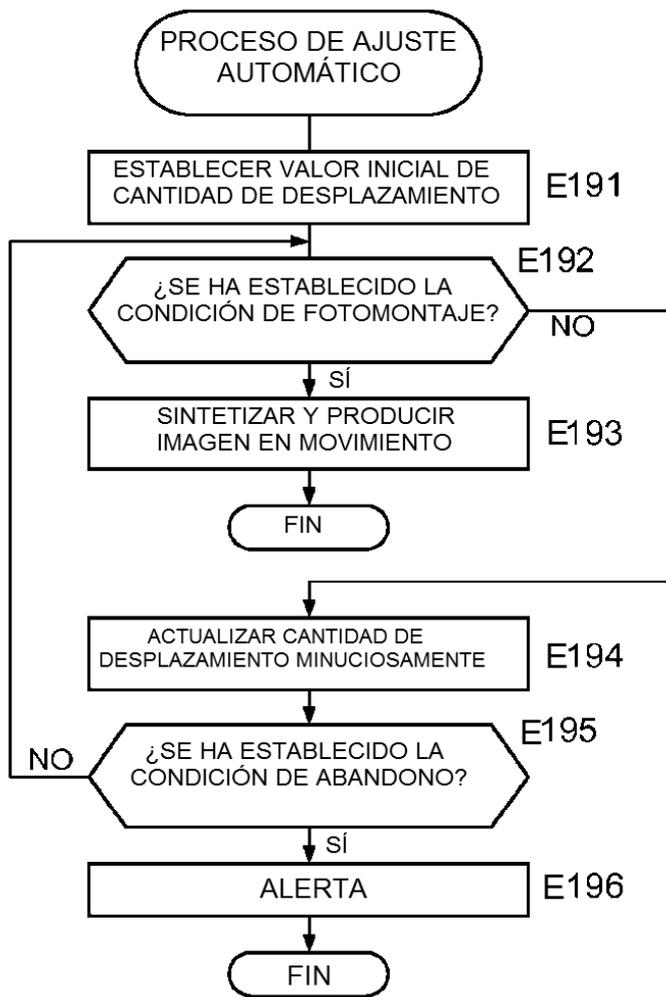


FIG. 9A

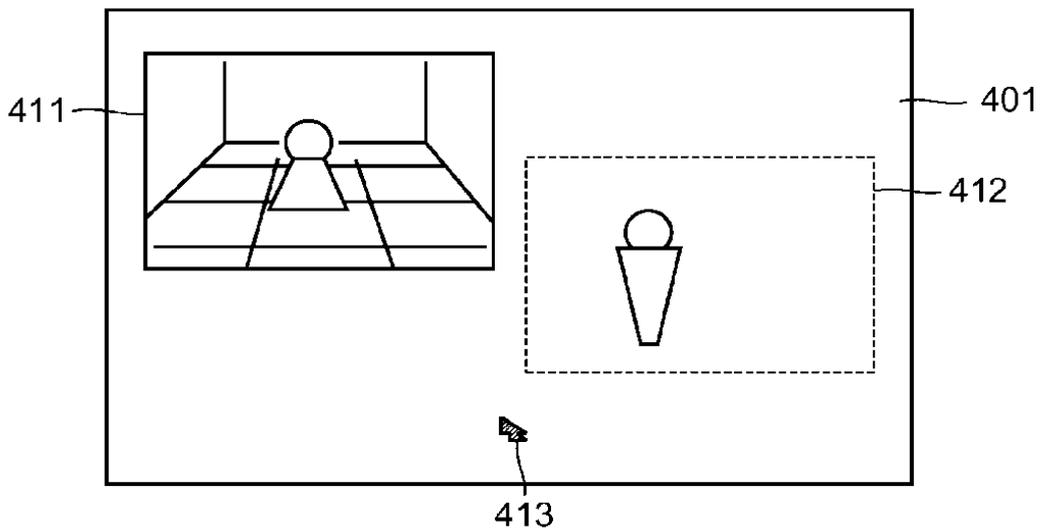


FIG. 9B

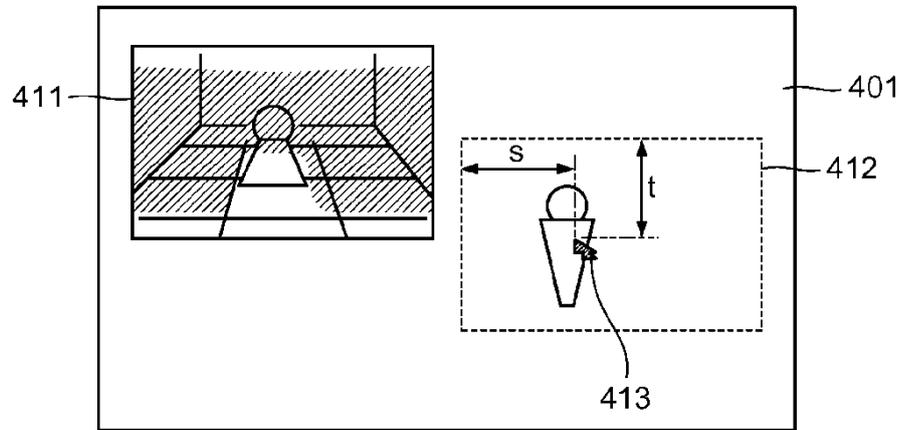


FIG. 9C

