

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 585 903**

51 Int. Cl.:

G06T 5/00 (2006.01)

G06T 7/00 (2006.01)

G06T 7/20 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.05.2007** **E 07720162 (2)**

97 Fecha y número de publicación de la concesión europea: **04.05.2016** **EP 2024937**

54 Título: **Método y sistema para generar una representación de una escena 3D dinámicamente cambiante**

30 Prioridad:

02.06.2006 EP 06405248

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

10.10.2016

73 Titular/es:

**LIBEROVISION AG (100.0%)
Technoparkstrasse 1
8005 Zürich, CH**

72 Inventor/es:

**WÜRMLIN, STEPHAN y
NIEDERBERGER, CHRISTOPH**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 585 903 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

Método y sistema para generar una representación de una escena 3D dinámicamente cambiante

Campo de la invención

5 La invención se refiere al campo del procesamiento de video y a la generación de imágenes virtuales, y es aplicable por ejemplo a la cobertura de deportes por televisión mediante reconstrucción basada en video en situaciones de juego en tres dimensiones. Se refiere a un método y un sistema para generar una representación 3D de una escena 3D dinámicamente cambiante, como se describe en el preámbulo de las reivindicaciones independientes correspondientes.

Antecedentes de la invención

10 El documento "A Video-Based 3D-Reconstruction of Soccer Games", T. Bebie y H. Bieri, EUROGRAPHICS 2000, Vol. 19 (2000), n.º 3, describe un sistema de reconstrucción diseñado para generar vistas en 3D virtuales, animadas, (tridimensionales) a partir de dos secuencias de video síncronas de parte de un juego de fútbol. Para crear una reconstrucción 3D de una escena dada, se ejecutan las siguientes etapas: 1) Se calculan los parámetros de cámara de todos los fotogramas de ambas secuencias (calibrado de la cámara). 2) Se extrae la textura del campo de juego
15 de las secuencias de video. 3) Se calculan las trayectorias de la pelota y las cabezas de los jugadores después de especificar manualmente sus posiciones de imagen en unos pocos fotogramas clave. 4) Se extraen las texturas de jugador automáticamente del video. 5) Se separan automáticamente las formas de jugadores que colisionan o se ocluyen. 6) Para la visualización, se mapea la textura de las formas de jugador sobre rectángulos colocados apropiadamente en un espacio virtual. Se supone que las cámaras permanecen en la misma posición durante toda
20 la secuencia de video que se está procesando.

El documento EP 1 465 115 A2 describe la generación de una vista deseada desde un punto de vista seleccionado. Se obtienen imágenes de escena a partir de varias cámaras con diferentes puntos de vista. Los objetos seleccionados se identifican en al menos una imagen, y se determina una estimación de la posición de los objetos seleccionados. Dado un punto de vista deseado, se determinan las posiciones de los objetos seleccionados en la
25 vista deseada resultante, y las vistas de los objetos seleccionados se interpretan usando los datos de imagen de las cámaras.

Se hace referencia adicional a las siguientes publicaciones de la técnica anterior:

- 30 - "Architecture and algorithms for tracking football players with multiple cameras". M. Xu, J. Orwell y D. Thirde; IEE Proceedings - Vision, Image, and Signal Processing - abril 2005 - Volumen 152, edición 2, pág. 232-241. Este documento describe la extracción de la posición de jugadores de fútbol y la posición 3D de una pelota, dadas por imágenes de video de cámaras estáticas. Sin embargo, no se aborda la síntesis de una representación 3D.
- 35 - "Multiple camera tracking of interacting and occluded human motion". Dockstader, S. y Tekalp, A. M.; 2001a. Proceedings of the IEEE 89, 1441-1455. Este documento se dirige a hacer un seguimiento de múltiples personas que se mueven e interactúan, dadas múltiples imágenes de cámara. No se aborda la síntesis de una representación 3D y la calibrado de la localización de la cámara y/u orientación.
- "Sports scene analysis and visualization from multiple-view video". Hideo Saito, Naho Inamoto, Sachiko Iwase; Int. Conf Multimedia and Expo (ICME) 2004: 1395-1398. Este documento describe una visualización desde un punto de vista libre de una escena de fútbol a partir de múltiples vistas de cámara.
- 40 - "Robust Tracking of Athletes Using Multiple Features of Multiple Views". Toshihiko Misu, Seiichi Gohsi, Yoshinori Izumi, Yoshihiko Misu, Masahide Naemura; Proceedings of WSCG'2004. Págs. 285-292. Para realizar el seguimiento de objetos móviles en imágenes 2D, se usa un número de programas de instalación para observar múltiples características tales como textura, color, movimiento y región, y la información de los diferentes programas de instalación se condensa entonces, ponderando la información adaptativamente según una fiabilidad proporcionada por cada programa de instalación.
- 45 - "Live mixed-reality 3D video in soccer stadium". Koyama, T. Kitahara, I. Ohta, Y.; Mixed and Augmented Reality, 2003. The Second IEEE and ACM International Symposium on mixed and augmented reality (ISMAR): 7-10 de octubre de 2003; páginas: 178-186. Se reconstruye un modelo 3D simplificado a partir de múltiples imágenes de video, y se proyecta la información de video seleccionada sobre el modelo. Una cámara especializada, instalada verticalmente por encima del campo de juego, se requiere para determinar la posición de un jugador. Se
50 requieren instrumentos de interrogación láser para calibrar las posiciones de las otras cámaras.
- "Immersive Observation of Virtualized Soccer Match at Real Stadium Model". Naho Inamoto, Hideo Saito; The Second IEEE and ACM International Symposium on mixed and augmented reality (ISMAR); 7-10 de octubre de 2003; páginas 188-197. Se genera una representación de una escena de fútbol virtual a partir de imágenes reales, y se muestra a una persona que lleva un dispositivo montado en la cabeza (HMD).

Descripción de la invención

Es un objeto de la invención crear un método y un sistema para generar una representación 3D de una escena 3D dinámicamente cambiante del tipo mencionado inicialmente, que proporciona una mejora sobre sistemas conocidos. En este contexto, la representación que se va a generar se entiende por sí misma como una representación 3D, es decir, un modelo de la escena que comprende información sobre la posición y orientación 3D de los objetos modelados. Esto es diferente de las representaciones 2D, en las cuales se manipula y analiza una imagen 2D, por ejemplo, identificación de objetos y seguimiento del movimiento de uno o más objetos, sin determinar la posición 3D del objeto.

Estos objetos pueden conseguirse por un método y un sistema para generar una representación 3D de una escena dinámicamente cambiante de acuerdo con las reivindicaciones independientes correspondientes.

El método para generar una representación 3D de una escena dinámicamente cambiante comprende las etapas de

- adquirir al menos dos emisiones de video, preferiblemente sincronizadas desde al menos dos cámaras localizadas en diferentes localizaciones, y que observan la misma escena 3D;
- determinar parámetros de cámara, que comprenden posición, orientación, y parámetros internos tales como, por ejemplo, ajuste de zoom, para dichas al menos dos cámaras;
- seguir el movimiento de objetos en las al menos dos emisiones de video;
- determinar la identidad de dichos objetos en las al menos dos emisiones de video;
- determinar la posición 3D de los objetos combinando la información desde al menos dos emisiones de video;

en el que al menos una de las etapas mostradas anteriormente depende de la información obtenida a partir de las al menos dos emisiones de video por una de las etapas posteriores.

Como resultado, se implementa un bucle de retroalimentación de información, es decir, una etapa particular en la secuencia de las etapas de procesamiento de información usa información obtenida en una o más etapas posteriores del proceso. Esta información típicamente se obtiene a partir de información de video que se generó en el instante anterior en el tiempo. Tal uso de una información, a priori, mejora la calidad y fiabilidad de los resultados de dicha etapa particular. En una realización preferida de la invención, la posición 3D de al menos un objeto se extrapola de al menos dos posiciones 3D anteriores del objeto.

En una realización preferida de la invención, la etapa de seguir el movimiento de objetos en las al menos dos emisiones de video usa información que se obtiene en una de las etapas posteriores de las al menos dos emisiones de video y parte de uno o más instantes anteriores en el tiempo.

En una realización preferida adicional de la invención, la información obtenida del uno o más instantes anteriores en el tiempo es la posición 3D del objeto.

