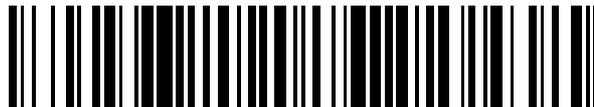


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 618 075**

51 Int. Cl.:

**G06F 17/30** (2006.01)

**G06K 9/00** (2006.01)

**G06T 7/20** (2006.01)

**H04N 5/76** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.10.2012 PCT/JP2012/076480**

87 Fecha y número de publicación internacional: **25.07.2013 WO2013108448**

96 Fecha de presentación y número de la solicitud europea: **12.10.2012 E 12866068 (5)**

97 Fecha y número de publicación de la concesión europea: **07.12.2016 EP 2733931**

54 Título: **Dispositivo de búsqueda de vídeo, método de búsqueda de vídeo, medio de registro y programa**

30 Prioridad:

**20.01.2012 JP 2012010492**

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

**20.06.2017**

73 Titular/es:

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

72 Inventor/es:

**HIRANO HIROMI;  
SANJO MASAHIRO;  
KAWATE YUTA;  
OKABE MAKOTO y  
ONAI RIKIO**

74 Agente/Representante:

**FÚSTER OLAGUIBEL, Gustavo Nicolás**

ES 2 618 075 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 búsqueda de vídeo, procedimiento de búsqueda de vídeo, medio de registro y programa

5 **Campo técnico**

La presente invención se refiere a un dispositivo de búsqueda de vídeo, a un procedimiento de búsqueda de vídeo, a un medio de registro y a un programa para efectuar búsquedas de vídeo, permitiendo que un usuario especifique fácilmente un objeto que debería aparecer en los vídeos, junto con la posición, la orientación y el movimiento del mismo.

**Técnica anterior**

15 Anteriormente, la SIFT (transformación de características invariantes a escala), divulgada en la bibliografía no de patente 1, y los HOG (histogramas de gradientes orientados), divulgados en la bibliografía no de patente 2, se han propuesto como técnica para reconocer objetos representados en imágenes fijas y los puntos característicos de los mismos. Aplicando esta técnica, es fácil determinar si el aspecto externo de los objetos representados en dos imágenes fijas es similar o no, incluso si se rotan los objetos, si difiere la proporción de ampliación o si difiere la luminosidad. Además, aplicando la técnica mencionada anteriormente después de invertir la imagen de objeto que muestra el objeto representado, es posible determinar que las imágenes son similares incluso cuando se representa la imagen especular del objeto.

25 Por otra parte, se propone el vídeo en partículas en las bibliografías no de patente 3 y 4 como técnica para rastrear el movimiento de un objeto en un vídeo. Aplicando esta técnica, es posible adquirir la posición en la que se representa una ubicación dada en la superficie de un objeto en el mundo real, en cada imagen de trama incluida en el vídeo, y es posible identificar una huella de movimiento que incluye esa posición representada. Es decir, con esta técnica, es posible identificar una huella que expresa cómo se desplaza un píxel dado en una imagen de trama dada, con el transcurso hacia delante o hacia atrás del tiempo.

30 Además, en la bibliografía de patente 1 se propone una técnica para buscar otros vídeos similares al vídeo especificado por el usuario.

**Lista de citas**

35 **Bibliografía de patente**

Bibliografía de patente 1: publicación denominada "kokai" de solicitud de patente japonesa no examinada nº 2001-134589.

40 **Bibliografía no de patente**

Bibliografía no de patente 1: D. G. Lowe, Object recognition from Local Scale-Invariant Features [Reconocimiento de objetos a partir de rasgos locales invariantes a escala], Anales de la Conferencia Internacional del IEEE sobre visión por ordenador, págs. 1150-1157, 1999.

45 Bibliografía no de patente 2: N. Dalal y B. Triggs, Histograms of Oriented Gradients for Human Detection [Histogramas de gradientes orientadas para detección humana], Anales de la Conferencia de la Sociedad del Ordenador del IEEE sobre visión por ordenador y reconocimiento de patrones, págs. 886-893, 2005.

50 Bibliografía no de patente 3: Peter Sand y Seth Teller, Particle Video: Long-Range Motion Estimation using Point Trajectories [Vídeo en Partículas: Estimación de movimiento a larga distancia usando trayectorias puntuales], Revista Internacional de Visión por Ordenador, vol. 80, nº 1, págs. 72-91, Springer, 2008.

55 Bibliografía no de patente 4: Dan B. Goldman, Chris Gonterman, Brian Curless, David Salesin y Steven M. Seitz, Video Object Annotation, Navigation and Composition [Anotación, navegación y composición de objetos de vídeo], UIST 2008, Anales del simposio de ACM sobre Software y Tecnología de Interfaces de Usuarios, págs. 3-12, ACM 978-1-59593-975-3/08/10, <http://www.danbgoldman.com/uw/papers/ivoa.uist08.pdf>, 2008.

**Sumario de la invención**

60 **Problema técnico**

65 Sin embargo, cuando se hacen búsquedas de vídeo, sigue siendo grande la demanda de los usuarios que quieren poder especificar fácilmente información tal como el aspecto exterior, la posición, la orientación y/o similares de un objeto representado en el vídeo.

En particular, cuando ya se ha preparado una imagen fija en la que se representa el objeto deseado, existe la demanda de que los usuarios quieren poder especificar fácilmente la posición y la orientación cuando ese objeto aparece en el vídeo, usando esa imagen fija.

5 Incluso en vídeos en los que aparece un objeto tomado en una dirección de toma diferente o con una posición u orientación que difiere de la orientación del objeto representado en la imagen fija, existe una fuerte demanda de los usuarios que quieren poder obtener la misma como resultado de búsqueda.

10 Considerando lo anterior, un objetivo de la presente invención es proporcionar un dispositivo de búsqueda de vídeo, un procedimiento de búsqueda de vídeo, un medio de registro y un programa para efectuar búsquedas de vídeo de manera que sea posible para un usuario especificar fácilmente un objeto que debería aparecer en los vídeos, junto con la posición, la orientación y, además, el movimiento del mismo.

15 Solución al problema

El objeto se resuelve mediante las características de las reivindicaciones independientes. Los modos de realización preferidos se describen en las reivindicaciones dependientes. El dispositivo de búsqueda de vídeo según un primer aspecto de un ejemplo comprende:

20 un receptor para recibir la entrada de una imagen fija en la que se representa un objeto de referencia, una primera posición de referencia y una segunda posición de referencia contenidas en una región de referencia en la que se representa el objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

25 un extractor para extraer una imagen de referencia representada en la región de referencia, basándose en la primera posición de referencia y la segunda posición de referencia; y

30 un buscador para buscar, entre múltiples vídeos, imágenes de trama similares que sean imágenes de trama contenidas en los vídeos y en las que se representan imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio correspondientes a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local en la que se representa la imagen local en la imagen de trama similar buscada, rastrear una primera huella de movimiento a lo largo de la cual un primer píxel significativo, representado en la primera posición de inicio, se desplaza en la imagen de trama similar buscada, y una segunda huella de movimiento a lo largo de la cual se desplaza un segundo píxel significativo representado en la segunda posición de inicio, explorando imágenes de trama contenidas en el vídeo, que incluye la imagen de trama similar buscada en el sentido de avance o retroceso cronológico con respecto a la imagen de trama similar buscada, buscar una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino, y producir como resultado de búsqueda un vídeo que incluye la imagen de trama similar buscada y la imagen de trama de destino buscada.

Además, en el dispositivo de búsqueda de vídeo de un ejemplo, es posible tener una composición de manera que:

45 el receptor reciba adicionalmente la entrada de una huella de destino en una imagen de trama que incluye el vídeo; y

50 el buscador, para cada uno de los vídeos buscados, rastree una huella representativa a lo largo de la cual se desplaza un píxel representativo de una imagen local representada en la imagen de trama de destino buscada, explorando imágenes de trama contenidas en el vídeo buscado en el sentido cronológico desde una imagen de trama de destino incluida en el vídeo, y produzca como resultado de búsqueda un vídeo en el que la huella de destino y la huella representativa son similares.

55 Además, en el dispositivo de búsqueda de vídeo de un ejemplo, es posible tener una composición de manera que el buscador, para cada uno de los vídeos buscados, rastree una huella representativa a lo largo de la cual se desplaza un píxel representativo de una imagen local representada en la imagen de trama de destino buscada, explorando imágenes de trama contenidas en el vídeo buscado en el sentido cronológico desde una imagen de trama de destino incluida en el vídeo, y produzca como resultado de la búsqueda la huella representativa de esos vídeos junto con aquellos vídeos.

60 Además, en el dispositivo de búsqueda de vídeo de un ejemplo, es posible tener una composición de manera que:

el extractor obtenga cantidades características de imagen para la imagen de referencia; y

65 el buscador busca las imágenes de trama similares buscando imágenes de trama que tengan cantidades características de imagen similares a las cantidades características de imagen de la imagen de referencia.

Además, en el dispositivo de búsqueda de vídeo de un ejemplo, es posible tener una composición de manera que:

el extractor obtenga adicionalmente cantidades características de imagen de una imagen especular que es una inversión de la imagen de referencia; y

- 5 el buscador busque las imágenes de trama similares buscando imágenes de trama similares a las cantidades características de imagen de la imagen de referencia o las cantidades características de imagen de la imagen especular.

El procedimiento de búsqueda de vídeo según un segundo aspecto de un ejemplo comprende las etapas de:

- 10 recibir la entrada de una imagen fija en la que se representa un objeto de referencia, una primera posición de referencia y una segunda posición de referencia contenidas en una región de referencia en la que se representa el objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

- 15 extraer una imagen de referencia representada en la región de referencia basándose en la primera posición de referencia y en la segunda posición de referencia; y

- 20 buscar, entre múltiples vídeos, imágenes de trama similares que sean imágenes de trama contenidas en los vídeos y en las que se representen imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio correspondientes a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local en la que la imagen local se representa en la imagen de trama similar buscada, rastrear una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, representado en la primera posición de inicio, en la imagen de trama similar buscada y una segunda huella de movimiento a lo largo de la cual se desplaza un segundo píxel significativo representado en la segunda posición de inicio, explorando imágenes de trama contenidas en el vídeo que incluyen la imagen de trama similar buscada en el sentido de avance o retroceso cronológico con respecto a la imagen de trama similar buscada,

- 30 buscar una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino, y producir como resultado de búsqueda un vídeo que incluye la imagen de trama similar buscada y la imagen de trama de destino buscada.

- 35 El medio de registro legible por ordenador, según un tercer aspecto de un ejemplo, está conformado de manera que registre un programa que hace que un ordenador funcione como:

- 40 un receptor para recibir la entrada de una imagen fija en la que se representa un objeto de referencia, una primera posición de referencia y una segunda posición de referencia contenidas en una región de referencia en la que se representa el objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

un extractor para extraer una imagen de referencia representada en la región de referencia, basándose en la primera posición de referencia y la segunda posición de referencia; y

- 45 un buscador para buscar, entre múltiples vídeos, imágenes de trama similares que sean imágenes de trama contenidas en los vídeos y en las que se representen imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio correspondientes a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local en la que la imagen local se representa en la imagen de trama similar buscada,

- 50 rastrear una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, representado en la primera posición de inicio en la imagen de trama similar buscada, y una segunda huella de movimiento a lo largo de la cual se desplaza un segundo píxel significativo representado en la segunda posición de inicio, explorando imágenes de trama contenidas en el vídeo que incluye la imagen de trama similar buscada en el sentido de avance o retroceso cronológico con respecto a la imagen de trama similar buscada, buscar una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino, y producir como resultado de búsqueda un vídeo que incluye la imagen de trama similar buscada y la imagen de trama de destino buscada.

- 60 El programa según un cuarto aspecto de un ejemplo está conformado para hacer que un ordenador funcione como:

- 65 un receptor para recibir la entrada de una imagen fija en la que se representa un objeto de referencia, una primera posición de referencia y una segunda posición de referencia contenidas en una región de referencia en la que se representa el objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

un extractor para extraer una imagen de referencia representada en la región de referencia basándose en la primera posición de referencia y en la segunda posición de referencia; y

5 un buscador para buscar, entre múltiples vídeos, imágenes de trama similares que sean imágenes de trama contenidas en los vídeos y en las que se representen imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio correspondientes a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local en la que la imagen local se representa en la imagen de trama similar buscada,

10 rastrear una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, representado en la primera posición de inicio, dentro de la imagen de trama similar buscada y una segunda huella de movimiento a lo largo de la cual se desplaza un segundo píxel significativo representado en la segunda posición de inicio, explorando imágenes de trama contenidas en el vídeo que incluye la imagen de trama similar buscada en el sentido de avance o retroceso cronológico con respecto a la imagen de trama similar buscada, buscar una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino, y producir como resultado de búsqueda un vídeo que incluye la imagen de trama similar buscada y la imagen de trama de destino buscada.

20 Es posible registrar el programa de un ejemplo en un medio de registro no transitorio legible por ordenador, tal como un disco compacto, un disco flexible, un disco duro, un disco magneto-óptico, un disco de vídeo digital, cinta magnética, memoria semiconductora y/o similares. Además, es posible que tales medios de registro se distribuyan y se vendan independientemente de un ordenador.

25 Además, es posible que el programa de un ejemplo se cargue desde el medio de registro descrito anteriormente, a un medio de registro legible y grabable por un ordenador, tal como una RAM (memoria de acceso aleatorio) y/o similares, y se almacene temporalmente, y que una CPU (unidad central de procesamiento) lea, interprete y ejecute el programa registrado en la RAM y/o similares.

30 Además, es posible que el programa de un ejemplo se distribuya y se venda mediante un medio de transmisión transitorio, tal como una red de comunicación de ordenadores y/o similar, independiente del ordenador que ejecuta el programa.

35 Efectos ventajosos de la invención

Con la presente invención, es posible proporcionar un dispositivo de búsqueda de vídeo, un procedimiento de búsqueda de vídeo, un medio de registro y un programa para efectuar búsquedas de vídeo de manera que sea posible para un usuario especificar fácilmente un objeto que debería aparecer en los vídeos, junto con la posición, la orientación y, además, el movimiento del mismo.