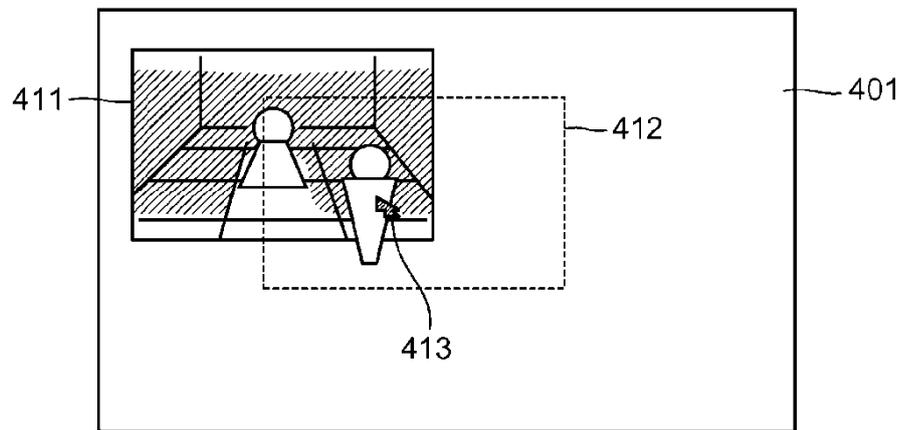


FIG. 9D

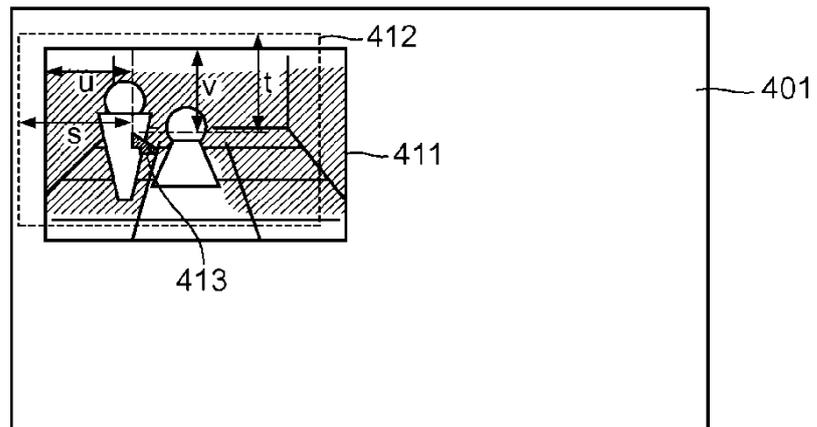


FIG. 9E

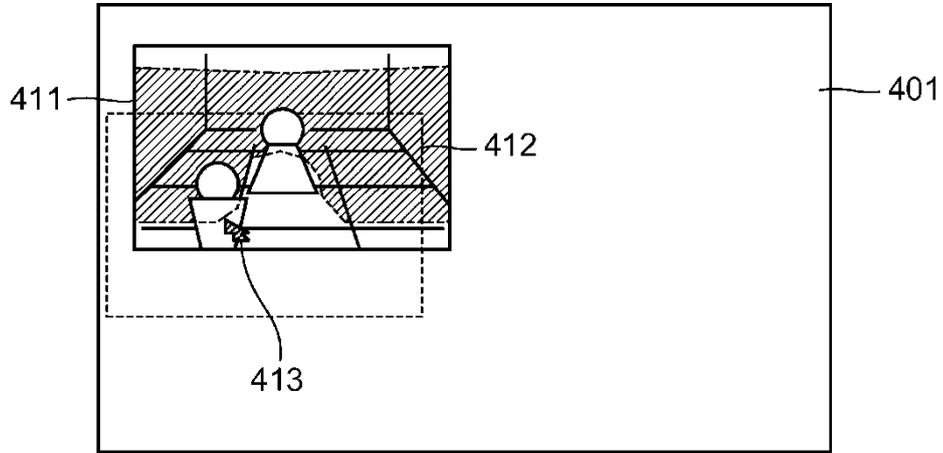


FIG. 10

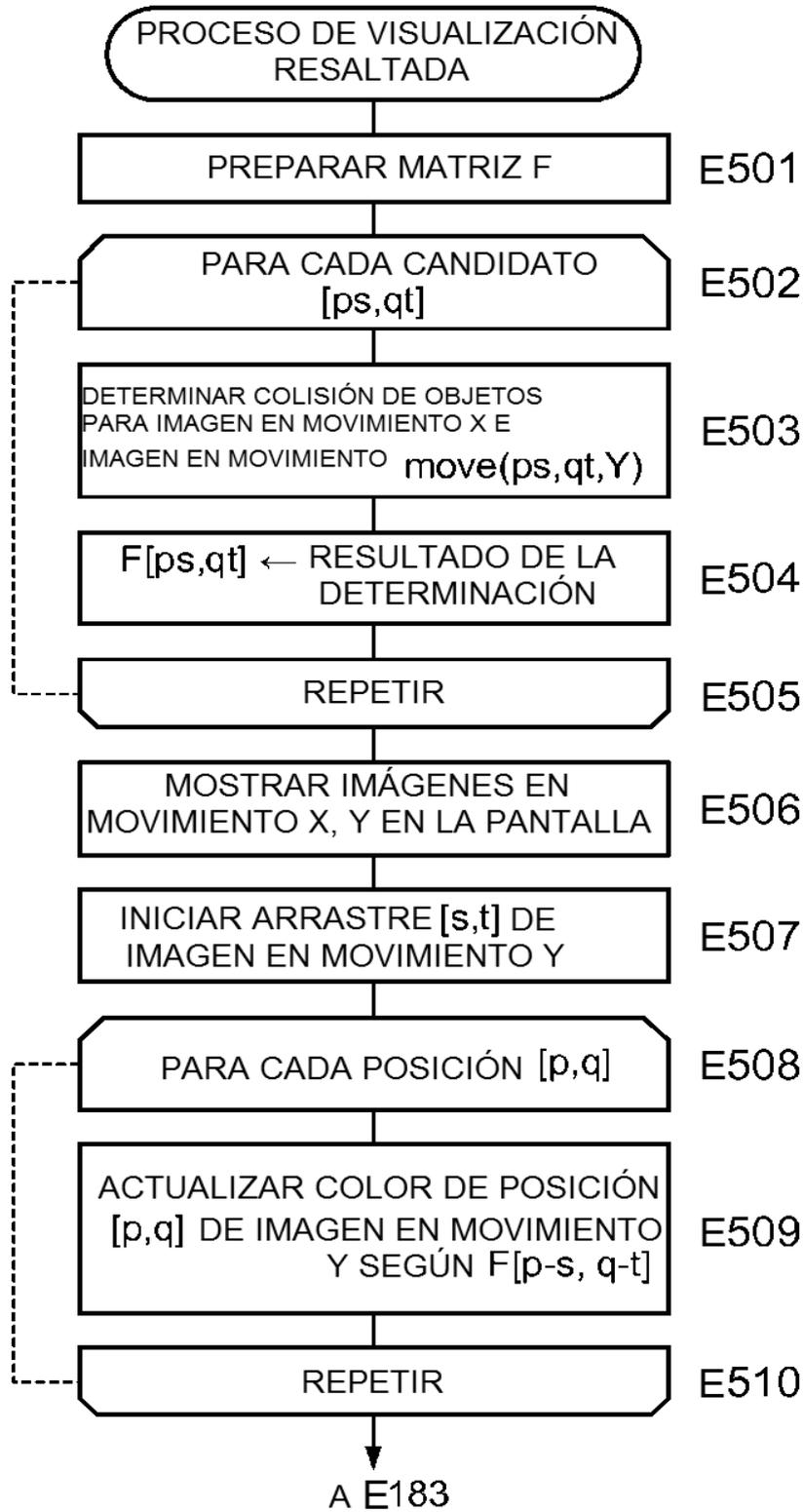