Como resultado, la calidad y robustez del seguimiento 2D en una emisión de video mejora. Según el estado de la técnica, el seguimiento en una emisión de video solo está basado en la información disponible a partir de la propia emisión de video. Esto conduce a ambigüedades cuando dos objetos colisionan o solapan en una de las emisiones de video. Dada una posición 3D de un objeto y una información de calibrado de cámara, el movimiento de la proyección de los objetos en el espacio de imagen puede predecirse con mayor precisión usando retroproyección de la posición 3D de los objetos en la imagen. Además, es posible desambiguar entre diferentes interpretaciones de la misma imagen. Proyectando la posición 3D del objeto de vuelta a la imagen, el algoritmo de seguimiento puede detectar objetos que colisiona/solapan y mantiene la identificación correcta después de que ambos objetos se separen de nuevo.

En una variante preferida de la invención, la posición 3D de al menos un objeto se extrapola de al menos dos posiciones 3D anteriores del objeto. Esto se realiza, por ejemplo, por interpolación de primer orden o superior de la trayectoria anterior del objeto sobre los últimos fotogramas de la imagen. El término "objetos" en este punto y más adelante se refiere a objetos móviles de interés en la escena que son observados por la cámara. En un contexto deportivo, los objetos son, por ejemplo, los jugadores, una pelota y un árbitro. Otros elementos de imagen clave son el fondo, que es esencialmente estacionario, y que por un lado puede comprender la cancha o campo de juego, elementos característicos tales como líneas y otras marcas en la cancha, paredes y un entorno (por ejemplo un estadio) que rodea la cancha. Si se requiere, se denominarán "objetos de fondo". Los espectadores en el entorno, aunque están en movimiento, no se consideran como "objetos móviles" con el fin de esta solicitud. En general "2D" se refiere a las posiciones o formas de objetos o elementos en una imagen de video, mientras que "posiciones 3D" se refiere a la posición de un objeto en un modelo computacional tridimensional de una escena. En una realización preferidas de la invención, tal modelo 3D se mantiene y actualiza dinámicamente. La información de imagen o textura de video a partir de varias alimentaciones de video se interpreta sobre superficies definidas por este modelo 3D. Esto permite sintetizar vistas arbitrarias, dinámicas del modelo 3D, generando alimentaciones de video sintéticas

o virtuales adicionales.

5 En este punto y en el resto de la solicitud, el término "reproducción" se entiende que significa, como es habitual en gráficos informáticos, el proceso de producir los píxeles de una imagen a partir de una descripción de mayor nivel de sus componentes. En este caso, la descripción de mayor nivel es la combinación del modelo 3D y su información de textura asociada, y la reproducción se consigue mapeando la información de textura sobre el modelo 3D y determinando qué cámara virtual se ve en el modelo 3D.

10 En una variante preferida de la invención, el calibrado de la cámara se basa en el seguimiento de los elementos de referencia. Para inicializar el proceso de seguimiento, es necesario identificar, en cada entrada de video, un número de elementos de referencia, es decir, por asociación de cada uno de ellos con un identificador único. Cada elemento de referencia está asociado con un elemento de la escena (por ejemplo, posición, línea, esquina, círculo, etc.) en el espacio 3D. Dados estos elementos y la proyección de los elementos de referencia, como se ve en las imágenes de video, se determinan los parámetros de cámara y, en particular, la posición, orientación y ajuste de zoom y posiblemente otros parámetros ópticos de la cámara. Todo el conjunto de estos parámetros se denominará de aquí en adelante como "parámetros de cámara". La etapa de determinar estos parámetros se denomina calibrado de la cámara y, típicamente, se realiza para cada fotograma de cada entrada de video que se procesa. Para hacer esto, el movimiento de los elementos de referencia dentro de cada emisión de video puede seguirse por un método y sistema de seguimiento apropiado, como se conoce en la técnica.

20 En una variante preferida de la invención, se calcula una posición de imagen esperada del elemento de referencia en una de las imágenes de video a partir de la localización 3D conocida del elemento de referencia. Esto es particularmente útil cuando, después de mover o ampliar el zoom de la vista, un elemento de referencia que no era visible durante un tiempo resulta visible de nuevo: a partir del modelo 3D y dados los parámetros de cámara actuales, se determina que un elemento de referencia particular (con identidad conocida) debería ser visible en una posición esperada en el fotograma actual. La proximidad de la posición esperada se examina por extracción del elemento, y a un elemento mostrado en la imagen (y preferiblemente que es el mismo tipo que el elemento de referencia esperado) se le asigna automáticamente la identidad del elemento de referencia esperado.

En una variante preferida de la invención, el calibrado se consigue en base a las imágenes de video en solitario. La etapa de calibrado de la cámara puede conseguirse alternativamente determinando la posición, orientación y ajuste de zoom de la cámara por medios de medición (electro)mecánicos, o extrayendo los valores correspondientes de un sistema de control que controla estos parámetros.

30 En una variante preferida adicional de la invención, para cada emisión de video, a los elementos de referencia se les asigna su identidad en una imagen de referencia. En un escenario de juego típico, los elementos de referencia son líneas y otras marcas en la cancha. Un usuario los identifica en una imagen estática de un video

- apuntando, con un dispositivo de entrada gráfica, a una representación del mismo elemento de referencia en una representación esquemática de la cancha, y seleccionando dicha representación del elemento de referencia;
- 35 • apuntando, con un dispositivo de entrada gráfica, a un elemento de referencia particular, como se ve en la imagen estática del video y seleccionando dicho elemento de referencia; y
- asociando la identidad de la representación del elemento de referencia con el elemento de referencia visto en la imagen estática.

40 La acción de seleccionar una representación o elemento se efectúa mediante una orden o entrada de usuario, tal como por ejemplo clicar un botón del ratón o pulsar una tecla predeterminada después de apuntar a la representación o elemento. El orden de selección del elemento de referencia (primero o segundo) y su representación esquemática (segunda o primera) puede variar. La representación del elemento de referencia y la imagen de video pueden mostrarse en pantallas diferentes o dentro de ventanas diferentes en la misma pantalla. Este método para inicializar la asociación entre elementos de imagen y elementos del modelo permite un rápido ajuste del sistema.

En otra variante preferida de la invención, en la etapa de seleccionar dicho elemento de referencia en la imagen estática de video, se realizan las siguientes subetapas para determinar la posición exacta del elemento de referencia en la imagen estática del video:

- 50 • realizar automáticamente, en las proximidades de la posición seleccionada por el usuario, una extracción de elemento y, en particular, una extracción de líneas, intersecciones y esquinas;
- determinar la posición del elemento de referencia como la posición de uno de los elementos extraídos y, en particular, de un elemento cuyo tipo es el mismo que el seleccionado en la representación esquemática de la cancha.

55 Esto permite "ajustar" automáticamente la selección a la mejor posición del elemento, según se determina a partir del fotograma del video. Esto corrige pequeños errores de posición, realizados por el usuario cuando apunta al

elemento y, por lo tanto, simplifica y acelera el proceso de inicialización.

En otra realización preferida adicional más de la invención, la etapa de seguir el movimiento de los objetos comprende la etapa de incorporar parámetros de cámara dinámicamente cambiantes en la función de seguimiento, de manera que la función de seguimiento compense cambios en los parámetros de cámara. Esto significa que, para cada etapa de localización de un objeto seguido particular, se calcula no solo su posición esperada y preferiblemente también su tamaño, basándose en la posición y preferiblemente también su tamaño en fotogramas anteriores, sino que dicha posición y tamaño esperados se corrigen según los cambios conocidos de los ajustes de la cámara, incluyendo el zoom entre el fotograma de video anterior y actual. La corrección significa que se usa el movimiento de la cámara (paneo e inclinación) para determinar el cambio esperado de posición en la imagen, mientras la cantidad de zoom tiene influencia en el tamaño esperado del objeto. Esto conduce a una predicción más precisa del algoritmo de seguimiento, que aumenta el tiempo de procesamiento de todo el método de seguimiento.

Los parámetros de cámara tenidos en cuenta en la etapa de compensación o corrección de la función de seguimiento son cualquiera de

- parámetros de cámara determinados por una etapa de calibrado de cámara realizada para el mismo fotograma de video para el cual se realiza el seguimiento, o
- parámetros de cámara determinados por una etapa de calibrado de cámara realizada para uno o más fotogramas de video previos.