#### Breve descripción de los dibujos

45 La figura 1A es un dibujo explicativo que muestra un ejemplo de una imagen fija en la que se representa un objeto de referencia;

la figura 1B es un dibujo explicativo que muestra el estado de una región de referencia en la imagen fija;

50 la figura 1C es un dibujo explicativo que muestra el estado de una imagen de referencia;

la figura 2A es un dibujo explicativo que muestra un ejemplo de la posición y orientación en la que debe situarse un objeto de destino en una imagen de trama incluida en un vídeo;

55 la figura 2B es un dibujo explicativo que muestra el estado de una región local en la imagen de trama;

la figura 2C es un dibujo explicativo que muestra el estado de una imagen local;

60 la figura 3 es un dibujo explicativo que muestra el estado en el que un usuario especifica la relación de correspondencia entre la imagen de referencia y la imagen de destino;

la figura 4 es un dibujo explicativo que muestra el estado en el que un usuario especifica la relación de correspondencia entre la imagen de referencia y la imagen de destino;

65 la figura 5 es un dibujo explicativo que muestra la composición resumida de un dispositivo de búsqueda de vídeo según un ejemplo de la presente invención;

la figura 6 es un diagrama de flujo que muestra el flujo de control del proceso de búsqueda de vídeo ejecutado por un dispositivo de búsqueda de vídeo según un ejemplo de la presente invención;

5 la figura 7 es un dibujo explicativo que muestra el estado de la imagen de referencia obtenida mediante la especificación del usuario;

la figura 8 es un dibujo explicativo que muestra un ejemplo de un vídeo que contiene una imagen de trama que tiene una imagen local similar a la imagen de referencia;

10 la figura 9 es un dibujo explicativo que muestra la relación de correspondencia entre la imagen de referencia y la imagen de trama similar;

la figura 10 es un dibujo explicativo que muestra el estado de movimiento de píxeles en el vídeo;

15 la figura 11 es un dibujo explicativo que muestra la huella del movimiento de píxeles en el vídeo;

la figura 12 es un dibujo explicativo que muestra una explicación de la relación de posición relativa de las posiciones de referencia y la relación de posición relativa de la posición de destino;

20 la figura 13 es un dibujo explicativo que muestra una explicación de la relación de posición relativa de las posiciones de referencia y la relación de posición relativa de las posiciones de destino;

la figura 14A es un dibujo explicativo que muestra el estado de una huella representativa;

25 la figura 14B es un dibujo explicativo que muestra el estado de una imagen resumida;

la figura 15 es un dibujo explicativo que muestra el estado cuando un usuario especifica una huella de destino; y

30 la figura 16 es un diagrama de flujo que muestra el contenido de un proceso seguido en el proceso de búsqueda de vídeo.

### **Descripción de modos de realización**

35 A continuación, se explica el modo de realización preferido de la presente invención. Sin embargo, se pretende que este modo de realización preferido sea ilustrativo y que no limite el alcance de la presente invención. Por consiguiente, el modo de realización preferido puede modificarse en cuanto a su disposición y detalle sin apartarse de los principios divulgados en el presente documento, y se pretende que se interprete la solicitud como incluyente de todas dichas modificaciones y variaciones.

#### **40 Modo de realización 1**

Es posible realizar un dispositivo de búsqueda de vídeo según la presente invención, ejecutando un programa prescrito en diversos ordenadores, tales como un ordenador servidor o un ordenador personal, y/o similares.

45 Un ordenador es hardware para recibir instrucciones desde un usuario mediante un dispositivo de entrada tal como un teclado, un ratón y/o similares, emitir los resultados de los procesos a un dispositivo de salida tal como un dispositivo de visualización y/o similares, y efectuar la entrada y salida descritas anteriormente comunicándose con otros equipos mediante una NIC (tarjeta de interfaz de red), ejecutando la CPU un programa, usando una RAM como región de memoria temporal o designación de salida para los resultados del proceso, y también es posible omitir de manera adecuada equipos para entrada y salida.

50 Un programa ejecutado por la CPU y una base de datos en la que están almacenados registros que expresan diversos tipos de información, que es el objetivo de procesamiento por la CPU, se registran en una unidad de disco duro y/o similar del ordenador. La CPU efectúa procesos tales como buscar registros que coinciden con una consulta de búsqueda de la base de datos y añadir nuevos registros a la base de datos.

55 En este modo de realización preferido, además de múltiples vídeos, se registra en la base de datos información local, a partir de la cual las diversas imágenes de trama contenidas en el vídeo se consideran imágenes fijas.

60 Además, también es posible acelerar los procesos del dispositivo de búsqueda de vídeo según la presente invención, haciendo que múltiples ordenadores, conectados mediante una red de comunicación de ordenadores, tal como líneas de conexión dedicada, circuitos de comunicación, Internet y/o similares, ejecuten el proceso descrito anteriormente en paralelo, de manera descentralizada y de manera concurrente.

65 Además de esto, también es posible realizar el dispositivo de búsqueda de vídeo de la presente invención creando un diseño de circuito electrónico a partir de programas, mediante la aplicación de tecnología tal como FPGA

(formación de compuertas programables *in situ*) y/o similar, y organizando circuitos electrónicos especiales basándose en este diseño.

(Procedimiento de búsqueda de vídeo)

5 Con este modo de realización preferida, el usuario especifica qué tipo de vídeo debe obtenerse especificando información diversa junto con imágenes fijas. A continuación, se explicarán en primer lugar procedimientos para especificar dicha información.

10 Con este modo de realización preferido, el usuario especifica la posición deseada y la orientación deseada del objeto deseado en la trama de la pantalla. Además, el dispositivo de búsqueda de vídeo busca vídeos en los que aparecen objetos (a continuación en el presente documento denominados "objetos de destino") que se asemejan al objeto (a continuación en el presente documento, denominado el "objeto de referencia") que tiene la orientación deseada en la ubicación deseada.

15 Pensando en esto en comparación con búsquedas de documentos, el objeto de referencia corresponde a la cadena de caracteres que es la clave de búsqueda y los objetos de destino corresponden a cadenas de caracteres que aparecen en el documento obtenido como resultado de efectuar la búsqueda.

20 En una búsqueda de documentos, es posible buscar documentos en los que aparecen cadenas de caracteres que coinciden perfectamente con la clave de búsqueda y también es posible buscar documentos en los que aparecen cadenas de caracteres similares a la clave de búsqueda (que tienen diversos estados, tales como sinónimos o variaciones en cuanto a la ortografía). Es decir, la clave de búsqueda y las cadenas de caracteres que aparecen en cadenas de caracteres que son resultados de búsqueda no coinciden necesariamente.

25 De manera similar, con este modo de realización preferido, el usuario especifica como clave de búsqueda un objeto de referencia representado en una imagen fija, una ubicación deseada y una orientación deseada.

30 Acto seguido, el dispositivo de búsqueda de vídeo busca vídeos en los que un objeto de destino, que se asemeja al objeto de referencia, aparece en la posición deseada con la orientación deseada.

Para facilitar la comprensión, la explicación a continuación usará un automóvil como ejemplo de objeto.

35 En primer lugar, el usuario introduce una imagen fija en la que se representa el objeto de referencia correspondiente a la clave de búsqueda, usando, por ejemplo, un motor de búsqueda de imágenes y/o similares.

La figura 1A es un dibujo explicativo que muestra un ejemplo de una imagen fija en la que se representa un objeto de referencia.

40 Tal como se muestra en este dibujo, un automóvil 102, que es el objeto de referencia, se representa en la imagen fija 101. En este dibujo, el automóvil 102 se representa orientado hacia la izquierda en el centro de la imagen. Es decir, la imagen fija 101 se obtuvo fotografiando el automóvil 102 desde una cámara situada directamente a la izquierda, tal como se ve con respecto al conductor del automóvil 102.

45 A continuación, para facilitar la comprensión, la región en la que el objeto de referencia se representa en la imagen fija se denomina la región de referencia y la imagen representada en esta región de referencia se denomina la imagen de referencia.

50 La figura 1B es un dibujo explicativo que muestra el estado de la región de referencia dentro de la imagen fija 101, y la figura 1C es un dibujo explicativo que muestra el estado de la imagen de referencia. Estos se explican a continuación con referencia a los dibujos.

55 En la figura 1B, una región de referencia 122 rodeada por una línea discontinua se muestra dentro de una región total 121 de la imagen fija 101. La región total 121 corresponde a la imagen fija 101 como un todo, y la región de referencia 122 corresponde a la región en la que el automóvil 102 (el objeto de referencia) se representa en la imagen fija 101.

60 La figura 1C muestra una imagen de referencia 123 con el aspecto externo del automóvil 102 (objeto de referencia) solo, recortado de la imagen fija 101. En este modo de realización preferido, se realiza una búsqueda de un vídeo en el que aparece un objeto de destino que se asemeja al objeto de referencia expresado por la imagen de referencia 123.

65 Tal como se ha descrito anteriormente, el usuario especifica como claves de búsqueda una posición deseada y una orientación deseada. A continuación, la explicación es para un ejemplo en el que el usuario ha especificado la parte superior derecha de la trama de la pantalla como la posición deseada y hacia abajo a la izquierda como la orientación deseada.

En este ejemplo, se considera que el usuario quiere, como resultados de búsqueda, vídeos en los que un objeto de destino, que se asemeja al objeto de referencia, aparece en la parte superior derecha de la trama de la pantalla con una orientación hacia abajo a la izquierda.

5 La figura 2A es un dibujo explicativo que muestra un ejemplo de la posición y orientación en que debería estar situado el objeto de destino en la imagen de trama contenida en el vídeo. La explicación a continuación hace referencia a esta imagen.

10 En el ejemplo mostrado en este dibujo, un automóvil 202 que se asemeja a este automóvil 102 (que puede ser el mismo que el automóvil 102) se representa como el objeto de destino en la imagen de trama 201, del que se piensa que sea lo que el usuario quiere buscar.

15 Es decir, con este modo de realización preferido, los vídeos que incluyen la imagen de trama 201, en la que aparece el objeto de destino (el automóvil 202) que se asemeja al objeto de referencia (el automóvil 102) representado en la imagen fija 101, son candidatos de resultados de búsqueda.

20 En la imagen de trama 201, el automóvil 202 aparece en la parte superior derecha de la trama de la pantalla, orientado hacia abajo a la izquierda. Es decir, el vídeo deseado es un vídeo que incluye un instante en el que el automóvil 202 fue fotografiado por una cámara situada en el frente y mirando hacia abajo a la izquierda, tal como se ve desde el conductor del automóvil 202.

25 Para facilitar la comprensión a continuación, la región en la que el objeto de destino se representa en la imagen de trama se denomina la región local y la imagen representada en esta región local se denomina la imagen local.

La figura 2B es un dibujo explicativo que muestra el estado de la región local en la imagen de trama 201, y la figura 2C es un dibujo explicativo que muestra el estado de la imagen local. La explicación a continuación hace referencia a estas imágenes.

30 En la figura 2B, una región de destino 222, delimitada por una línea discontinua, se muestra en la región total 221 de la imagen de trama 201. La región total 221 corresponde a la imagen de trama 201 como un todo y la región de destino 222 corresponde a la región en la que se representa el automóvil 202 (objeto de destino) dentro de la imagen de trama 201.

35 La figura 2C muestra la imagen de destino 223 como el aspecto externo del automóvil 202 (objeto de destino) recortado con respecto a la imagen de trama 201.

40 Con este modo de realización preferido, se busca un vídeo de manera que el objeto de referencia expresado mediante la imagen de referencia 123 y el objeto de destino expresado mediante la imagen de destino 223 sean similares.

45 Sin embargo, la imagen fija 101 es el automóvil 102 fotografiado desde el lado izquierdo, de modo que no es posible obtener una composición similar al objeto de destino 223, incluso implementando transformaciones afines que incluyen rotación, inversión, desplazamientos de plano, ampliación/reducción, poda y/o similares de la imagen de referencia 123.

En este ejemplo, mientras que la parte frontal del automóvil 102 no puede verse claramente en la imagen fija 101, la parte frontal del automóvil 202 puede verse correctamente en la imagen de trama 201.

50 En general, si se usa una técnica de búsqueda de imágenes, no es imposible obtener una imagen fija en la que aparece el objeto de referencia deseado por el usuario.

55 Sin embargo, prácticamente no hay casos en los que la posición y orientación con las que se representa el objeto de referencia en esa imagen fija, junto con la dirección de toma y el ángulo del objeto de referencia, coincidan tal cual con la posición y orientación con las que se representa el objeto de destino en la imagen de trama contenida en el vídeo junto con la dirección de toma y el ángulo del objeto de destino.

60 Por tanto, con este modo de realización preferido, el usuario especifica una relación de correspondencia entre la imagen de referencia en la imagen fija 101 y la imagen de destino en la imagen de trama 201.

Las figuras 3 y 4 son dibujos explicativos que muestran el estado en el que un usuario especifica la relación de correspondencia entre la imagen de referencia y la imagen de destino. La explicación a continuación hace referencia a estos dibujos.

65 En los ejemplos mostrados mediante estos dibujos, se exhibe una trama de vídeo 331 que expresa una región en la que se debería reproducirse un vídeo, que incluye la imagen fija 101 y la imagen de trama 201, dentro de la pantalla

301.

En el ejemplo mostrado en la figura 3, la imagen fija 101 se amplía o reduce y se coloca a fin de ocupar el centro de la trama de vídeo 331, pero en el ejemplo mostrado en la figura 4, la imagen fija 101 está situada separada de la trama de vídeo 331.

Esto es debido a que, en general, la proporción de aspecto difiere entre la imagen fija 101 y la trama de vídeo 331.

En este caso, el usuario efectúa:

(a) una designación trazando una flecha 371 desde una posición de referencia 311, en la que el borde frontal del automóvil 102 se representa en la imagen fija 101, hasta una posición de destino 351, en la que el borde frontal del automóvil 202 debería representarse en la trama de vídeo 331; y

(b) una designación trazando una flecha 372 desde una posición de referencia 312, en la que el borde posterior del automóvil 102 se representa en la imagen fija 101, hasta una posición de destino 352, en la que el borde posterior del automóvil 202 debería representarse en la trama de vídeo 331.

Estas designaciones se efectúan, por ejemplo, mediante una operación de arrastrar y soltar, usando un ratón o una operación de deslizamiento sobre una pantalla táctil o panel táctil (una operación de tocar la pantalla con un dedo, y luego retirar el dedo después de hacer que el dedo se mueva mientras todavía está en contacto).

Con este modo de realización preferido, el usuario selecciona las ubicaciones (el borde frontal y el borde posterior) que han de ser marcas del objeto de referencia (el automóvil 102) designando las posiciones de referencia 311 y 312 en las que esas ubicaciones (el borde frontal y el borde posterior) se representan en la imagen fija.

Además, el usuario designa las posiciones de destino 351 y 352, en las que las ubicaciones (el borde frontal y el borde posterior) correspondientes al objeto de destino (el automóvil 202) deberían representarse en la trama de vídeo 331.

En los ejemplos mostrados en estos dibujos, la proporción de aspecto de la trama de vídeo 331 es 9:16, y coincide en cuanto a su tamaño con la imagen de vídeo que es el objetivo de búsqueda. Por consiguiente, las posiciones de destino 351 y 352 designadas por el usuario en la trama de vídeo 331 se usan como posiciones que expresan directamente la posición y la orientación en las que debería aparecer el automóvil 202 que es el objeto de destino 202.