FIG. 11A

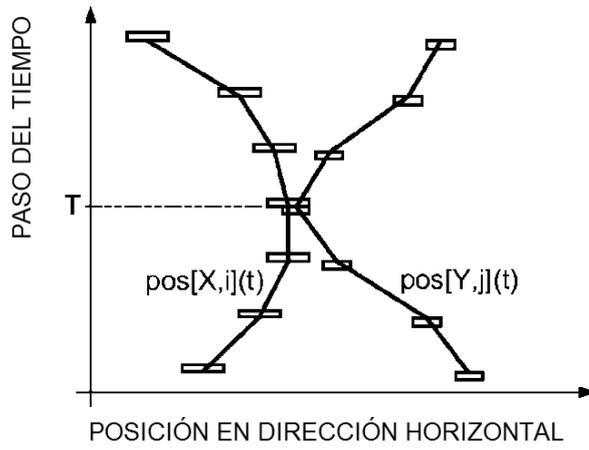


FIG. 11B

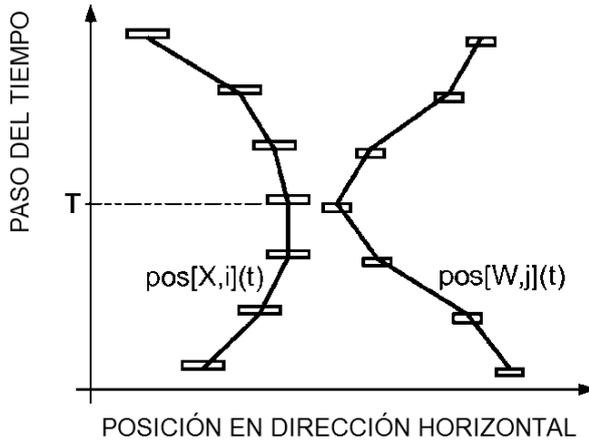


FIG. 12A

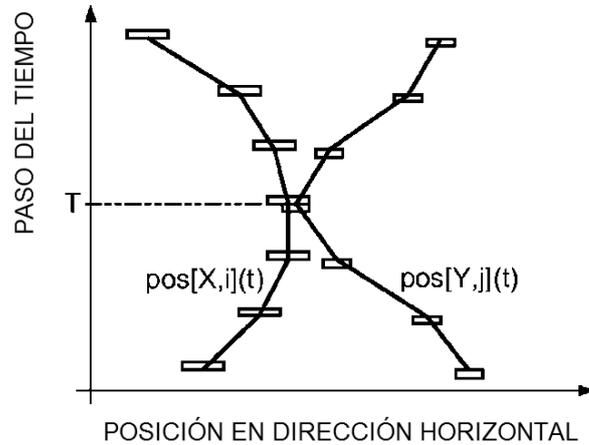