En el primer caso, la etapa de seguimiento tiene que esperar a que se complete el calibrado de la cámara, en el último caso, las etapas de seguimiento y calibrado pueden realizarse en paralelo. En el último caso, los parámetros de cámara opcionalmente pueden extrapolarse, por ejemplo por interpolación lineal o de mayor orden, a partir de dos o más ajustes anteriores de los parámetros de cámara.

En una variante preferida adicional de la invención, para inicializar identificaciones de objeto, se realizan las siguientes etapas:

- un usuario selecciona, en una primera imagen estática de la primera de las emisiones de video, un objeto y asigna a este un identificador único; y
- determinar automáticamente, en otra imagen estática de al menos otra emisión de video, un objeto cuya identidad es la misma.

Esta "asistencia de identificación automática" reduce el trabajo de identificar cada objeto en cada uno de los conjuntos de imágenes estáticas (uno para cada emisión de video, y bajo para la precondición de que todas las imágenes se toman en el mismo momento). Identificar un objeto en una imagen significa que un objeto, como se ve en una imagen, está asociado con ser un "árbitro" o una "pelota" o el "jugador Vroomfondel" u otro jugador, etc. Idealmente, la identificación de un objeto puede realizarse clicando sobre (o seleccionando de otra manera) un objeto solo en una de las imágenes estáticas. Suponiendo que todos los objetos están localizados a nivel del campo de juego, la posición 3D del objeto seleccionado en el campo de juego se determina intersecando el vector que apunta desde la cámara hacia la posición del objeto según se ve por la cámara con el plano del campo de juego. Cuando no se supone que los objetos están localizados al nivel del terreno, el punto 3D más cercano a todos dichos vectores puede calcularse como la posición del objeto. Desde esta posición, se calcula la posición esperada en el otro video estático. Para cada uno de los otros videos estáticos, si un objeto está en la posición esperada, entonces su identidad se ajusta para que sea la misma que la del objeto en la primera imagen estática. El sistema preferiblemente muestra la otra imagen estática que incorpora una representación visual de dicha identidad y permite al usuario confirmar o rechazar la asociación con dicha identidad. En los casos en los que los objetos solapan en una o más imágenes estáticas, en ocasiones puede detectarse automáticamente, por ejemplo si el área total, la forma o el histograma de color, etc., no es como se esperaba. En tales casos, se informa al usuario, y la identidad del objeto en la imagen respectiva se ajusta manualmente. En otros casos, puede ser necesario identificar el objeto en otro fotograma a partir de la misma emisión de video, en el cual en dicho otro fotograma el objeto es más fácil de separar de los otros objetos.

Puede darse una situación en la cual un objeto, típicamente un jugador, no es visible en ninguna de las entradas de video y, de esta manera, no se puede realizar su seguimiento nunca más. Como resultado, cuando el objeto reaparece en una de las emisiones de video, su identidad no se reconoce. En este caso, el sistema realiza las etapas de:

- alertar al usuario de la presencia de un objeto móvil no identificado; y
- permitir al usuario asociar un identificador con el objeto.

La etapa de detectar un objeto móvil comprende detectar la presencia de movimiento significativo entre fotogramas (compensando los cambios en los parámetros de cámara, como se ha esbozado anteriormente, descontando objetos que ya se han seguido, las áreas restantes de movimiento pueden corresponder al objeto de reentrada.

Después de superar un umbral predeterminado respecto al tamaño del objeto y, por ejemplo, criterios de coincidencia con un modelo estadístico del conjunto de objetos conocidos, se indica al usuario la presencia del objeto.

5 En otra variante preferida de la invención, no es necesario mantener la identidad completa de un objeto, es decir, no es importante conocer que el objeto X representa al "jugador Vroomfondel". Es suficiente con saber que el objeto es de una categoría particular, por ejemplo, un jugador del equipo A, que puede determinarse automáticamente a partir de los colores del objeto o simplemente un jugador. Para este escenario, un objeto móvil identificado preferiblemente se asocia automáticamente con un identificador único generado por el sistema. Esto permite generar una reconstrucción y representación 3D continua, sin intervención manual.

10 En el transcurso del seguimiento del movimiento, los objetos comúnmente se asocian con "cuadros de delimitación", que son áreas de píxeles rectangulares en las cuales se sabe o se espera que esté situado el objeto. En un sentido más general, un cuadro de delimitación puede reemplazarse por una silueta de diferentes formas alrededor del objeto. Los cuadros de delimitación (o siluetas) preferiblemente se generan por seguimiento y refinado mediante la última etapa de segmentación, como se explica más adelante.

15 Alternativamente, en una realización preferida adicional de la invención, una o más cámaras de visión general se sitúan para cubrir siempre todo el campo de juego. No se mueven y no utilizan zoom. La vista desde esas cámaras permite seguir todos los objetos, siempre y cuando no dejen el campo, y elimina la necesidad de identificar objetos cuya identidad se ha perdido. Esta cámara puede ser una cámara de menor coste que las cámaras usadas para generar las imágenes de TV que finalmente se difunden.

20 La etapa de localizar la posición de los objetos móviles puede conseguirse también o estar asistida por la incorporación de uno o más transpondedores de RF (radiofrecuencia) en los objetos, y midiendo sus posiciones con el sistema de localización RF. De esta manera, la identidad y posición del objeto se conocen con precisión en cada momento temporal.

25 En una variante preferida de la invención, los objetos se clasifican como pertenecientes a una de al menos dos categorías. Las categorías preferiblemente están basadas en un modelo estadístico, tal como un modelo gaussiano mixto, y comprenden al menos dos del primer equipo, segundo equipo, primer portero del equipo, segundo portero del equipo, pelota y árbitro. Los parámetros incorporados por el modelo estadístico preferiblemente son el color o colores de los objetos. Se sabe por ejemplo que el Equipo A viste de un primer conjunto de colores, el Equipo B de un segundo conjunto, los porteros de los equipos tienen diferentes colores que los de ambos equipos, y que el árbitro predominantemente viste de negro u otro color, y el verde de fondo, blanco y una diversidad de otros colores (color del césped, marcas, postes de gol y espectadores). De esta manera, la imagen se segmenta no solo separando los objetos del fondo, sino que los objetos se clasifican en diferentes conjuntos. El modelo estadístico preferiblemente se genera a partir de una imagen estática de una cámara y después se aplica a las emisiones de video de todas las cámaras. El modelo estadístico se genera, para cada categoría de objetos, cuando el usuario mueve, mediante un dispositivo de señalización, una marca de referencia a lo largo de una trayectoria sobre una diversidad de puntos que pertenecen a dicha categoría. Los colores de los puntos en dicha trayectoria forman una muestra que representa dicha categoría en la generación del modelo estadístico.

35 Cuando se segmentan los fotogramas de video en cada una de las entradas de video, la separación de los objetos del fondo y la distinción de un objeto comprende preferiblemente la etapa de

- 40
- usar un canal alfa coincidente para asignar, a cada elemento de la imagen o píxel, un valor que expresa la probabilidad con la cual el píxel es parte de un objeto o parte del fondo.

45 Tal valor se conoce como valor Alfa, y se almacena por ejemplo como información de canal alfa asociada con la imagen o emisión. Como resultado, no hay límites estrictos (binarios) alrededor de los objetos, lo que mejora la calidad de las últimas etapas de reproducción en las que las imágenes objeto se combinan y/o mapean sobre una superficie 3D.

50 En una realización preferida adicional de la invención, la etapa de segmentación comprende, después de recortar los objetos del fondo, llenar los huecos correspondientes que quedan en el fondo mediante restauración de la imagen, véase por ejemplo "View Interpolation for Image Synthesis", Chen y Williams, ACM SIGGRAPH 1993, págs. 279-288. Tales elementos de imagen restaurada se marcan como datos de imagen sintéticos. Esto se realiza ajustando un bit correspondiente para cada píxel restaurado en la imagen llena o asociando la entrada de video con un canal adicional que comprende dicha información de marcado. Esta información adicional permite, en un momento posterior en el método inventivo, ignorar los datos de imagen sintéticos para elementos de imagen para los cuales están disponibles datos de imagen reales.