Por otra parte, para vídeos que tienen diferentes proporciones de aspecto o vídeos con diferentes números de píxeles verticales y horizontales, los píxeles verticales y horizontales de esos vídeos se hacen coincidir estirando o contrayendo horizontal y verticalmente la trama de vídeo 331 de proporción de aspecto 9:16. Es decir, como posición de destino, en un vídeo que tiene una proporción de aspecto, o números de píxeles verticales y horizontales, diferentes, sería conveniente usar, no las posiciones de destino 351 y 352 que el usuario designó directamente en la trama de vídeo 331, sino, en cambio, las desplazadas mediante expansión o contracción, tal como se ha descrito anteriormente.

Con este modo de realización preferido, es posible buscar imágenes de vídeo especificando, mediante una operación sencilla, en qué posición y orientación debería aparecer el objeto de destino en una imagen de trama contenida en el vídeo. A continuación, se explican detalles de la técnica de búsqueda de vídeo.

(Dispositivo de búsqueda de vídeo)

La figura 5 es un dibujo explicativo que muestra la composición resumida de un dispositivo de búsqueda de vídeo según un ejemplo de la presente invención. La figura 6 es un diagrama de flujo que muestra el flujo de control en un proceso de búsqueda de vídeo ejecutado por el dispositivo de búsqueda de vídeo, según un modo de realización a modo de ejemplo de la presente invención.

Un dispositivo de búsqueda de vídeo 501 según este modo de realización preferido es realizado por un programa prescrito que se ejecuta en un ordenador, y comprende un receptor 502, un extractor 503 y un buscador 504.

El receptor 502 recibe la entrada de la información a continuación como condiciones de búsqueda (etapa S601).

(a) imagen fija 101.

(b) posiciones de referencia 311 y 312 dentro de la imagen fija 101 (en este modo de realización preferido, los puntos de inicio de las flechas 371 y 372).

(c) posiciones de destino 351 y 352 dentro de la trama de vídeo 331 (en este modo de realización preferido, los

puntos finales de las flechas 371 y 372).

Tal como se ha descrito anteriormente, las funciones del receptor 502 se realizan mediante el trabajo conjunto de un visor o un ratón, una pantalla táctil o un panel táctil, bajo el control de la CPU.

5 A continuación el extractor 503 extrae la imagen de referencia representada en la región de referencia basándose en las posiciones de referencia 311 y 312 (etapa S602).

10 Tal como se ha descrito anteriormente, con este modo de realización preferido, el usuario especifica la imagen fija 101 cuando se representa el objeto de referencia. Además, el usuario especifica las posiciones de referencia 311 y 312 en las que se representan las ubicaciones que el usuario piensa que son las marcas del objeto de referencia, en la imagen fija 101.

15 El extractor 503 extrae la imagen de referencia a partir de la información en relación con la imagen fija 101 y las posiciones de referencia 311 y 312, usando tecnología de procesamiento de imágenes.

La figura 7 es un dibujo explicativo que muestra el estado de la imagen de referencia, obtenida mediante la especificación del usuario, mostrada en las figuras 3 y 4. La explicación a continuación hace referencia a este dibujo.

20 Tal como se muestra en el dibujo, una imagen de referencia 701 muestra el aspecto externo del objeto de referencia según fue fotografiado desde una dirección de toma dada. En este ejemplo, el automóvil 101 está fotografiado desde el lado izquierdo, y el fondo está eliminado de la imagen fotografiada. Es decir, el borde exterior de la región de referencia corresponde a una línea limitrofe que divide el objeto de referencia del fondo en la imagen fija 101.

25 Al extraer la imagen de referencia 701, es posible usar diversos tipos de tecnología de reconocimiento de imágenes. Por ejemplo, después de extraer el contorno de la imagen fija 101 y efectuar la separación de regiones, se seleccionan las regiones que incluyen las posiciones de referencia 311 y 312. Además, se unen las regiones interpuestas entre las regiones seleccionadas, las regiones adyacentes a las regiones seleccionadas y las regiones con pequeñas diferencias en el color medio, y/o similares, y se obtiene una región de referencia repitiendo esto.  
30 Además, también sería conveniente usar la técnica divulgada en la bibliografía de patente 1.

La imagen de referencia 701 es la imagen de referencia que incluye las posiciones de referencia 311 y 312 recortadas de la imagen fija 101, de modo que las posiciones de referencia 311 y 312 estén contenidas ambas dentro de la región de referencia y sean ubicaciones de posición que son marcas del objeto de referencia en la  
35 imagen de referencia 701.

El extractor 503 está realizado mediante el trabajo conjunto de la CPU con la RAM y/o similares cuando se registra información de píxeles de la imagen fija 101.

40 Además, el buscador 504 repite el proceso a continuación para cada uno de los múltiples vídeos (etapa S603).

En primer lugar, se busca una imagen de trama que sea una imagen de trama en la que esté incluido el vídeo y en la que se represente una imagen local similar a la imagen de referencia 701 (a continuación en el presente documento, denominada "imagen de trama similar"), y se toma una determinación en cuanto a si una imagen de trama similar  
45 está contenida o no en el vídeo (etapa S604).

El vídeo está compuesto por una cadena de múltiples imágenes de trama en orden cronológico. En la trama de vídeo 331 preparada en la trama 301, es posible reproducir el vídeo exhibiendo los vídeos de trama en orden en un intervalo de tiempo prescrito (por ejemplo, son comunes un treintavo de segundo o un sesentavo de segundo). Es decir, puede pensarse una imagen de trama como un tipo de imagen fija.  
50

Con este modo de realización preferido, los múltiples vídeos se registran de antemano en una base de datos de vídeo y/o similares. Se incluyen múltiples imágenes de trama para cada vídeo. En este caso, se toma una determinación en cuanto a si una imagen local que se asemeja a la imagen de referencia 701 se representa o no en  
55 cada imagen de trama aplicando diversos tipos de tecnología de determinación de similitud de imágenes y de tecnología de reconocimiento de imágenes.

La figura 8 es un dibujo explicativo que muestra un ejemplo de un vídeo que contiene imágenes de trama que tienen una imagen local que se asemeja a la imagen de referencia 701. La explicación a continuación hace referencia a este dibujo.  
60

Tal como se muestra en los dibujos, un vídeo 801 comprende una cadena de múltiples imágenes de trama 802a, 802b, ..., 802f, ..., 802i, 802j, ... (en este dibujo, el intervalo de tiempo de las imágenes de trama se alarga para facilitar la comprensión. Es decir, las imágenes de trama se muestran reducidas con respecto a los vídeos típicos).  
65 En la explicación a continuación, cuando se mencionan estas múltiples imágenes de trama en términos generales, la referencia será a "imágenes de trama 802", omitiendo la letra al final del número de referencia.

En las imágenes de trama 802, se representa el automóvil 202 que se asemeja al automóvil 102, que es el objeto de referencia.

5 Sin embargo, en las imágenes de trama 802a, 802b, 802i y 802j, la dirección de toma difiere, de modo que, en las imágenes de trama 802a, 802b, 802i y 802j, no se representa una imagen que se asemeja a la imagen de referencia 701.

10 Por otra parte, en la imagen de trama 802f, se representa el automóvil 202 con prácticamente la misma dirección de toma, y esto corresponde a una imagen local 803. Es decir, la imagen local 803 y la imagen de referencia 701 tienen puntos característicos correspondientes de manera mutua, y la imagen local 803 y la imagen de referencia 701 pueden solaparse acometiendo una transformación tal como ampliación o reducción, desplazamiento en paralelo, rotación o inversión de la imagen de referencia 701.

15 Para detectar una imagen que se asemeja a la imagen de referencia desde dentro de la imagen de trama 802, se aplican diversos tipos de tecnología de reconocimiento de imágenes, de detección de imágenes y de reconocimiento de patrones.

20 Por ejemplo, en cada una de las imágenes de trama 802 incluidas en los diversos vídeos, se calculan de antemano las cantidades características locales tales como SIFT, HOG y/o similares, y se almacenan en una base de datos. Además, las cantidades características locales se calculan para la imagen de referencia 701 y se busca en la base de datos una imagen de trama 802 que tenga una cantidad característica local similar a esta cantidad característica local.

25 Cuando la cantidad característica local de la imagen de trama, incluida en un vídeo dado, se registra en la base de datos, si las cantidades características locales de imágenes 802 de trama mutuamente adyacentes son similares, es posible controlar volúmenes de datos y acelerar la búsqueda registrando las mismas con un afinamiento apropiado.

30 De esta manera, si una imagen de trama similar, en la que se representa una imagen local 803, similar a la imagen de referencia 701 (en el ejemplo mostrado en los dibujos, la imagen de trama 802f corresponde a una imagen de trama similar) (etapa S604; Sí), el buscador 504 determina dónde están situadas en la imagen local 803 las posiciones de referencia 311 y 312 cuando la imagen de referencia 701 se solapa con la imagen local 803 (a continuación, esta posición se denomina la "posición de inicio"). Es decir, el buscador 504 encuentra la posición de inicio correspondiente a las posiciones de referencia 311 y 312 en la región de referencia, en la región local en la que la imagen local 803 se representa en la imagen de trama 802f similar buscada (etapa S605).

La figura 9 es un dibujo explicativo que muestra la relación de correspondencia entre la imagen de referencia 701 y la imagen de trama 802f similar. La explicación a continuación hace referencia a este dibujo.

40 En este dibujo, se presentan visualmente puntos negros como las marcas de las posiciones de referencia 311 y 312 en el borde frontal y el borde posterior del automóvil 102 en la imagen de referencia 701.

45 Por otra parte, se presentan visualmente marcas negras como las marcas de las posiciones de inicio 811 y 812 en el borde frontal y el borde posterior del automóvil 202 en el interior del automóvil 202 correspondiente a la imagen local 803 de la imagen de trama 802f similar.

Las posiciones de referencia 311 y 312 son posiciones en las que se representan las posiciones seleccionadas como marcas por el usuario en el objeto de referencia (el automóvil 102) en la imagen de referencia 701.

50 Por otra parte, las ubicaciones correspondientes a las marcas en el objeto de referencia (el automóvil 102) se representan como las posiciones de inicio 811 y 812 en la imagen local 803 en el objeto de destino (el automóvil 202 similar al automóvil 102).

55 Cuando se detecta la imagen local 803 similar a la imagen de referencia 701, se realizan transformaciones tales como ampliar o reducir, rotar, invertir y/o similares, de modo que la imagen de referencia 701 se solape con la imagen local 803, e implementando de manera similar estas transformaciones en las posiciones de referencia 311 y 312, se obtienen las posiciones de inicio 811 y 812.

60 A continuación de esto, el buscador 504 rastrea cómo se desplazan los píxeles situados en las posiciones de inicio 811 y 812 en la imagen de trama 802f similar (a continuación en el presente documento, denominados "píxeles significativos") en el vídeo 801 y obtiene la huella de movimiento de los mismos. Es decir, explorando las imágenes de trama 802a, 802b, ..., 802i, 802j, ..., contenidas en el vídeo 801 que contiene la imagen de trama 802f similar buscada en el sentido de avance cronológico (el orden de las imágenes de trama 802g, 802h, 802i, 802j, ...) o en el sentido de retroceso (el orden de las imágenes de trama 802e, 802d, 802c, 802b,...) con respecto a la imagen de trama similar, el buscador 504 rastrea la huella de movimiento a lo largo de la cual los píxeles significativos representados en las posiciones de inicio 811 y 812 en la imagen de trama 802f similar buscada se desplazan,

mediante su representación en otras imágenes de trama 802 (etapa S606).

La figura 10 es un dibujo explicativo que muestra el estado de movimiento de píxeles en el vídeo 801. La explicación a continuación hace referencia a este dibujo.

5 Tal como se muestra en este dibujo, los píxeles significativos 901 y 902 se representan en las posiciones de inicio 811 y 812 de la imagen de trama 802f similar, pero estas representan respectivamente el borde frontal y el borde posterior del automóvil 202.

10 En este dibujo, las posiciones de los píxeles significativos 901 y 902 que representan el borde frontal y el borde posterior del automóvil 202 se muestran en las imágenes de trama 802a, 802b, ..., 802f, ..., 802i, 802j, ..., contenidas en el vídeo 801. En cada imagen de trama 802a, 802b, ..., 802f, ..., 802i, 802j, ..., se añade una letra correspondiente a esa imagen de trama 802 al final de los números de referencia de los píxeles significativos 901 y 902.

15 Además, en este dibujo, se presentan visualmente las posiciones de destino 351 y 352 a las que se hace referencia en procesos posteriores. Las posiciones de destino 351 y 352 son designadas por el usuario en el trama de vídeo 331, y las posiciones de destino 351 y 352 se obtienen en cada imagen de trama 802 mediante ampliación/reducción y/o similares, de modo que cada imagen de trama 802 quepa exactamente en el interior del trama de vídeo 331.

20 Es posible obtener este tipo de rastreo de los píxeles significativos 901 y 902 encontrando la relación de correspondencia entre los puntos característicos en imágenes de trama 802 adyacentes, usando una técnica tal como SIFT, HOG y/o similares, e interpolando entonces cambios en las posiciones de los puntos característicos correspondientes, y calculando, mediante esto, cómo se desplazan las posiciones de los píxeles significativos 901 y 902 en una imagen de trama 802, en la otra imagen de trama 802. Además, es posible un rastreo de los píxeles significativos 901 y 902, incluso aplicando la técnica de vídeo en partículas.

25 El rastreo de los píxeles significativos 901 y 902 no significa necesariamente rastrear la huella a lo largo de la cual se desplazan píxeles del mismo color. Es decir, esto significa rastrear donde la ubicación del objeto de destino, representado en las imágenes de trama 802 mediante los píxeles significativos 901 y 902, está representada en otra imagen de trama 802. Por consiguiente, en función del estado de la fotografía del objeto de destino y/o similares, el color de los píxeles significativos 901 y 902 puede cambiar gradualmente.

30 De esta manera, las imágenes de trama 802 se exploran en el sentido de avance cronológico o en el sentido de retroceso con respecto a la imagen de trama 802 similar, contenida en el vídeo 801, y se rastrea dónde están situados los píxeles significativos 901 y 902 en cada imagen de trama 802. Por consiguiente, se obtienen las huellas de movimiento de los píxeles significativos 901 y 902.

35 La figura 11 es un dibujo explicativo que muestra las huellas de movimiento de los píxeles significativos 901 y 902 en el vídeo 801. La explicación a continuación hace referencia a este dibujo.

40 Tal como se muestra en este dibujo, en el vídeo 801, los píxeles significativos 901 y 902 representan unas huellas de movimiento 951 y 952 que se desplazan suavemente desde la parte superior derecha de la pantalla (imagen de trama 802a) hasta la parte inferior izquierda de la pantalla (desde la imagen de trama 802j en adelante) con el paso del tiempo. Las huellas de movimiento 951 y 952 pasan a través de las posiciones de inicio 811 y 812, respectivamente, en la imagen de trama 802f similar.