FIG. 12B

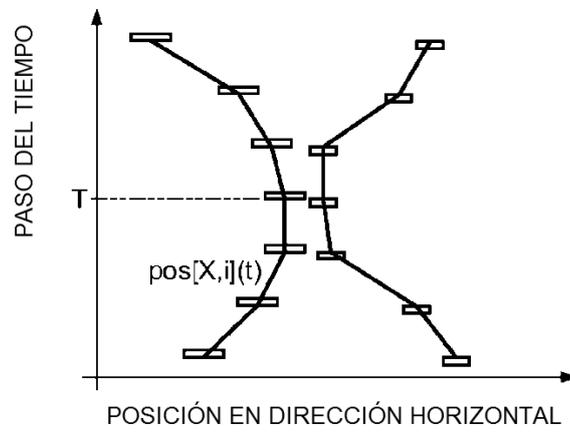


FIG. 12C

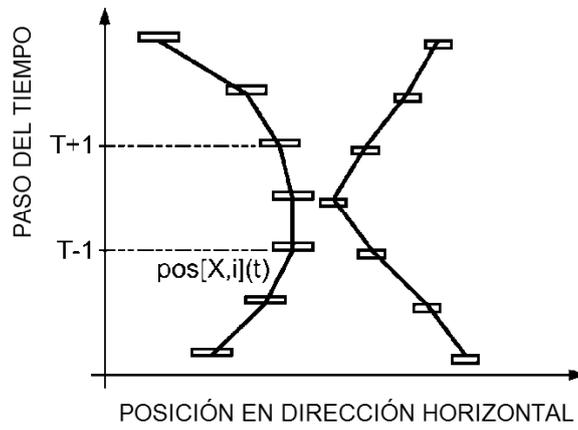


FIG. 12D

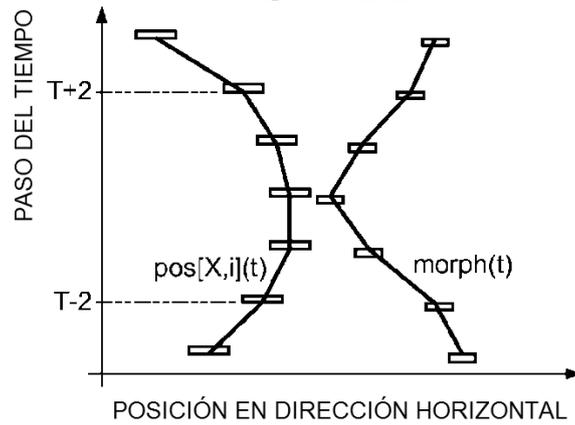


FIG. 13