55 Por favor, obsérvese que los términos "vista sintética", "síntesis de imagen", etc. - a diferencia de "datos de imagen sintéticos" - se refieren a algo diferente, en concreto a las imágenes virtuales o vistas virtuales de una escena 3D que se generan de acuerdo con la invención, y que están basadas tanto en datos de imagen reales como en datos de imagen sintéticos.

De esta manera, "los datos de imagen reales" relacionados con una localización de superficie o parche de superficie en la escena 3D típicamente es información de imagen a partir de una o más emisiones de video que proceden de la observación de este parche en el instante de tiempo en el que se representa. En contraste, los "datos de imagen sintéticos" son información de imagen generada, por ejemplo

- 5 - por restauración de imagen mediante técnicas 2D, tal como llenando un hueco, partiendo en el límite, con colores o diseños encontrados en el límite;
- usar datos de imagen procedentes del mismo parche de superficie pero obtenidos a partir de una emisión de video en otro instante de tiempo;
- 10 - usar datos de imagen procedentes del mismo parche de superficie pero obtenidos de otra emisión de video (en el mismo o en otro instante de tiempo);
- usar datos de imagen obtenidos de una emisión de video en el mismo instante de tiempo pero procedentes de otro parche de superficie;
- restaurar una imagen por técnicas 3D, tal como llenando un hueco con diseños de la escena que rodea el hueco mientras se considera la localización 3D de los parches de superficie llenos y usados para el llenado.

15 Se presentan ahora los modos preferidos de generación de datos de imagen sintéticos (fondo), que típicamente combinan aspectos de los dos últimos puntos. Tal restauración de imagen 3D usa no solo partes de imagen de la cercanía de un hueco para llenar el hueco, sino que también tiene en cuenta el conocimiento sobre relaciones espaciales (es decir, 3D) entre parches de superficie. Es decir, cuando se usan datos de imagen de un "parche de origen" para llenar datos de imagen en un "parche de destino", entonces se considera la relación espacial entre
20 estos parches, que puede implicar rotación y/o una perspectiva diferente.

El objetivo es llenar los huecos con colores o texturas que se parezcan a la imagen esperada real tanto como sea posible. Un enfoque conocido es elegir un parche de destino que se extienda sobre el límite de un hueco, de esta manera comprende partes conocidas y desconocidas de la imagen. Se realiza entonces una búsqueda para encontrar un parche de origen, con una parte del parche de origen que coincide con las partes conocidas. La parte restante del parche de origen se usa después para llenar las partes desconocidas por copiado. Sin embargo, esta búsqueda y copiado solo tiene lugar en el plano de una imagen de video.
25

De acuerdo con la invención, se usa la posición y/u orientación conocidas de los parches de origen y destino en el espacio 3D cuando se buscan parches de origen y cuando se copian datos de imagen de origen a un parche de destino.

30 En una realización preferida adicional de la invención, la información sobre la localización de marcas se usa para guiar la restauración de la imagen. Las marcas son elementos característicos inmóviles, típicamente elementos que se extiende linealmente, tal como líneas rectas o circulares sobre el campo de juego, cuya localización nominal es conocida. Este enfoque incorpora preferiblemente conocimiento a partir de un modelo de campo que comprende la localización y orientación de las líneas en el campo de juego. Preferiblemente, esto se combina con la información
35 3D sobre la localización de parches a lo largo de la marca.

Por ejemplo, se implementan uno o más de los siguientes enfoques:

- si un parche de destino comprende una parte de una marca, entonces los parches de origen se buscan a lo largo de la marca y se mapean sobre localizaciones dentro del hueco en el cual se espera que esté la marca;
- 40 - si un parche de destino está situado en un círculo, entonces el parche de destino se transforma (por ejemplo, se rota y cambia de escala, típicamente sin preservar su relación de aspecto) antes de compararlo con un parche de origen potencial a lo largo del círculo o viceversa. Igualmente, el parche de origen se transforma cuando se copia al parche de destino.
- Si un parche de destino está situado lo largo de una línea recta, entonces el parche de destino se cambia de escala (típicamente preservando su relación de aspecto) antes de compararlo con un parche de origen potencial
45 más distante - desde el punto de vista de la cámara - a lo largo de la línea. Igualmente, el parche de origen se cambia de escala de vuelta cuando se copia al parche de destino.

La restauración de la imagen preferiblemente se realiza individualmente para varias emisiones de video, dando como resultado un conjunto de emisiones de video que muestran la misma escena de fondo, sin los objetos (móviles) desde diferentes perspectivas. En un último punto de procesamiento, estas imágenes de fondo de imagen restauradas se usan para colorear o "pintar" el modelo de fondo 3D. Sin embargo, en otra realización de la invención, la restauración de imagen tiene lugar cuando el modelo de fondo 3D está coloreado: el modelo de fondo se colorea en primer lugar de acuerdo con la información de imagen de fondo de una o más emisiones de video, y después se pinta cualquier parche no coloreado restante en el modelo 3D, de una manera similar a la descrita anteriormente, es decir, con información de imagen de otras localizaciones o instantes en el tiempo.
50

El método de llenado de huecos descrito anteriormente puede implementarse independientemente de los otros aspectos de la invención descrita en la presente memoria, en particular independientemente de las etapas de seguimiento e identificación del objeto 3D, calibrado de la cámara, etc....

5 Preferiblemente, la etapa de segmentación comprende la etapa de refinar la posición y dimensiones de los cuadros de delimitación. Esto significa que los cuadros de delimitación alrededor de los objetos, cuando estos se producen por el seguimiento del objeto, se adaptan de acuerdo con la información de segmentación: esto tiene sentido puesto que la calidad de los algoritmos de seguimiento convencionales está limitada en tanto que los cuadros de delimitación que se crean alrededor de objetos móviles a menudo tienden a quedar detrás de la posición del objeto real o tienden a aumentar de tamaño. Por lo tanto, el cuadro de delimitación está adaptado para contener el objeto
10 después de la segmentación de la imagen. La posición del objeto, que preferiblemente está definida como la parte media inferior (o alternativamente, una predeterminada de las esquinas del cuadro de delimitación) en consecuencia se adapta también.

En una variante preferida de la invención, se proporciona una vista sintetizada que muestra la escena desde un punto de vista habitual que es distinto de las posiciones de las cámaras reales. Esta comprende las etapas de:

- 15
- proporcionar parámetros de cámara de una cámara virtual;
 - determinar una imagen de fondo como vista por la cámara virtual;
 - determinar una proyección de cada uno de los objetos en la cámara virtual y superponerlos en la imagen de fondo;
 - producir o almacenar la imagen combinada para almacenamiento o para procesamiento adicional.

20 Como resultado, se mantiene y actualiza dinámicamente una representación 3D global de toda la escena, basándose en la entrada desde las emisiones de video. La información de video o imagen a partir de las emisiones de video se reproduce en la representación 3D, permitiendo generar imágenes desde los puntos de vista que difieren de las localizaciones físicas reales de las cámaras que proporcionan las entradas de video de origen. Esto está en contraste con la técnica anterior, donde solo se colocan rectángulos aislados en el espacio 3D y se mapea la
25 textura de las imágenes en estos rectángulos, pero sin considerar la cancha y los objetos de fondo restantes.

En este punto y en el resto de la solicitud, el término "textura" significa datos de imagen o dibujo desde la superficie de un objeto real (o, alternativamente, de uno generado por ordenador). Los datos de textura, por ejemplo, se observan mediante una o más cámaras o se recuperan de una base de datos, y pueden transformarse geoméricamente y reproducirse mapeándolos sobre la superficie de un objeto en el modelo 3D.

- 30 En una variante preferida adicional de la invención, la etapa de determinar una imagen de fondo como se ve por la cámara virtual comprende las etapas de
- combinar, para cada elemento de la imagen de fondo, información de imagen desde las diferentes emisiones de video que corresponden a la misma localización de fondo;
 - dar prioridad a la información de imagen que no está marcada como datos de imagen sintéticos; y
- 35
- reproducir la información de imagen sobre un modelo de fondo que comprende una o más superficies que representan el fondo, es decir, objetos de fondo.