45 El buscador 504 rastrea estas huellas de movimiento 951 y 952 con el paso del tiempo, busca imágenes de trama (a continuación en el presente documento, denominadas "imágenes de trama de destino") allí donde estas llegan simultáneamente cerca de las posiciones de destino 351 y 352, y determina si el vídeo 801 contiene o no la imagen de trama de destino (etapa S607).

50 Volviendo a la figura 10, cuando las posiciones de los píxeles significativos 901 y 902 en cada imagen de trama 802 se comparan con las posiciones de destino 351 y 352, se entiende en la imagen de trama 802b que estas están situadas cerca. Por consiguiente, en este ejemplo, la imagen de trama 802b se encuentra como resultado de búsqueda, como la imagen de trama de destino.

55 De esta manera, si se encuentra una imagen de trama de destino (etapa S607; Sí), este vídeo 801 se emite como un resultado de búsqueda (etapa S608).

60 Además, para todos los vídeos 801 registrados en la base de datos, el proceso desde la etapa S603 en adelante se repite (etapa S609) y este proceso concluye cuando el proceso se ha ejecutado para todos los vídeos.

65 Sería conveniente presentar visualmente la salida como resultado de búsqueda en la etapa S608, enviándola inmediatamente a la pantalla de un dispositivo de visualización y/o similar, de modo que se presenten visualmente de manera adicional los resultados de búsqueda en pantallas sucesivas cada vez que se descubre un vídeo

deseado. Además de esto, sería conveniente acumular, en la etapa S608, resultados de búsqueda enviándolos temporalmente a una RAM y/o similares, y después de que se haya completado la repetición de las etapas S603 a S609, para presentarlos visualmente todos a la vez en la pantalla de un dispositivo de visualización y/o similar.

- 5 Por otra parte, cuando no se descubren imágenes de trama similares (etapa S604; No), o cuando no se descubren imágenes de trama de destino (etapa S607: No), el proceso avanza a la etapa S609. Por consiguiente, este vídeo 801 no pasa a ser un resultado de búsqueda.

10 De esta manera, en este modo de realización preferido, incluso cuando la composición en la que el objeto de referencia se representa en la imagen fija 101 y la composición en la que el objeto de destino aparece en el vídeo 801 son completamente diferentes, es posible para un usuario efectuar fácilmente la búsqueda de vídeos introduciendo simplemente la relación de correspondencia entre las posiciones de referencia 311 y 312, que son marcas en la imagen fija 101, y las posiciones de destino 351 y 352 en estas tramas de vídeo 331.

- 15 En particular, en este modo de realización preferido, es posible efectuar fácil y rápidamente búsquedas de vídeo, dado que las búsquedas se ejecutan mediante un proceso de imágenes bidimensionales.

#### Modo de realización 2

20 Actualmente, es posible determinar si una imagen local 803 es similar o no a una imagen de referencia 701, incluso cuando está rotada aproximadamente 30 grados, usando tecnología de reconocimiento de imágenes, tal como SIFT y/o similares. Sin embargo, cuando el ángulo de rotación supera este valor, o cuando las imágenes están invertidas y tienen una relación de imagen especular, la detección es imposible.

25 En tal caso, sería conveniente si, además de las cantidades características locales de la imagen de referencia 701, se preparase una imagen especular de esta imagen de referencia 701 invertida y se encontrasen también las cantidades características locales de esta imagen especular, y para imágenes 802 de trama que tienen cantidades características locales, similares a cualquiera de estas dos cantidades características locales que van a buscarse desde la base de datos.

30 A la hora de buscar imágenes de trama similares, sería conveniente usar constantemente tanto la imagen de referencia 701 como la imagen especular de la misma, pero también es posible usar de manera selectiva una entre la imagen de referencia 701 y la imagen especular.

35 Por ejemplo, es posible usar la relación de posición relativa de las posiciones de referencia 311 y 312 y la relación de posición relativa de las posiciones de destino 351 y 352.

40 Las figuras 12 y 13 son dibujos explicativos que explican la relación de posición relativa de las posiciones de referencia 311 y 312 y la relación de posición relativa de las posiciones de destino 351 y 352. La explicación a continuación hace referencia a estos dibujos.

45 En estos dibujos, se muestra un vector de dirección 971 que va desde la posición de referencia 311 hasta la posición de referencia 312 y un vector de dirección 972 que va desde la posición de destino 351 hasta la posición de destino 352.

El ejemplo mostrado en la figura 12 hace referencia al ejemplo mostrado en la figura 3 (y la figura 4), y los vectores de dirección 971 y 972 forman un ángulo agudo. En este tipo de caso, la imagen de referencia 701 se usa a la hora de buscar imágenes de trama similares.

50 Por otra parte, en el ejemplo mostrado en la figura 13, las posiciones de destino 351 y 352 están designadas para cambiar el borde frontal y el borde posterior de las posiciones de referencia 311 y 312 del automóvil 102, que se ven desde el lado izquierdo en la imagen fija 101, para buscar un vídeo en el que aparece un automóvil 202 visto desde el lado derecho y ligeramente hacia arriba.

55 Por consiguiente, los vectores de dirección 971 y 972 forman un ángulo obtuso. En este tipo de caso, se usa una imagen especular para buscar una imagen de trama similar.

60 Además de esto, existe también un procedimiento de determinación del ángulo que es el límite adecuado, usando, por ejemplo, la imagen de referencia 701 si el ángulo formado por los vectores de dirección 971 y 972 es de 60 grados o menor, usando la imagen especular si el ángulo formado por los vectores de dirección 971 y 972 es de al menos 120 grados, y usando tanto la imagen de referencia 701 como la imagen especular si el ángulo formado por los vectores de dirección 971 y 972 es mayor de 60 grados pero menor de 120 grados.

65 Estos procedimientos están basados en la heurística, tal como pensar que la inversión en la dirección horizontal es más típica que la inversión en la dirección vertical cuando se toma el vídeo y, además, que esto es lo que el usuario quiere buscar, y con esto es posible aumentar la velocidad de búsquedas de vídeo.

Modo de realización 3

5 Con la información obtenida a partir de los ejemplos descritos anteriormente, cuando se reproduce el vídeo 801 buscado, es posible encontrar la huella (a continuación en el presente documento denominada la “huella representativa”) sobre la que se desplaza el propio objeto de destino 202.

10 Por ejemplo, en cada imagen de trama 802, se encuentra la región en la que se representa el objeto de destino 202 y el píxel en el punto central (centroide de la región) de la misma se denomina un píxel representativo. Rastreando este píxel representativo, es posible crear la huella representativa sobre la que se desplaza el objeto de destino 202 en el vídeo 801.

15 Además, como procedimiento con cálculos más sencillos, existe un procedimiento que toma como el píxel representativo el píxel en la posición del punto medio de los píxeles significativos 901 y 902 en cada imagen de trama 802.

La figura 14A es un dibujo explicativo que muestra el estado de la huella representativa para el ejemplo mostrado en la figura 11. La explicación a continuación hace referencia a este dibujo.

20 Tal como se muestra en este dibujo, una huella representativa 981 atraviesa el punto medio 908 de los píxeles significativos 901 y 902 en cada una de las imágenes de trama 802. Por consiguiente, la huella representativa 981 es un promedio de las huellas de desplazamiento 951 y 952 del desplazamiento de los píxeles significativos 901 y 902.

25 En el dispositivo de búsqueda de vídeo 501 según este modo de realización preferido, se busca por parte del usuario un vídeo 801 en el que el objeto de destino 202, similar al objeto 102 de referencia, aparece en una posición y situación deseadas por el usuario.

30 Por tanto, en este modo de realización preferido, la huella representativa 981 se emite como un resultado de búsqueda junto con el vídeo buscado 801.

35 La figura 14B es un dibujo explicativo que muestra el estado de una imagen resumida para explicar un resumen de un vídeo 801 cuando el vídeo 801 y la huella representativa 981 se emiten como resultados de búsqueda. La explicación a continuación hace referencia a este dibujo.

40 Tal como se muestra en este dibujo, una imagen resumida 991 representa la imagen de trama 802b, en la que el objeto de destino 202, que es similar al objeto de referencia 102, fuera del vídeo 801 que es el resultado de búsqueda, aparece en la posición y situación deseadas por el usuario, estando superpuesta la huella representativa 981 en forma de flecha.

Es posible que la imagen resumida 991 use imágenes en miniatura y/o similares cuando muestra una presentación visual resumida del vídeo resultado de la búsqueda, y es posible proporcionar información más detallada, con respecto al vídeo 801 obtenido, al usuario mostrando la huella representativa 981 como resultado de búsqueda.

45 Modo de realización 4

50 En los ejemplos descritos anteriormente, la huella representativa 981, en la que se desplaza el objeto de destino 202, se emite como resultado de búsqueda, pero en este modo de realización preferido, se hace que el usuario introduzca la huella en la que debería desplazarse el objeto 202 de destino (a continuación en el presente documento denominada la “huella de destino”) y los vídeos en los que la huella representativa 981 es similar a la huella de destino pasan a ser los resultados de búsqueda.

55 Es decir, en este modo de realización preferido, el receptor 502, además de la información descrita anteriormente, recibe también la entrada de una huella de destino en las imágenes de trama contenidas en el vídeo.

La figura 15 es un dibujo explicativo que muestra el estado cuando el usuario ha designado adicionalmente la huella de destino en el ejemplo de entrada mostrado en la figura 3. La explicación a continuación hace referencia a este dibujo.

60 Tal como se muestra en este dibujo, el usuario introduce una flecha 982 que se inicia cerca de las posiciones de destino 351 y 352 en la trama de vídeo 331, además de las flechas 371 y 372. Esta flecha 982 es la huella de destino introducida por el usuario.

65 En este modo de realización preferido, el siguiente proceso se ejecuta después de que la imagen de trama de destino 802b se descubra en el vídeo 801 (etapa S607; Sí).

La figura 16 es un diagrama de flujo parcial que muestra el contenido de un proceso añadido al proceso de búsqueda de vídeo en el diagrama de flujo mostrado en la figura 6. La explicación a continuación hace referencia a este dibujo.

5 Tal como se muestra en este dibujo, cuando se descubre la imagen de trama de destino 802b (etapa S607; Sí), el buscador 504 rastrea las imágenes de trama 802 en orden cronológico desde la imagen de trama de destino 802b para obtener la huella representativa 981 (etapa S621).

10 Además, se toma una determinación en cuanto a si la huella representativa 981 obtenida y la huella de destino designada mediante la flecha 982 por el usuario son similares o no (etapa S622).

15 En este caso, el procedimiento más sencillo de determinar la similitud entre la huella representativa 981 y la huella de destino es el siguiente. Es decir, la determinación es que las dos huellas son similares cuando la diferencia de ángulos formados por la orientación desde el punto de inicio hasta el punto final de la huella representativa 981 y la orientación desde el punto de inicio hasta el punto final de la huella de destino no es mayor que un valor de umbral prescrito, y que las dos huellas no son similares si se supera este valor de umbral prescrito. Además de esto, también es posible aplicar diversas tecnologías de determinación de similitud de curvas.

20 Cuando la huella representativa 981 y la huella de destino son similares (etapa S622; Sí), el proceso avanza a la etapa S608 y ese vídeo se emite como resultado de búsqueda.

Por otra parte, cuando la huella representativa 981 y la huella de destino no son similares (etapa S622; No), el proceso avanza a la etapa S609 y el proceso se repite con otro vídeo como objeto de proceso.

25 Con esta invención, es posible para el usuario designar fácilmente en qué posición está el objeto de destino, en qué orientación está situado el mismo y en qué dirección se mueve el mismo desde allí dentro de la trama de vídeo 331, por lo que es posible ejecutar fácil y rápidamente una búsqueda de vídeo.

30 En el ejemplo descrito anteriormente, se mencionó un vídeo obtenido filmando en el mundo real, usando una cámara, como un ejemplo del vídeo 801, pero también es posible aplicar el mismo procedimiento a un vídeo arbitrario, tal como un vídeo creado mediante gráficos de ordenador del estado de un mundo virtual visto desde un punto de vista virtual, o un vídeo animado obtenido creando y disponiendo imágenes de trama una cada vez mediante trabajo manual, y/o similares.

35 Modo de realización 5

En los modos de realización preferidos descritos anteriormente, las posiciones de referencia 311 y 312 y las posiciones de destino 351 y 352 se designaron introduciendo las flechas 371 y 372, pero también es posible usar otros parámetros cuando el usuario dibuja las flechas 371 y 372.

40 Por ejemplo, cuando se usa un panel táctil, un ratón, un controlador y/o similares, que es capaz de detectar la intensidad de la presión del usuario, es posible cambiar el tamaño de las proximidades de las posiciones de destino 351 y 352, según la intensidad de la presión del usuario cuando dibuja las flechas 371 y 372.

45 En este estado, la configuración de visualización de la pantalla 301 se fija de modo que si la fuerza de presión del usuario es fuerte en el ejemplo de entrada mostrado en las figuras 3 y 4, las flechas 371 y 372 se dibujan con trazo grueso, mientras que si la fuerza de presión del usuario es débil, las flechas 371 y 372 se dibujan con trazo fino.

50 Además, mediante la intensidad de la fuerza de presión del usuario, es decir, el grosor de las flechas 371 y 372, se determina el tamaño de un valor de umbral, que se refiere a si una posición está o no en las proximidades de la posición de destino.

Con este modo de realización preferido, es posible que la especificación detallada de condiciones de búsqueda de un vídeo se realice mediante operaciones sencillas.

55 Además de esto, tal como se muestra en la figura 3, cuando la trama de vídeo 331 se exhibe en una posición diferente de la imagen fija 101, es posible exhibir información diversa dentro de la trama de vídeo 331.

60 También existe un procedimiento de totalización de las cantidades de información local para las imágenes de trama 802 de todos los vídeos 801 registrados en la base de datos para cada posición dentro de la trama de vídeo 331, y hacer una presentación visual por pantalla dentro de la trama de vídeo 331 de la distribución de cada posición de los resultados totalizados.

65 En general, cuando la cantidad de información local es un valor escalar, cuanto mayor es ese valor, más fácil de ver es esa ubicación para una persona y más puede estimarse esta como una ubicación característica.

Además, cuando la cantidad de información local es un valor vectorial, cuanto mayor sea la longitud del mismo, más puede pronosticarse que esta ubicación sea característica.

5 Por consiguiente, haciendo que sea escalar la suma de las cantidades de información local, es posible obtener el grado para el que cada posición dentro de la trama de vídeo 331 es característica en el promedio de todos los vídeos.

10 Por consiguiente, con este procedimiento, el usuario pronostica más fácilmente con qué posición, como posición de destino, puede obtenerse un resultado de búsqueda más adecuado, y/o similares.

Esta solicitud reivindica el beneficio de la solicitud de patente japonesa nº 2012-010492, presentada el 20 de enero de 2012.

15 **Aplicabilidad industrial**

Con la presente invención, es posible proporcionar un dispositivo de búsqueda de vídeo, un procedimiento de búsqueda de vídeo, un medio de registro y un programa para efectuar búsquedas de vídeo de manera que un usuario pueda especificar fácilmente un objeto que debería aparecer en el vídeo, y la posición, la orientación y además el movimiento del mismo.