Como resultado, siempre y cuando estén disponibles datos de imagen real, el fondo, estará coloreado con una combinación de colores de los datos reales disponibles, puesto que a los datos reales se les da prioridad sobre los datos sintéticos. Solo cuando no están disponibles datos reales el fondo en el modelo 3D se coloreará con datos
40 sintéticos de uno o más de los fondos de las diferentes entradas de video después de la segmentación.

En una variante preferida de la invención, la superficie que representa el fondo es una superficie que representa el campo de juego o cancha y, opcionalmente, también comprende superficies que representan un modelo de entorno, almacenado por ejemplo, en un módulo de datos de entorno. El modelo de entorno puede ser un simple modelo genérico que se aplica a cualquier entorno, o un modelo de entorno derivado de datos de CEA del entorno real. En
45 otra variante preferida de la invención, la etapa de terminar una imagen de fondo como se ve por la cámara virtual comprende además reproducir datos de imagen predeterminados en el modelo de fondo, superponerlos sobre o reemplazar la información de imagen proporcionada por las emisiones de video.

En otra variante preferida de la invención, la etapa de determinar una proyección de cada uno de los objetos en la cámara virtual comprende las etapas de determinar la posición y orientación de un objeto de reproducción, y
50 después reproducir la información de video extraída de las diferentes entradas de video y asociado con este objeto sobre el objeto de reproducción.

En una realización preferida de la invención, el objeto de reproducción es lo que se denomina valla publicitaria, es

decir, un plano vertical situado en la cancha. El tamaño y orientación de la valla publicitaria se determina de acuerdo con el cuadro de delimitación para este objeto desde una de las entradas de video. La localización de la valla publicitaria está definida por la posición 3D calculada del objeto descrito anteriormente. La posición 3D se combina entonces con la proyección del cuadro de delimitación para producir cuatro vértices 3D de un rectángulo tridimensional. Para ello, la normal del rectángulo es igual al eje óptico o normal al plano de visión definido por la cámara real desde la cual se originó el cuadro de delimitación, o por la cámara virtual. En el primer caso, pueden usarse múltiples vallas publicitarias para cada objeto, cada una de las cuales corresponde a una cámara real. La imagen reproducida al final está compuesta de una combinación de las imágenes reproducidas sobre estas múltiples vallas publicitarias. La combinación se controla dependiendo de la localización de la cámara virtual.

5
10
15

En otra variante preferida de la invención, las vallas publicitarias aumentan con un campo de altura que define la geometría aproximada del objeto. Es decir, se determinan campos de altura a partir de dos o más vistas de cámaras reales, por ejemplo mediante forma a partir de siluetas o forma a partir de métodos en estéreo, como se describe por ejemplo en "Multiple View Geometry in Computer Vision", Richard Hartley y Andrew Zisserman, Cambridge University Press, 2000. Estos campos de altura se usan después preferiblemente para mejorar la calidad final de reproducción del objeto. Las vallas publicitarias pueden aumentarse también usando mapas de desplazamiento que representan detalles más finos de la geometría del objeto. Esto último es útil para una reproducción más rápida.

20

En una variante preferida adicional de la invención, la superposición en la cámara virtual de la proyección de los objetos y la imagen de fondo se realiza mediante combinación alfa. Los valores en los canales alfa de las vallas publicitarias del objeto se usan entonces directamente, o se ponderan adicionalmente basándose en similitud de ángulo, resolución o campo de vista entre las características ópticas de las cámaras reales y la cámara virtual.

El método para

- usar una representación de un elemento de referencia en una representación esquemática de la cancha para identificar elementos de referencia;
- seleccionar la posición de un elemento de referencia a la posición de uno de los elementos extraídos en el proceso de identificación de elemento de referencia;
- seguir la compensación y corrección según parámetros de cámara cambiantes;
- identificación automática de asistencia;
- alertar al usuario de la presencia de un objeto no identificado;
- clasificar elementos de imagen como pertenecientes al fondo o de una de al menos dos categorías de objetos;
- usar coincidencia de canal alfa;
- marcar elementos de imagen no restaurados como datos de imagen sintéticos;
- restaurar imagen por técnicas 3D;
- refinar la posición y dimensiones de los cuadros de delimitación alrededor de los objetos basándose en los resultados de la segmentación;
- reproducir información de video sobre un modelo de fondo 3D;

30
35

puede implementarse también, cada uno de ellos, en principio, en un sistema que no incorpora el seguimiento del movimientos de los objetos en las al menos dos emisiones de video usando información de posición derivada de la posición 3D de los objetos.

40

El sistema de acuerdo con la invención comprende un módulo de adquisición de datos provisto de emisiones de video a partir de una entidad de producción, y comprende además un módulo de calibrado de cámara, un módulo de seguimientos 2D, un módulo de identificación de objetos, una combinación 3D y un módulo de cálculo de posición del objeto 3D y, preferiblemente, un módulo de recortado de objetos y un módulo de síntesis de imagen que proporciona datos de video a un consumidor.

45

Un producto de programa informático para generar una representación 3D de una escena 3D dinámicamente cambiante según la invención puede cargarse en una memoria interna de un ordenador digital, y comprende medios de código de programa informático para hacer, cuando dichos medios de código de programa informático están localizado en el ordenador, que el ordenador ejecute el método según la invención. En una realización preferida de la invención, el producto del programa informático comprende un medio legible por ordenador, que tiene un medio de código de programa informático registrado en el mismo.

50

Son evidentes otras realizaciones preferidas a partir de las reivindicaciones de patente dependientes. Los elementos

de las reivindicaciones del método pueden combinarse con elementos de las reivindicaciones del sistema y viceversa.

Breve descripción de los dibujos

5 La materia objeto de la invención se explicará con mayor detalle en el siguiente texto con referencia a las realizaciones ejemplares preferidas que se ilustran en los dibujos adjuntos, en los que:

la Figura 1 muestra esquemáticamente una vista general de la estructura de un sistema y método de acuerdo con la invención;

la Figura 2 muestra esquemáticamente imágenes usadas en la fase de interacción del usuario del método de calibrado;

10 la Figura 3 muestra esquemáticamente imágenes que ilustran el método de seguimiento;

la Figura 4 muestra esquemáticamente una situación en la que dos (o más) objetos colisionan como proyecciones 2D en imágenes de video de una cámara pero no en las de otra cámara;

la Figura 5 muestra esquemáticamente una realización de una estructura de posición y tamaño de un objeto 2D, un cuadro de delimitación;

15 la Figura 6 ilustra esquemáticamente el refinado del cuadro de delimitación;

la Figura 7 muestra una vista global esquemática del sistema inventivo con dispositivos periféricos;

la Figura 8 muestra esquemáticamente imágenes con un hueco para llenar y un hueco con una marca que pasa a través del mismo;

la Figura 9 muestra esquemáticamente un círculo proyectado y la rotación de parches;

20 la Figura 10 muestra esquemáticamente una proyección en perspectiva de una marca y la asignación por píxeles para una marca;

la Figura 11 muestra esquemáticamente un modelo de entorno 3D; y

la Figura 12 muestra una imagen reconstruida generada según la invención.

25 Los símbolos de referencia usados en los dibujos y sus significados, se indican de forma resumida en la lista de símbolos de referencia. En principio, a las partes idénticas se les proporcionan los mismos símbolos de referencia en las figuras.

Descripción detallada de la realización preferida

30 La **Figura 1** muestra esquemáticamente una vista general de un sistema 100 y un método correspondiente con componentes de sistema y submétodos correspondientes. Los componentes están unidos mediante interfaces, es decir, entradas y salidas de datos. De esta manera, la figura muestra, por un lado, los componentes o módulos y el flujo de datos principal entre estos módulos. Por otro lado, los módulos corresponden a etapas de método, del método implementado por el sistema. Por lo tanto, los módulos se denominarán también métodos, dependiendo del contexto.

35 El sistema 100 y sus componentes pueden implementarse mediante módulos de software y/o módulos de hardware especializado, dependiendo de los requisitos de procesamiento real de los componentes individuales. De esta manera, el sistema 100 puede implementarse en un dispositivo de procesamiento de datos de propósito general, u ordenador que comprenda una unidad de procesamiento, una unidad de almacenamiento de datos y dispositivos de entrada/salida tales como una pantalla, teclado, dispositivo de señalización e interfaces de comunicación de datos.

40 El sistema comprende un módulo 102 de adquisición de datos provisto de entradas de video desde una entidad 101 de producción, y comprende además un módulo 103 de calibrado de cámara, un módulo 104 de seguimiento 2D, un módulo 105 de identificación de objetos, un módulo 106 de recortado de objetos, un módulo 107 de combinación 3D y cálculo de posición de objetos 3D, y un módulo 108 de síntesis de imágenes que proporciona datos de video a un consumidor 109. El sistema 100 puede comprender además, o estar relacionado con, un modelo 110 de datos de recursos y un módulo 113 de datos de entorno. Los diferentes tipos de datos que fluyen a través de las interfaces entre los módulos se muestran en la lista de designaciones.

45 En la **Figura 1**, la secuencia muestra, de arriba abajo, la dirección de avance del procesamiento de la información (con la etapa 103 de calibrado de cámara precediendo a la etapa 104 de seguimiento 2D, y cada una de las otras etapas precediendo a las etapas mostradas a continuación de las mismas). De esta manera, se considera que una etapa es una etapa posterior de cualquier etapa que la preceda en el orden mostrado. La flecha que surge del cálculo 107 de la posición del objeto 3D denota la posición 130 del objeto 3D, que es proporcionada a las etapas

precedentes tales como el módulo 104 de seguimiento 2D y/o el módulo 105 de identificación de objetos. De esta manera, la posición 130 del objeto 3D constituye una retroalimentación de información, que fluye contra la dirección comúnmente implementada del procesamiento de información.

5 La invención, en una realización preferida, funciona de la siguiente manera: se producen 101 dos o más emisiones de video 120 y se proporcionan en tiempo real al sistema 100. Normalmente, la entrada se proporciona mediante cables de hardware y los datos en su interior en el formato SDI (Interfaz Digital en Serie), como un sistema que se usa habitualmente por los productores de televisión en el sitio de producción para transporte de video. El formato o hardware no es esencial para la invención y puede diferir en otras realizaciones.

10 En una primera etapa, un método 102 de adquisición de datos usa un componente de hardware preferiblemente fuera de la plataforma, que captura (digitalizada) estas dos o más emisiones de video 120 en una memoria interna del ordenador. Este método puede convertir adicionalmente el formato de los datos en RGB o YUV u otro formato de representación de imágenes para procesamiento adicional. Adicionalmente, una etapa opcional de este método comprende un método de desentrelazado usando métodos convencionales, véase "De-Interlacing: A Key Technology for Scan Rate Conversion", Bellars y De Haan, Elsevier, 2000. La salida del método 102 se digitaliza como datos 121 de textura de color de todas las representaciones de video entrantes para todos los fotogramas.

15 Después, se usa el método 103 de **calibrado de cámara** para determinar cada posición de la cámara, orientación y parámetros internos 122 que se usan en otros métodos. La **Figura 2** muestra una vista esquemática de la interacción del usuario usada para inicializar o corregir, en caso de fallo, el método de calibrado. En el fotograma de cada entrada de video correspondiente a un tiempo de inicialización (por ejemplo t_{inic}), el usuario identifica elementos que pueden ser localizaciones 2D 203a,b exactas, líneas 203c, círculos 203d u otros elementos conocidos. Este proceso es asistido por un método de "selección" que identifica elementos de imágenes tales como líneas o círculos usando métodos conocidos, o puede obtenerse por ejemplo a partir de "una biblioteca de visión por ordenador de código abierto" en <http://www.intel.com/technology/computing/opencv/>.

25 Cuando el dispositivo de entrada de usuario (puntero del ratón) está cerca de uno de estos elementos, este "selecciona" dicho elemento cambiando su posición en el elemento en la imagen. Después de haber seleccionado (clicando, pulsando una tecla, etc.) el elemento 203a-d en la imagen 201 de video, el usuario selecciona el elemento 204a-d correspondiente de la vista 202 esquemática. Después de haber seleccionado algunos elementos 203a-d y su representación 204a-d esquemática correspondiente, un algoritmo de cálculo de calibrado, por ejemplo como en "A flexible new technique for camera calibration", Z. Zhang, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11): 1330-1334, 2000, calcula la posición, orientación y parámetros internos de la cámara para este fotograma preciso, conocido como información de calibrado. Estas etapas se llevan a cabo para cada fotograma de inicialización de la cámara en el momento t_{inic} .

30 Para fotogramas posteriores (en el momento $t(k) = t_{\text{inic}}+1, t_{\text{inic}}+2, \dots$) y para cada emisión de video, el método de calibrado sigue automáticamente los elementos identificados en la imagen 2D usando un algoritmo de seguimiento, por ejemplo "Condensation - conditional density propagation for visual tracking", Isard y Blake, International Journal of Computer Vision, 29, 1, 5-28, 1998.

35 En paralelo, o después del método 103 de calibrado, el método 104 de **seguimiento** usa los datos 121 de textura de color digitalizados, los datos de calibrado de la cámara de las etapas 131 actuales y/o (dependiendo de si es en paralelo o no) previas, y la posición 131 del objeto 3D extrapolada para determinar la posición 2D y la forma 123 de todos los objetos visibles en cada conjunto de datos 121 de textura de color.

40 En el fotograma de inicialización en t_{inic} , el usuario preferiblemente especifica la localización de los objetos en cada imagen de video de todas las cámaras. Puede usarse un método automático para proponer candidatos para este proceso. En una realización preferida del método, el usuario especifica un rectángulo 2D, lo que se denomina cuadro de delimitación 501, usando el dispositivo de entrada de ratón y/o el teclado. Este rectángulo 501 puede cambiarse por píxeles en la posición 509 y tamaño (anchura 510 y altura 511) usando el ratón y/o teclado en cualquier momento posterior. En otra realización preferida, el usuario clica/apunta sobre cada objeto y un método automatizado determina el cuadro de delimitación 501 u otra indicación de la forma (semi)automáticamente. Esto puede realizarse usando un método similar al usado en el método de refinado del método 106 de recortado de objetos, descrito más adelante de forma adicional.

45 Para otros fotogramas, el método de seguimiento funciona automáticamente. Básicamente, los métodos de seguimiento aplicables (también en tiempo real), por ejemplo, "Condensation - conditional density propagation for visual tracking", Isard y Blake, International Journal of Computer Vision, 29, 1, 5-28, 19.8, funcionan de la siguiente manera:

- 50 1. Extrapolar el estado 2D de los objetos en la imagen basándose en estados previos.
- 55 2. Buscar los alrededores de esta posición 2D extrapolada comparando los elementos de la muestra de búsqueda con los elementos del objeto. Tal comparación del elemento de muestra puede basarse en un histograma de color, un modelo gaussiano mixto o modelos similares ("Pattern Classification", Duda, Hart y Stork, Wiley Interscience, 2000).

3. Seleccionar una o más de las muestras de búsqueda como el estado actual.

La **Figura 3** muestra esquemáticamente nuestro método de seguimiento mejorado descrito en la presente memoria. A partir del método 107 de **cálculo de posición de objetos 3D**, el método 103 de calibrado consigue la posición del objeto 3D extrapolada de todos los objetos. Esta extrapolación se realiza en el modelo 3D de la escena 303, donde por ejemplo la posición 3D del fotograma anterior 331 a,b y del fotograma antes del fotograma anterior 330a,b, se usan para extrapolar la posición del objeto 3D en el fotograma actual 332a,b, basándose en un esquema de interpolación lineal o de mayor orden.

Obsérvese que la Figura 3 no representa un único fotograma, sino que muestra la posición de los objetos 310a,b, 312a,b ... superpuestos desde dos fotogramas diferentes con fines de explicación. Una emisión de video real mostraría, en un primer fotograma, los objetos 310a y 310b y, en un fotograma posterior, los objetos 311a, 311b. Los objetos corresponden, en un fotograma de video actual, a segmentos de imagen determinados por seguimiento y segmentación, como se describe adicionalmente más adelante.

Los algoritmos de seguimiento del estado de la técnica solo siguen un objeto en una única secuencia de video, permaneciendo de esta manera en el espacio de imagen 2D. Las trayectorias 2D del objeto seguido basándose solo en las posiciones 2D seguidas se representan como líneas con flechas 311a,b para los objetos A y B, dando como resultado una posición 312a,b 2D esperada en el fotograma previo 301, donde otro algoritmo de seguimiento empezaría a buscar el objeto.