20 **Lista de números de referencia**

- 101 Imagen fija
- 25 102 Automóvil objeto de referencia
- 121 Región total de imagen fija
- 122 Región de referencia
- 30 123 Imagen de referencia
- 201 Imagen de trama
- 35 202 Automóvil objeto de destino
- 221 Región total de imagen de trama
- 222 Región de destino
- 40 223 Imagen de destino
- 301 Pantalla
- 45 311 Posición de referencia
- 312 Posición de referencia
- 331 Trama de vídeo
- 50 351 Posición de destino
- 352 Posición de destino
- 55 371 Flecha
- 372 Flecha
- 501 Dispositivo de búsqueda de vídeo
- 60 502 Receptor
- 503 Extractor
- 65 504 Buscador

	701	Imagen de referencia
	801	Vídeo
5	802	Imagen de trama
	803	Imagen local
	811	Posición de inicio
10	812	Posición de inicio
	901	Píxel significativo
15	902	Píxel significativo
	951	Huella de movimiento
	952	Huella de movimiento
20	971	Vector de dirección relacionado con la posición de referencia
	972	Vector de dirección relacionado con la posición de destino
25	981	Huella representativa
	982	Flecha relacionada con la huella de destino
30	991	Imagen resumida

**REIVINDICACIONES**

1. Un dispositivo de búsqueda de vídeo (501), que comprende:
  - 5 un receptor (502) para recibir la entrada de una imagen fija en la que se representa un objeto de referencia, siendo una primera posición de referencia y una segunda posición de referencia marcas del objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;
  - 10 un extractor (503) para extraer una imagen de referencia, que es una imagen extraída del objeto de referencia y que es una imagen representada en una región de referencia, cuya línea limítrofe divide el objeto de referencia de un fondo en la imagen fija, basándose en la primera posición de referencia y la segunda posición de referencia; y
  - 15 un buscador (504) para buscar imágenes de trama similares contenidas en vídeos y en las que se representan imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio, cuyas ubicaciones corresponden a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local, que es una región en la que se representa la imagen local similar a la imagen de referencia, en la imagen de trama similar, rastrear una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, que es un píxel situado en la primera posición de inicio en la imagen de trama similar, y una segunda huella de movimiento a lo largo de la cual un segundo píxel significativo, que es un píxel situado en la segunda posición de inicio en la imagen de trama similar, se desplaza en el vídeo, explorando imágenes de trama contenidas en el vídeo que incluyen la imagen de trama similar, desde la imagen de trama similar, y buscando una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino.
2. El dispositivo de búsqueda de vídeo (501) según la reivindicación 1, en el que:
  - 30 el buscador (504) está adaptado además para emitir como resultado de búsqueda un vídeo que incluye la imagen de trama similar y la imagen de trama de destino buscada.
3. El dispositivo de búsqueda de vídeo (501) según una cualquiera de las reivindicaciones 1 a 2, en el que:
  - 35 el receptor (502) recibe además la entrada de una huella de destino en una imagen de trama que incluye el vídeo; y
  - 40 el buscador (504), para cada uno de los vídeos buscados, rastrea una huella representativa a lo largo de la cual se desplaza un píxel representativo de una imagen local representada en la imagen de trama de destino, explorando imágenes de trama contenidas en el vídeo buscado, en el sentido cronológico desde una imagen de trama de destino incluida en el vídeo, y emite como resultado de búsqueda un vídeo en el que la huella de destino y la huella representativa son similares.
4. El dispositivo de búsqueda de vídeo (501) según una cualquiera de las reivindicaciones 1 a 2, en el que el buscador (504), para cada uno de los vídeos buscados, rastrea una huella representativa a lo largo de la cual se desplaza un píxel representativo de una imagen local representada en la imagen de trama de destino buscada, explorando imágenes de trama contenidas en el vídeo buscado, en el sentido cronológico desde una imagen de trama de destino incluida en el vídeo, y emite como resultado de búsqueda la huella representativa de esos vídeos junto con esos vídeos.
5. El dispositivo de búsqueda de vídeo (501) según una cualquiera de las reivindicaciones 1 a 4, en el que:
  - 55 el extractor (503) obtiene cantidades características de imagen para la imagen de referencia; y
  - el buscador (504) busca las imágenes de trama similares buscando imágenes de trama que tengan cantidades características de imagen similares a las cantidades características de imagen de la imagen de referencia.
6. El dispositivo de búsqueda de vídeo (501) según la reivindicación 5, en el que:
  - 60 el extractor (503) obtiene adicionalmente cantidades características de imagen de una imagen especular que es una inversión de la imagen de referencia; y
  - 65 el buscador (504) busca las imágenes de trama similares buscando imágenes de trama similares a las cantidades características de imagen de la imagen de referencia o a las cantidades características de

imagen de la imagen especular.

7. Un procedimiento de búsqueda de vídeo que comprende las etapas de:

5 recibir (S601) la entrada de una imagen fija en la que se representa un objeto de referencia, siendo una primera posición de referencia y una segunda posición de referencia marcas del objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

10 extraer (S602) una imagen de referencia, que es una imagen extraída del objeto de referencia y que es una imagen representada en una región de referencia, cuya línea limítrofe divide el objeto de referencia de un fondo en la imagen fija, basándose en la primera posición de referencia y en la segunda posición de referencia; y

15 buscar (S604) imágenes de trama similares contenidas en vídeos y en las que se representan imágenes locales similares a la imagen de referencia, encontrar (S605) una primera posición de inicio y una segunda posición de inicio, cuyas ubicaciones corresponden a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local, que es una región en la que se representa la imagen local similar a la imagen de referencia, en la imagen de trama similar, rastrear (S606) una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, que es un píxel situado en la primera posición de inicio en la imagen de trama similar, y una segunda huella de movimiento a lo largo de la cual un segundo píxel significativo, que es un píxel situado en la segunda posición de inicio, se desplaza en el vídeo, explorando imágenes de trama contenidas en el vídeo que incluye la imagen de trama similar, desde la imagen de trama similar, y buscar (S607) una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino.

8. El procedimiento de búsqueda de vídeo según la reivindicación 7, en el que:

30 el procedimiento de búsqueda de vídeo se ejecuta en un dispositivo de búsqueda de vídeo (501) que incluye un receptor (502), un extractor (503) y un buscador (504), comprendiendo el procedimiento la etapa adicional de:

35 emitir (S608) como resultado de búsqueda un vídeo que incluye la imagen de trama similar y la imagen de trama de destino buscada.

9. Un programa que hace que un ordenador funcione como:

40 un receptor (502) para recibir la entrada de una imagen fija en la que se representa un objeto de referencia, siendo una primera posición de referencia y una segunda posición de referencia marcas del objeto de referencia en la imagen fija, y una primera posición de destino y una segunda posición de destino dentro de una imagen de trama contenida en un vídeo;

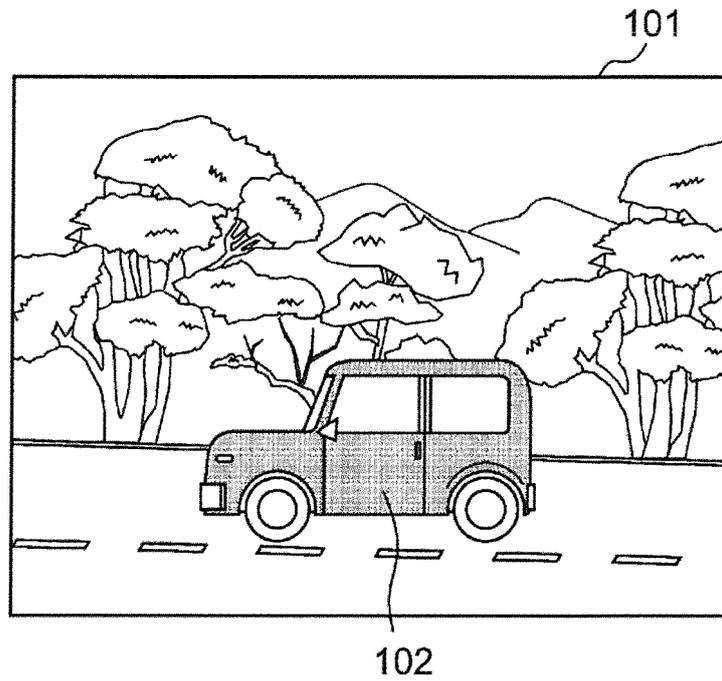
45 un extractor (503) para extraer una imagen de referencia, que es una imagen extraída del objeto de referencia y que es una imagen representada en una región de referencia, cuya línea limítrofe divide el objeto de referencia de un fondo en la imagen fija, basándose en la primera posición de referencia y en la segunda posición de referencia; y

50 un buscador (504) para buscar imágenes de trama similares contenidas en vídeos y en las que se representan imágenes locales similares a la imagen de referencia, encontrar una primera posición de inicio y una segunda posición de inicio, cuyas ubicaciones corresponden a la primera posición de referencia y a la segunda posición de referencia en la región de referencia, en una región local, que es una región en la que se representa la imagen local similar a la imagen de referencia, en la imagen de trama similar, rastrear una primera huella de movimiento a lo largo de la cual se desplaza un primer píxel significativo, que es un píxel situado en la primera posición de inicio en la imagen de trama similar, y una segunda huella de movimiento a lo largo de la cual un segundo píxel significativo, que es un píxel situado en la segunda posición de inicio en la imagen de trama similar, se desplaza en el vídeo, explorando imágenes de trama contenidas en el vídeo que incluye la imagen de trama similar, desde la imagen de trama similar, y buscando una imagen de trama de destino en la que la primera huella de movimiento alcanza las proximidades de la primera posición de destino y la segunda huella de movimiento alcanza las proximidades de la segunda posición de destino.

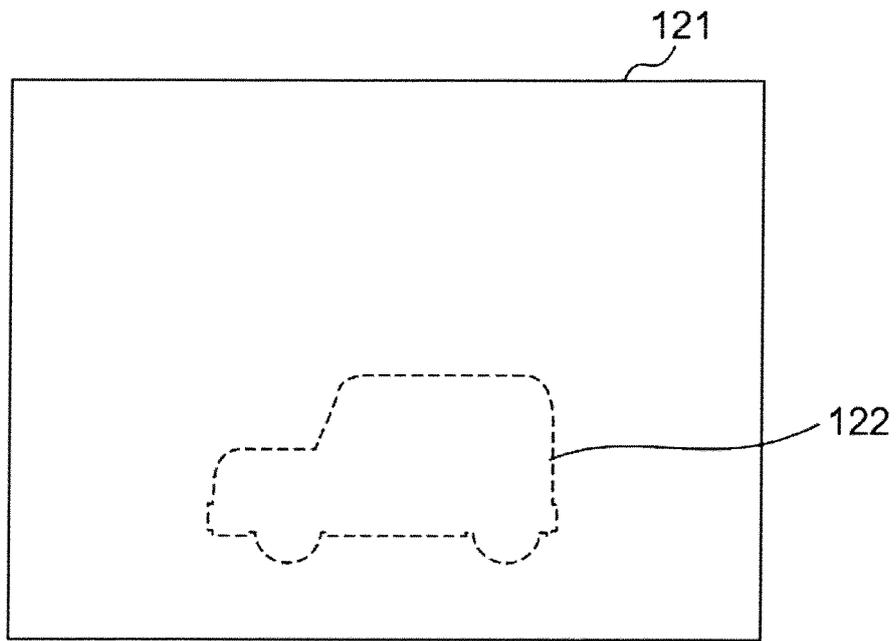
10. El programa según la reivindicación 9, en el que:

65 el buscador (504) está adaptado además para emitir como resultado de búsqueda un vídeo que incluye la imagen de trama similar y la imagen de trama de destino buscada.