En nuestro caso, no se espera que las cámaras estén fijas, de esta manera, pueden cambiar su orientación y parámetros internos (tal como el zoom) con el tiempo, dando como resultado una vista diferente de la escena en el momento actual (fotograma 302) que previamente (fotograma 301). Cuando solo se tienen en cuenta las posiciones 2D resultantes de otro algoritmo de seguimiento, la búsqueda del método de seguimiento empezará en las mismas posiciones 2D 321a,b que las mostradas en el fotograma previo 312a,b. Usando los datos de calibrado 131 reales o extrapolados del método 103 de calibrado y la posición 130 del objeto 3D extrapolada del método 107 de cálculo de posición de objetos 3D, se calculan las posiciones 2D de las proyecciones 3D extrapoladas y las posiciones 320a,b en la imagen real. Esto no solo afecta a la posición 2D de la estimación del objeto, sino también al tamaño o forma del objeto de la imagen de video. De esta manera, la etapa de búsqueda de seguimiento comenzará con una posición 2D y estimación de forma mejoradas, dando como resultado un menor espacio de búsqueda que reduce el esfuerzo computacional del método de seguimiento.

Otra mejora del método de seguimiento es que los objetos que entran o salen del área visible de la cámara pueden detectarse fácilmente si se conoce su posición 3D. Proyectando las posiciones 130 del objeto 3D extrapoladas de todos los objetos en la imagen 2D, deben considerarse todas las posiciones 2D situadas dentro de los límites del fotograma durante el seguimiento. Esto puede realizarse automáticamente el método o puede estar asistido por el usuario, según las circunstancias y la información de video disponible.

Otra mejora del método de seguimiento es la manipulación de colisiones de objetos en el espacio de la imagen 2D, como se muestra esquemáticamente en la **Figura 4**. En la vista esquemática de la escena 403, las posiciones 3D de los objetos A, B se muestran para un momento previo $t(k-1)$ 430a,b y para el tiempo actual $t(k)$ 431a,b. Puede verse la misma situación en la vista 401 desde una cámara con ambos objetos 411a,b colisionando en 2D, mientras que la misma situación desde la vista 402 de otra cámara no muestra colisión de los objetos 421a,b. Esto conduce a dos tipos adicionales de información para el método de seguimiento:

1. El método de seguimiento puede usar la información de que dos o más objetos colisionan en una cierta vista 401 y, de esta manera, sabe que hay dos o más objetos aunque únicamente "un objeto" es reconocible de la vista 2D en solitario.

2. El método de seguimiento puede mantener fácilmente el seguimiento de los objetos después de una colisión, puesto que sabe dónde están localizados los objetos o se espera que estén en un espacio 3D y, por lo tanto, también en la vista 2D y, de esta manera, continuará el seguimiento de cada objeto individual y correctamente.

Finalmente, el método de seguimiento determina la posición del objeto en 2D y la forma 123 para todos los objetos visibles en todos los fotogramas de todas las cámaras, y se las proporciona al módulo 105 de identificación de objetos.

Después del método de seguimiento, el método 105 **de identificación de objetos** asocia, para cada objeto visible en cada emisión de video, la posición 2D del objeto y la forma en los datos 123 de textura de color con un objeto real (por ejemplo, jugadores, porteros, árbitros, pelota, etc.) basándose en los datos 122 de calibrado de la cámara, la información sobre objetos 132 del mundo real contenidos en un módulo de datos de recursos (o simplemente "recurso") 110, y posiblemente también la posición 130 del objeto 3D extrapolada y la posición 2D y la forma 123 para esencialmente todos los objetos en todos los fotogramas de todas las cámaras proporcionados por el método 104 de seguimiento. En este punto y en cualquier otro, "todos" significa "esencialmente todos", es decir, sin defectos o entidades patológicas o que funcionen mal (objetos, cámaras).

En el transcurso de la inicialización para el fotograma en el momento t_{inic} , el usuario asocia cada información 123

de posición 2D y forma de una cámara con un objeto 132 del mundo real específico, que todos conocen previamente (nombres de jugador, nombres de los porteros, árbitros, pelota, etc.) desde el recurso 110. Para los fotogramas de la otra cámara en el momento t_{inic} , el método de identificación de objetos preferiblemente sugiere automáticamente la identificación (por ejemplo un nombre) llevando a cabo las siguientes etapas:

- 5 1. Para todos los objetos identificados
 - a. Calcular una aproximación de la posición 3D (usando la suposición a nivel de suelo como se describe en el método 107 de cálculo de la posición 3D) si no hay posición 3D disponible por extrapolación 130.
 - b. Calcular una posición 2D proyectando dicha posición 3D en cada fotograma de la cámara en una imagen usando los datos 122 de calibrado.
- 10 c. Para todas dichas posiciones 2D dentro del límite del fotograma de la cámara respectiva
 - i. Determinar la información 123 de posición 2D y forma que está localizada cerca de dicha posición 2D.
 - ii. Si solo hay una posición 2D dentro de algún umbral definido por el usuario respecto a la distancia, esta información 123 de posición 2D y forma está asociada con el objeto del mundo real asociado con dicho objeto identificado.
 - 15 iii. Si hay más de una posición 2D dentro de dicho umbral, presentar al usuario una lista acortada por la distancia desde la cual puede seleccionar activamente a mano la segunda o más distante. En caso de que no haya acción por parte del usuario, permanece seleccionada el primera. Asociar la información 123 de posición 2D y forma con el objeto de mundo real asociado con dicho objeto seleccionado.
- 20 d. Para todas las posiciones 2D y forma de cada cámara que no se hayan asociado con un objeto del mundo real en una de las etapas previas, determinar una lista de posibles candidatos
 - i. Eliminando aquellos objetos del mundo real que ya estén asociados con información 123 de posición 2D y forma en dicho fotograma de la cámara.
 - ii. Ordenando a los candidatos restantes aumentando la distancia de la proyección 2D desde su posición 3D en la imagen (como en 1b).
 - 25 iii. Los objetos cuya posición 3D aún no es calculable se colocan en la parte delantera o trasera de la lista, dependiendo de los ajustes de usuario.

Para todos los fotogramas posteriores, el método de identificación de objetos lleva a cabo las siguientes etapas:

1. Si la información 123 de posición y forma del objeto 2D ya se ha identificado y asociado en el fotograma previo, usar la información 132 consecuente de nuevo.
- 30 2. Si la información 123 de posición y forma del objeto 2D aún no se ha identificado en la etapa previa del método del seguimiento, llevar a cabo las mismas etapas que durante el fotograma de inicialización para objetos no identificados.
3. Si la información 123 sobre la posición y forma del objeto 2D aún no se ha asociado con una información 132 del mundo real, marcar los fotogramas que requieren la interacción del usuario.
- 35 4. Para todos los fotogramas, donde se ha encontrado una identificación no ambigua de toda la información 123 sobre posición y forma del objeto 2D proporcionada por el método 104 de seguimiento, marcar dicho fotograma como realizado y que no requiere interacción adicional del usuario.

El usuario asocia después manualmente la información 123 sobre posición y forma del objeto 2D en todos los fotogramas marcados según requieran interacción del usuario usando las listas de posibles candidatos hasta que todos los fotogramas contengan una identificación no ambigua de toda la información 123 de posición y forma del objeto 2D.

El método 105 de identificación de objetos da como resultado la posición y forma del objeto 2D y la identificación 124 del objeto para todas las posiciones y formas del objeto 2D encontradas en el método 104 de seguimiento de todas las emisiones de video.

Los datos proporcionados por la adquisición de datos 102 y la identificación de objetos 105, en concreto los datos 121 de textura de color y la posición y forma del objeto 2D incluyendo la identificación 124 de objetos se introducen en el método 106 de recortado de objetos.

El método calcula una segmentación o recorte de los datos de textura de color dentro del área definida por la posición y tamaño del objeto entre los píxeles en primer plano (objetos) y los píxeles de fondo, lo que se denomina máscara alfa. De esta manera, una salida del método de recortado de objetos son los datos de textura del objeto

asociados con una máscara alfa 126.