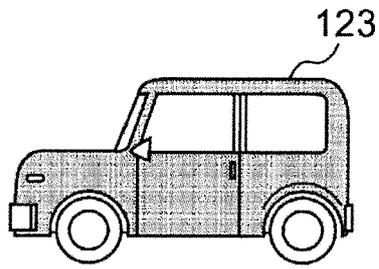
FIG.1A



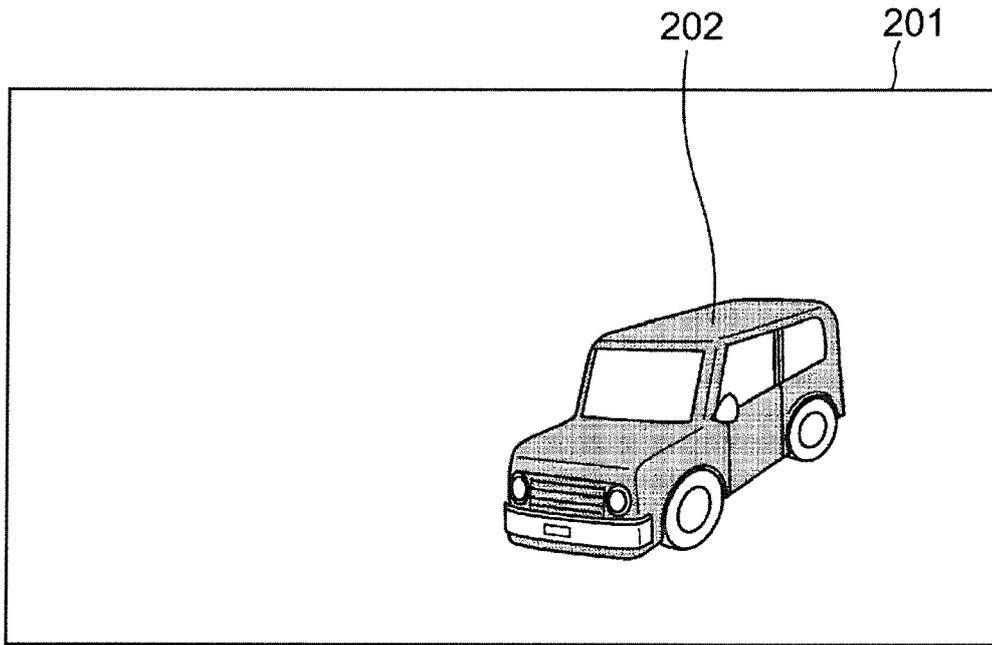
**FIG.1B**



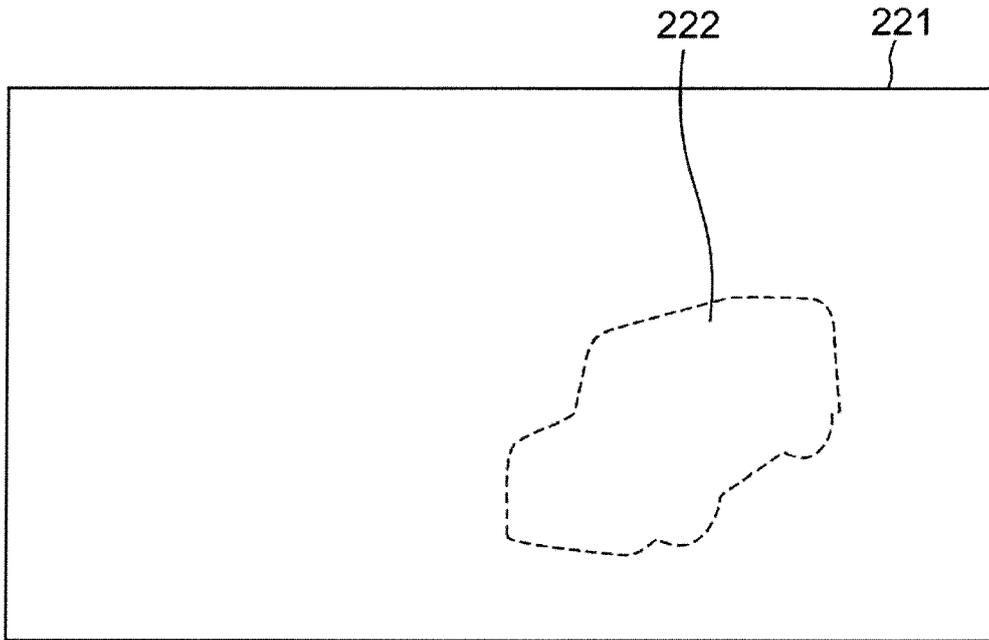
**FIG.1C**



**FIG.2A**



**FIG.2B**



**FIG.2C**

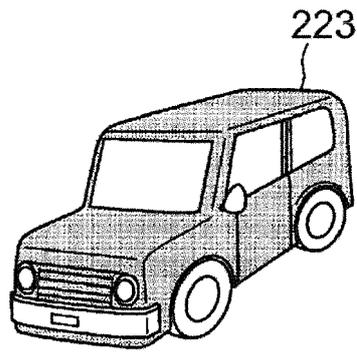


FIG.3

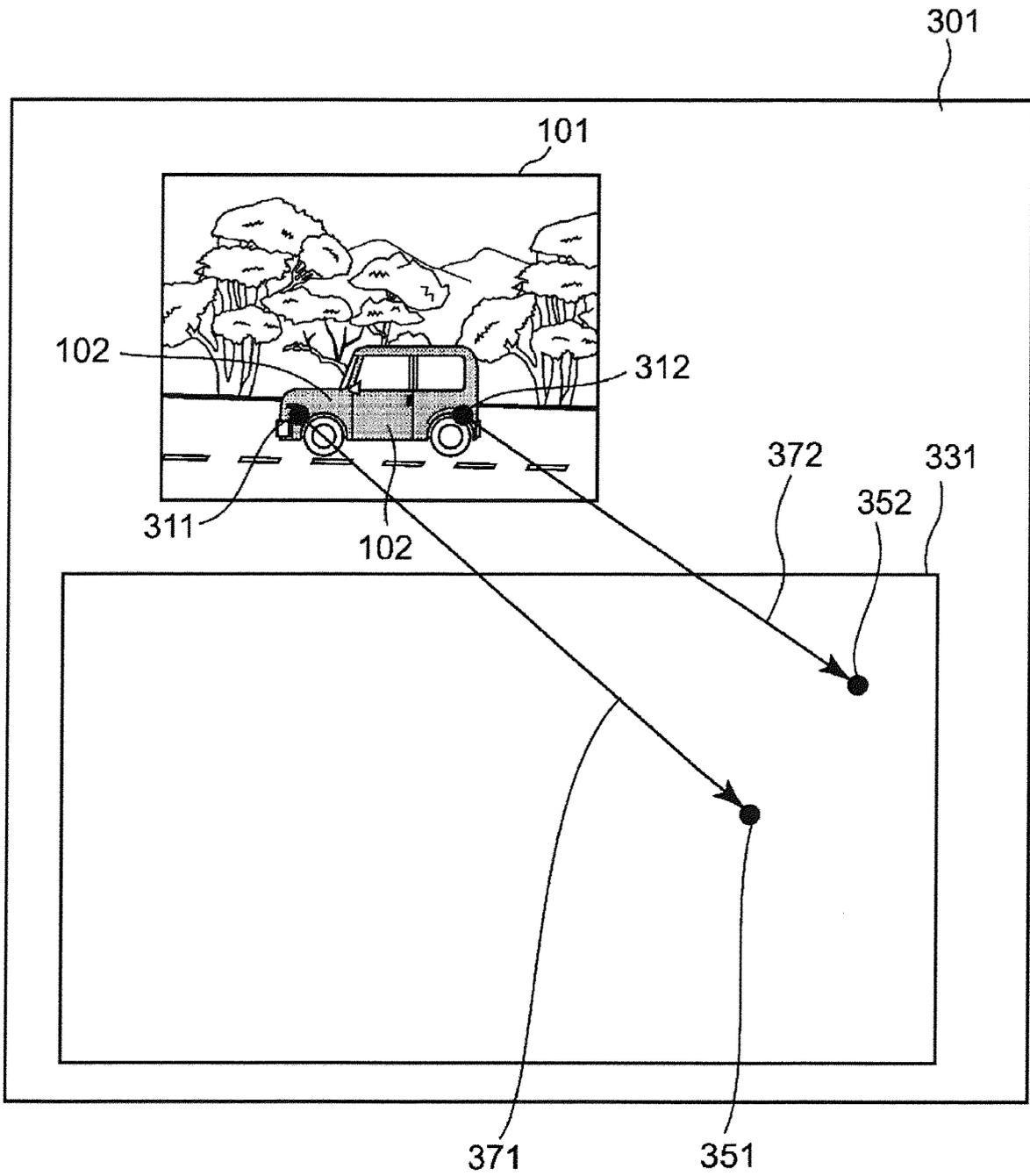
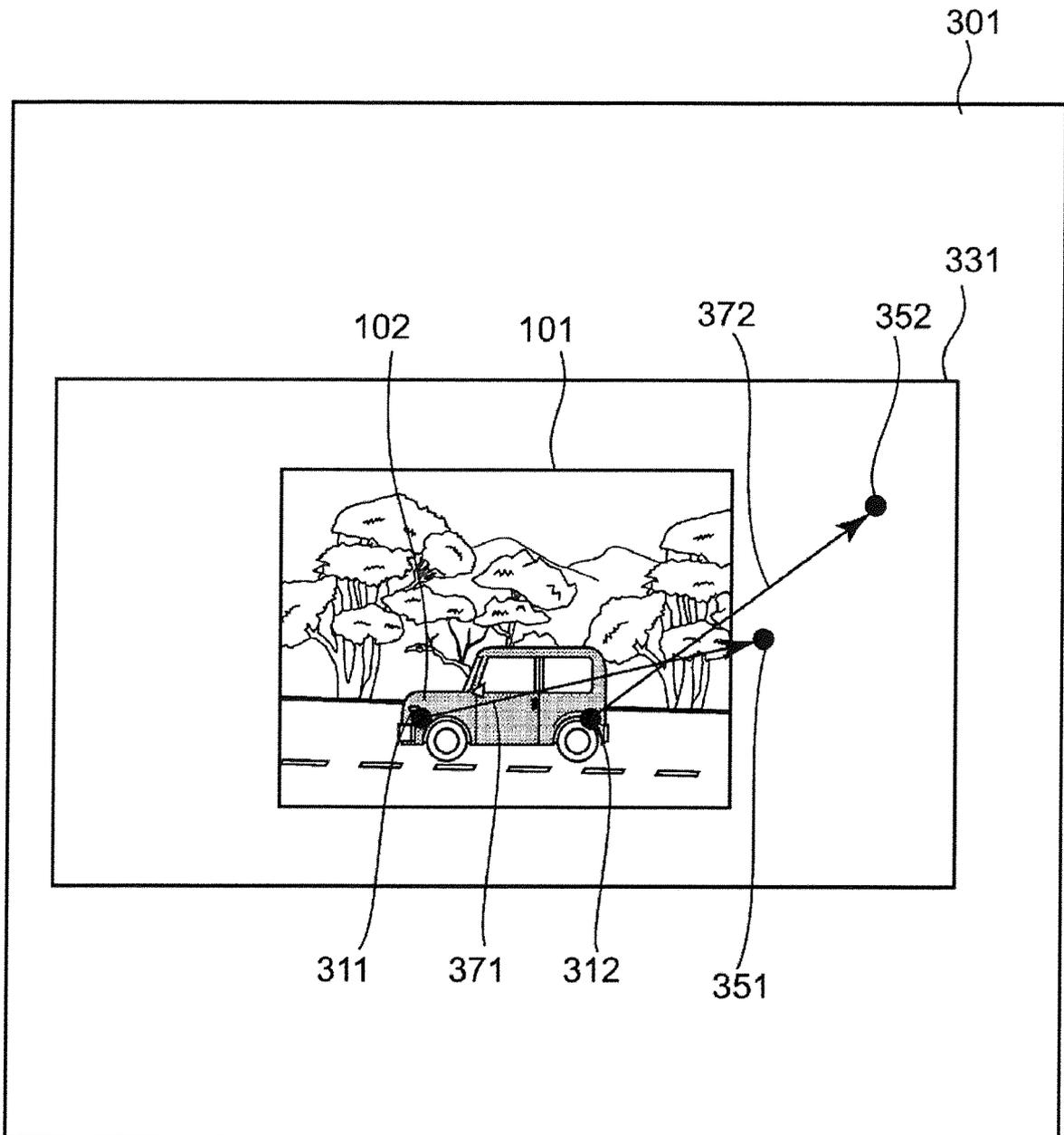


FIG.4



**FIG.5**

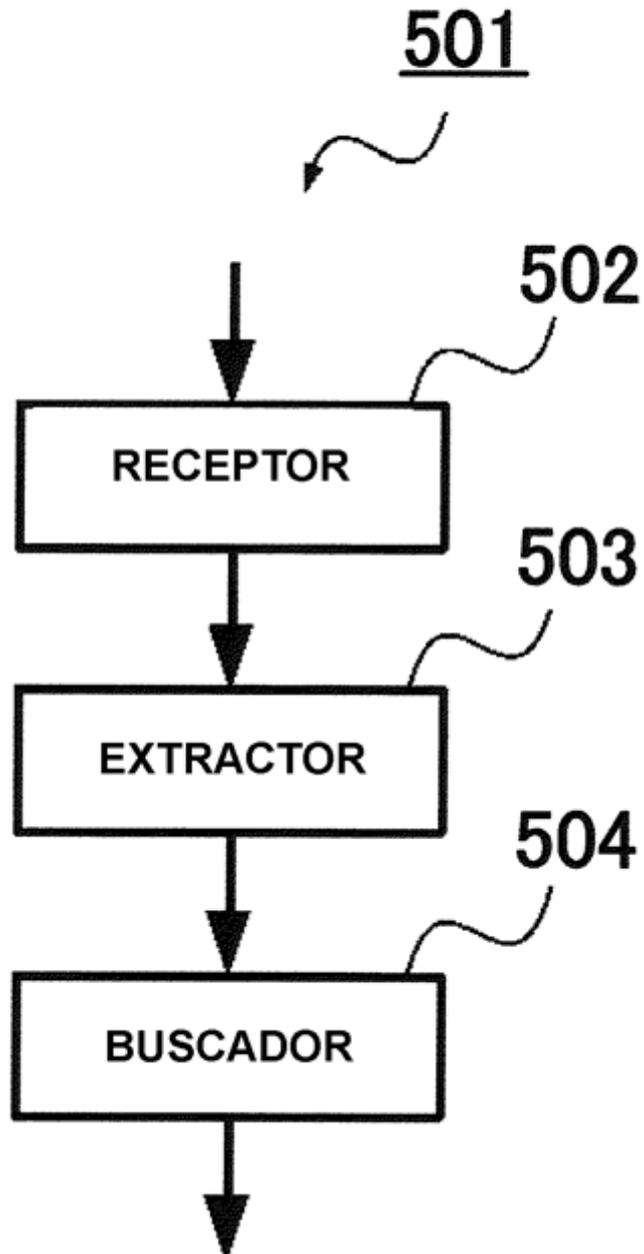
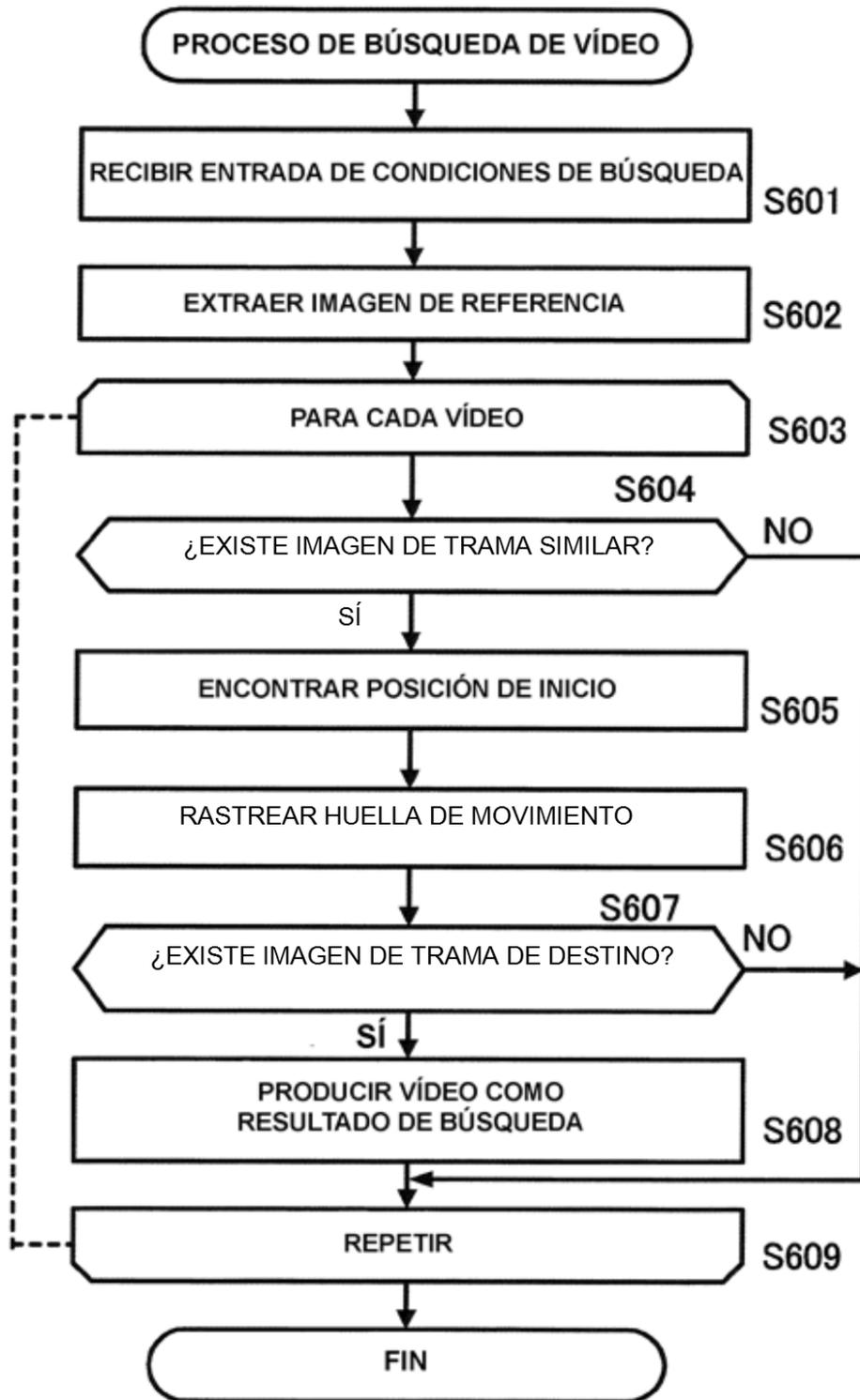


FIG.6



**FIG.7**

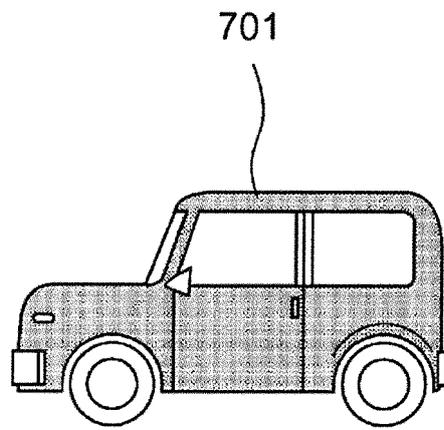


FIG.8

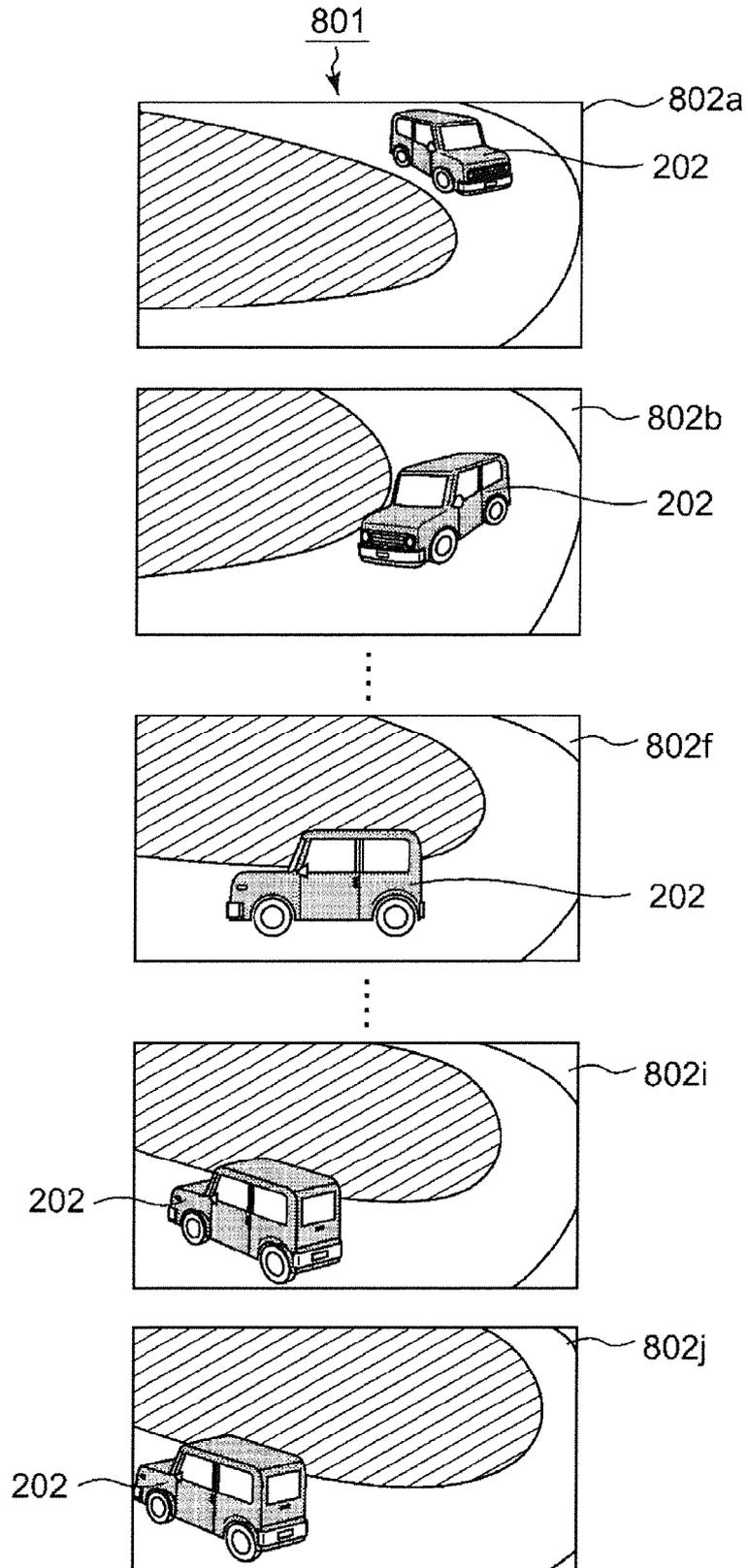


FIG.9

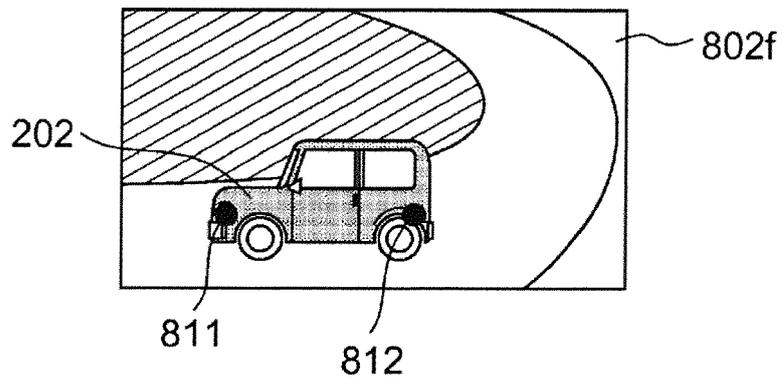
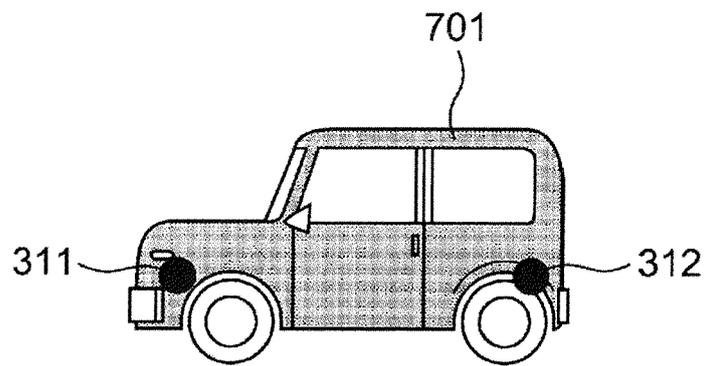


FIG.10

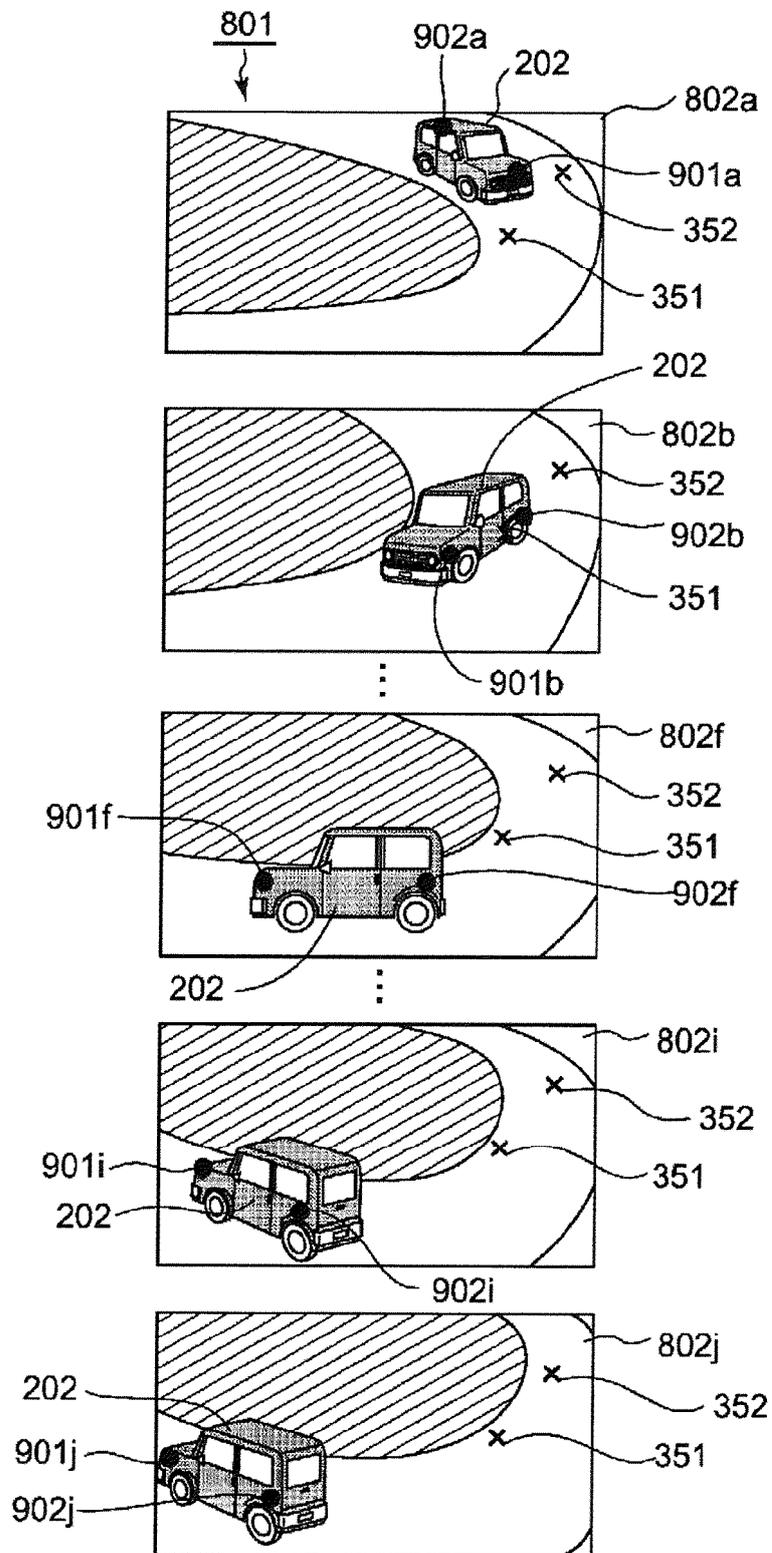


FIG.11

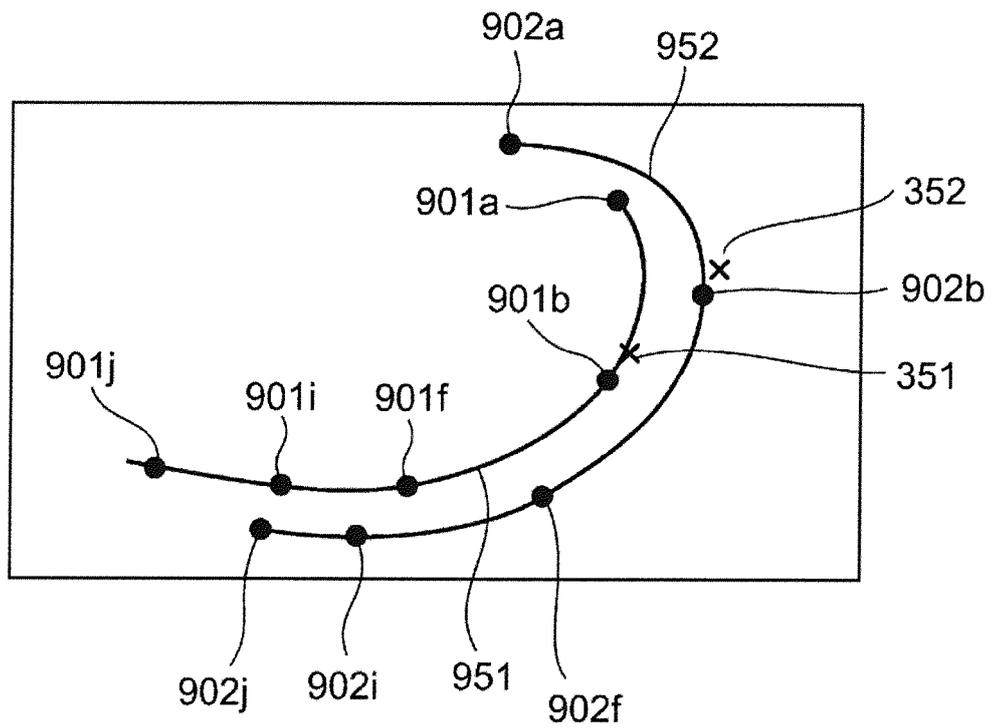


FIG.12

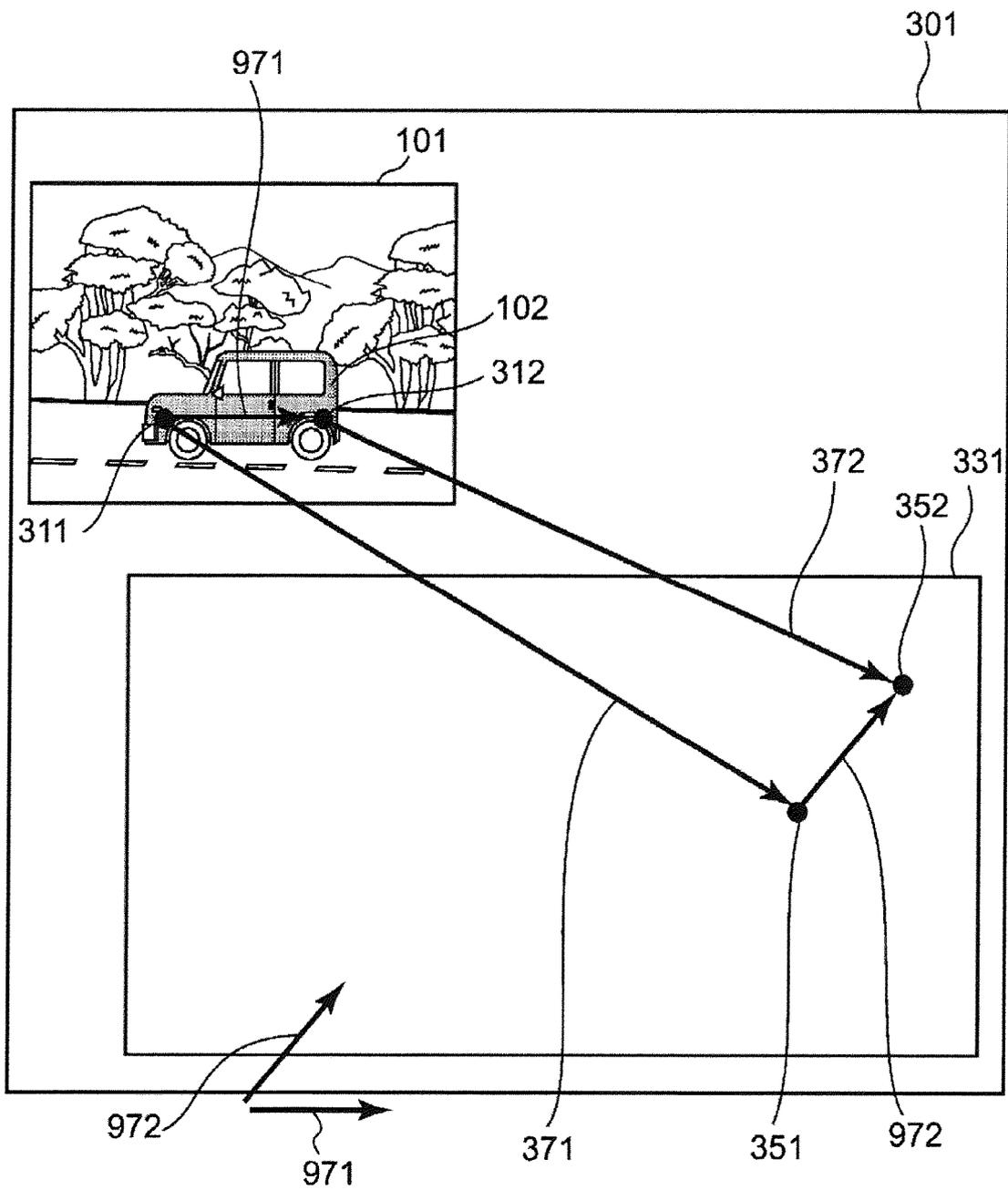


FIG.13

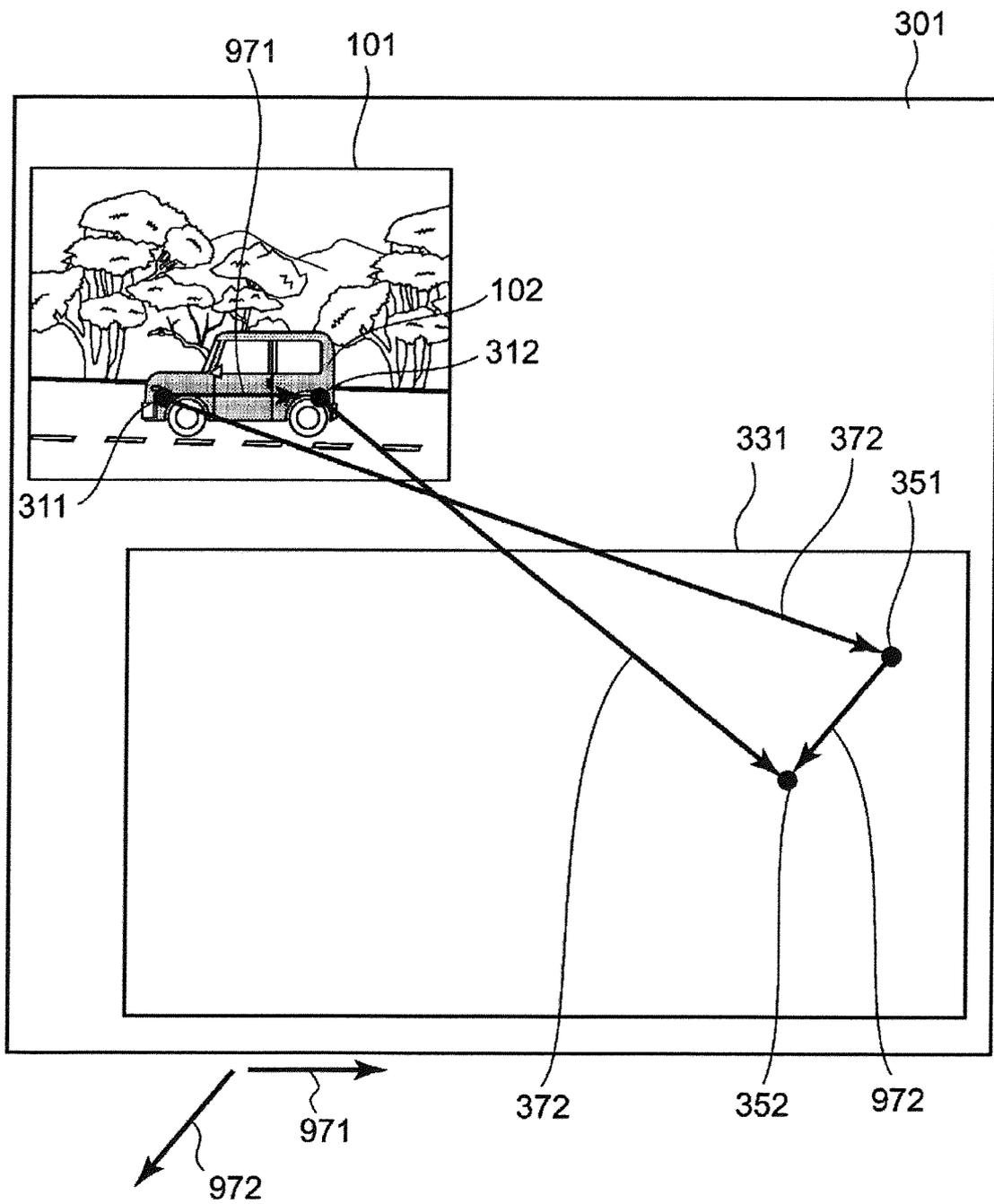


FIG.14A

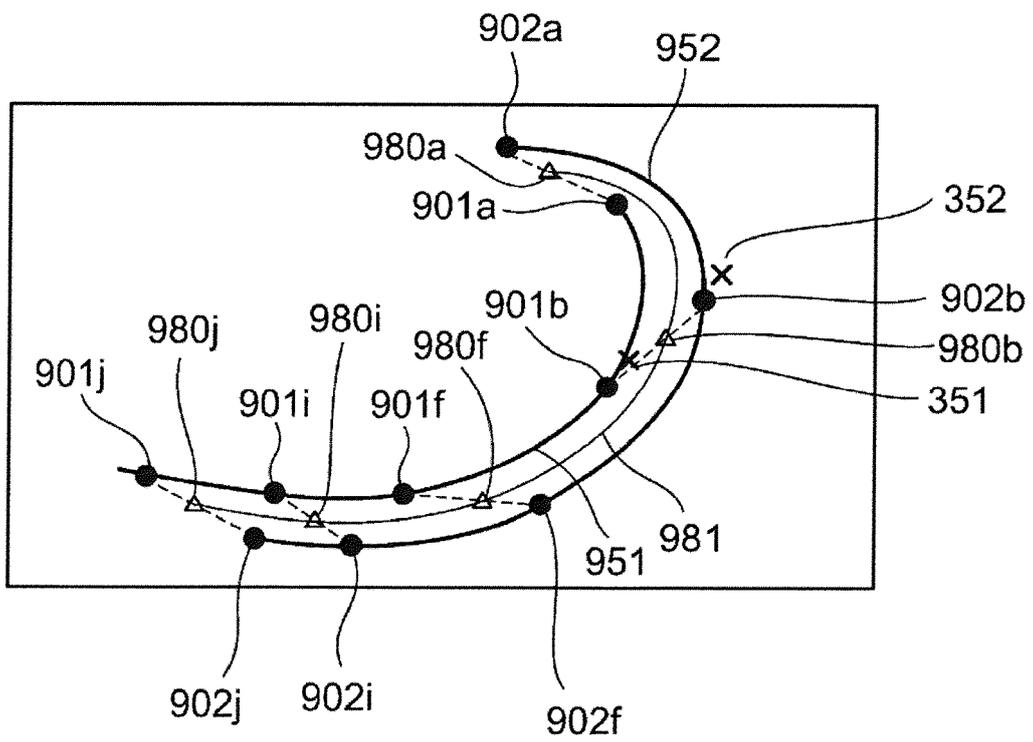


FIG.14B

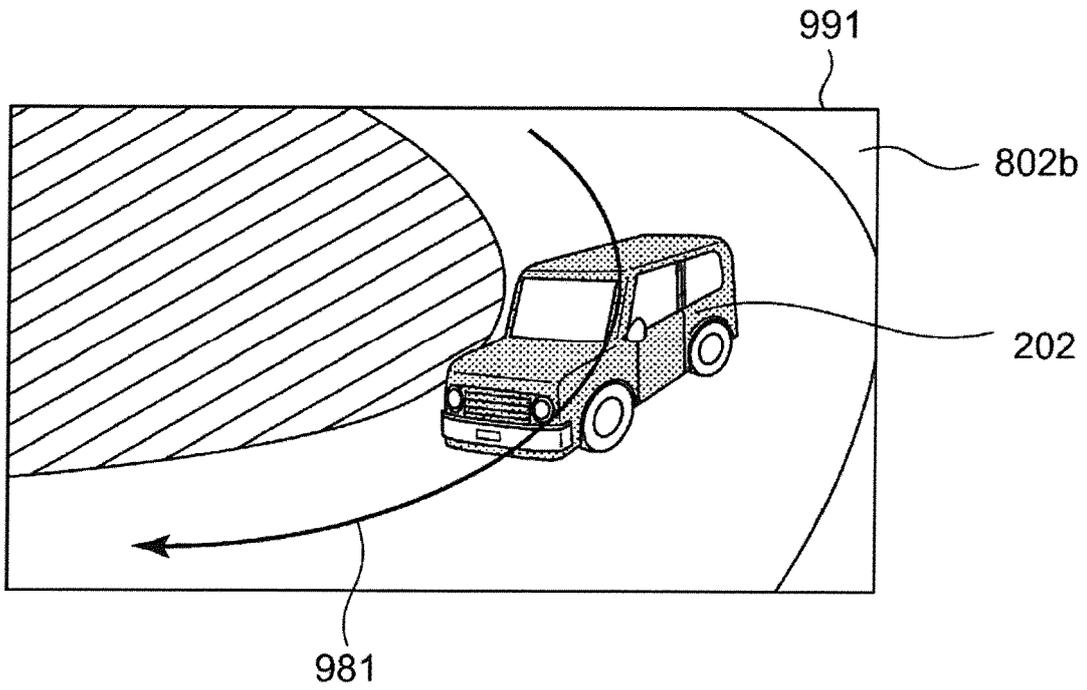


FIG.15

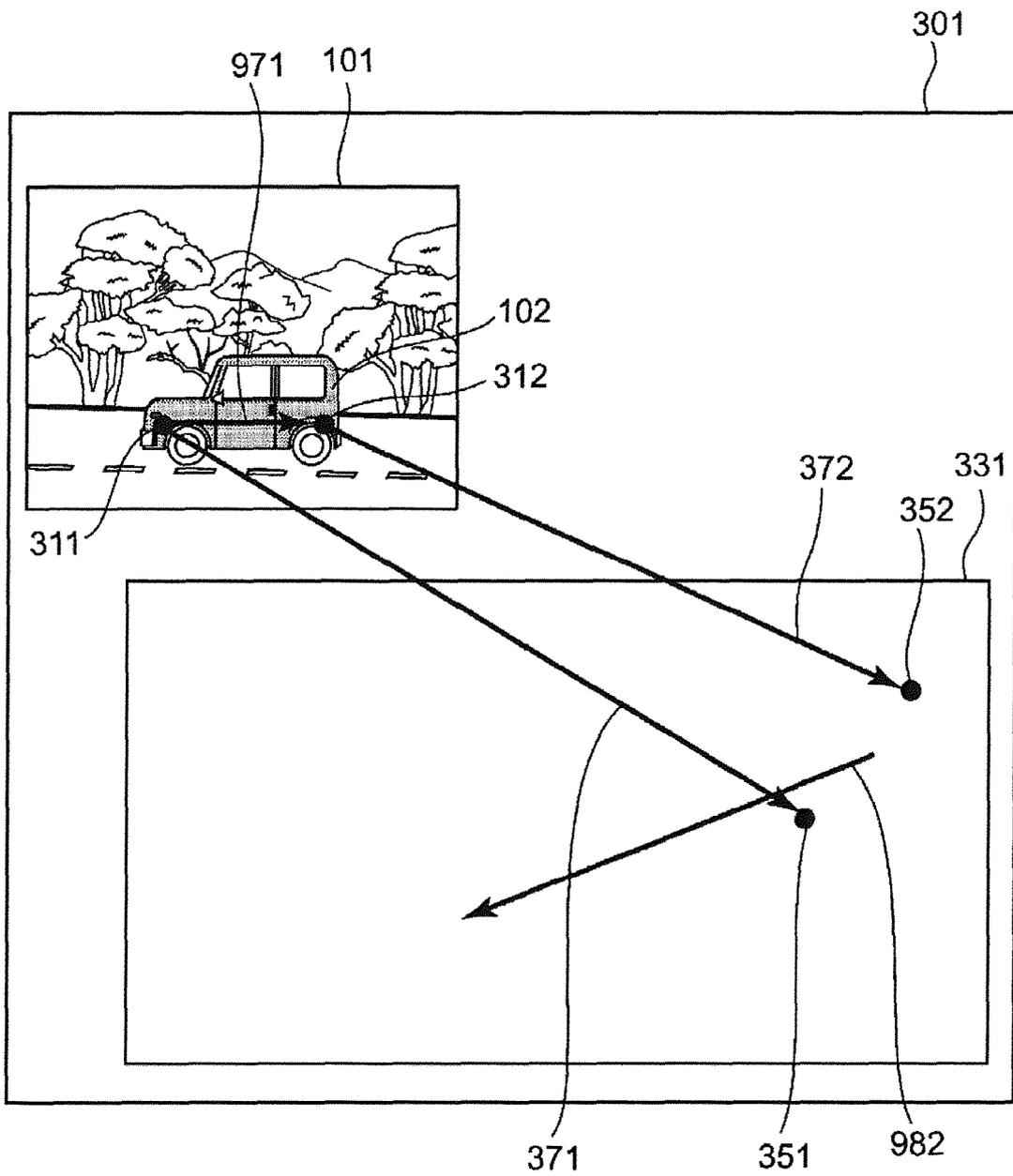


FIG.16