La **Figura 5** muestra esquemáticamente una realización de una posición y estructura de tamaño del objeto 2D, en un cuadro de delimitación 501. Representa también la diferencia entre el primer plano 502 y el fondo 503 de los valores 505 de la máscara alfa resultantes del método de recorte o segmentación.

- 5 Si el cuadro de delimitación 501 no contiene todo el objeto o interseca el cuadro de delimitación, dicho cuadro de delimitación 501 puede suponerse a priori o estar definido por un usuario, el cuadro de delimitación puede ampliarse en un cierto tamaño, ya sea por el usuario, en un valor fijo, o por un procedimiento automático que analiza el cuadro de delimitación 501.

10 En este método, los objetos se clasifican en primer lugar de acuerdo con al menos dos categorías. Las categorías preferiblemente están basadas en modelos de elementos estadísticos, tales como un histograma, un modelo gaussiano mixto, o métodos similares ("Pattern Classification", Duda, Hart y Stork, Wiley Interscience, 2000). Los elementos preferibles incluyen, aunque sin limitación, colores o bordes. Antes del procesamiento, el modelo estadístico para todas las categorías de objetos se construye usando imágenes estáticas de una o más cámaras y después se aplica a las emisiones de video de todas las cámaras. El modelo estadístico se genera, para cada categoría de objetos, moviendo el usuario, mediante un dispositivo de señalización, una marca de referencia a lo largo de una trayectoria sobre una diversidad de puntos que pertenecen a dicha categoría. Los elementos de los puntos de dicha trayectoria forman una muestra que representa dicha categoría en la generación del modelo estadístico. Para cada categoría de objetos, los puntos con la marca de referencia o los elementos de imagen (píxeles) se etiquetan como pertenecientes al fondo o al primer plano. El usuario marca el fondo preferiblemente para píxeles fuera del cuadro de delimitación 501 y dentro del cuadro de delimitación 501 pero en regiones en las que el fondo es realmente visible. El usuario marca el fondo para las partes dentro del cuadro de delimitación en regiones que pertenecen al primer plano. Preferiblemente, el modelo se genera usando una o múltiples imágenes 102 de entrada.

25 Posteriormente, se genera el modelo usando los métodos apropiados para los modelos usados, por ejemplo mediante agrupación de medios k o métodos de Expectación-Maximización, véase, "Pattern Classification", Duda, Hart y Stork, Wiley Interscience, 2000. Preferiblemente, el modelo se define por un número de parámetros estadísticos, por ejemplo por la media, varianzas o matrices de covarianza. El recorte se realiza procesando todos los píxeles en los cuadros de delimitación usando los parámetros del modelo estadístico correspondiente, teniendo en cuenta:

- 30 1. Similitud de datos, es decir, medición de la concordancia de color de una muestra de datos con los modelos de color, calculando probabilidades entre parámetros de modelo y muestras de datos.
2. Proximidad o similitud previa, es decir, evaluar las diferencias de color entre muestras espacial y temporalmente adyacentes. Se penalizan las desviaciones de color fuertes y se asegura coherencia espacial y temporal en el proceso de selección.

35 Para realizar el recorte, la segmentación se interpreta preferiblemente como un problema de etiquetado de gráficos que produce una solución usando un método de recorte mínimo. Por ejemplo, puede encontrarse una formulación y solución en "An experimental comparison of min-cut/max-flow algorithms for energy minimization in visión", Boykov y Kolmogorov, IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 9, 1124-1137, el resultado del recorte es una segmentación binaria (etiquetas) del cuadro de delimitación 501 en el primer plano 504 y el fondo 503 o 0 (fondo) y 1 (primer plano).

40 Preferiblemente, un método de post-procesamiento asigna valores alfa 505 a todos los píxeles en la proximidad del límite entre los píxeles del primer plano 508 y del fondo 506, incluyendo píxeles que tienen valores intermedios 507. Estos valores intermedios expresan el límite suave del objeto y pueden calcularse usando por ejemplo "A Bayesian Approach to Digital Matting", Chuang, Curless, Salesin y Szeliski, IEEE Computer Vision and Pattern Recognition, Vol. II, 264-271, 2001.

Usando la máscara alfa, la posición y forma del objeto 2D del objeto se refinan y proporcionan como salida adicional 127 del método 106 de recortado de objetos.

45 La **Figura 6** ilustra esquemáticamente tal refinado del cuadro de delimitación. Para ello, para cada cuadro de delimitación 601 dado inicialmente (línea de trazos) se atraviesa la máscara alfa y se calculan los valores mínimo y máximo en ambas dimensiones (horizontal y vertical, o u, v) para los que existen píxeles de primer plano. Se reduce el tamaño del cuadro de delimitación 601 (línea sólida) y hace coincidir mejor la forma del objeto actual 602. El mínimo sirve entonces como la nueva posición del cuadro de delimitación y las diferencias entre máximo y mínimo en cada dirección se toman como la nueva anchura (en la dirección u) y altura (en la dirección v) del cuadro de delimitación. Se toma la referencia o punto de anclaje 604 para el método de cálculo de la posición 3D, por ejemplo, para la dirección u como que es el punto medio entre el máximo y el mínimo, y en la dirección v como que es el mínimo.

Finalmente, el método determina todas las áreas del primer plano (un cierto conjunto de píxeles) en los datos de

5 textura de color de las emisiones de video 121 atravesando todos los objetos en cada emisión de video y marcando los píxeles que, en los datos 121 de textura de color, están marcados como primer plano con un banderín. Para conseguir datos de textura de fondo sin los objetos de primer plano y sin huecos, estas áreas se llenan usando un método de restauración de imagen, véase por ejemplo "View Interpolation for Image Synthesis", Chen y Williams, ACM SIGGRAPH 1993, págs. 279-288, que, para cada hueco-píxel en el borde de un hueco, selecciona posteriormente los parches más similares a partir de regiones no marcadas de la imagen y asigna el valor correspondiente a este píxel de borde, llenando así el hueco desde el borde hasta el medio. Todos estos píxeles generados sintéticamente se marcan especialmente como "restaurados", usando un banderín asociado con el píxel, dando como resultado una salida adicional del método de recortado que son unos datos 125 de textura de fondo llenos y marcados.

10 En resumen, la salida del módulo 106 de recortado comprende

- Datos 125 de textura de fondo llenos y marcados;
- Textura de objeto y máscara alfa por cámara y objeto del mundo real, por ejemplo información 126 de segmentación; y
- 15 • Posición y formatos del objeto 2D de refinado con identificación 127 del objeto del mundo real.

El **método 107 de cálculo de la posición del objeto 3D** usa la información de posición y tamaño 2D refinada con la información 127 del mundo real y los datos 122 de calibrado de la cámara de (todas) las cámaras para determinar la posición 3D de (todos) los objetos. El método lleva a cabo las siguientes etapas:

1. Para cada objeto del mundo real, se recoge la información asociada de posición y forma 2D.
- 20 2. Para objetos visibles sin cámara, la posición 3D de las etapas previas (si está disponible) se extrapola usando un esquema de interpolación de primer orden o superior.
3. Para objetos visibles en solo una cámara, se supone que el objeto está localizado a nivel de suelo, de esta manera, el vector desde la posición de la cámara que pasa a través de la posición 2D en la imagen 2D proyectada se interseca con el plano de suelo, dando como resultado la posición 3D del objeto.
- 25 4. Para objetos visibles en dos o más cámaras, pueden usarse dos métodos diferentes:
 - a. Se supone que todos los objetos están localizados a nivel de suelo, de esta manera, se usa el mismo método que el descrito anteriormente, dando como resultado dos o más posiciones 3D a nivel del suelo desde las cuales se calcula la media o promedio como la posición 3D final del objeto.
 - 30 b. El otro método no supone que los objetos están localizados a nivel de suelo y calcula el punto de aproximación más cercano a los vectores usando una técnica convencional, por ejemplo "Intersection of two lines in three space", Goldman, En Andrew Glassner, editor "Graphics Gems" página 304. Academic Press, 1990.

De esta manera, el método 107 de cálculo de la posición 3D asigna a cada objeto modelado una posición 3D para cada momento (fotogramas de video) y produce estas posiciones 3D de todos los objetos 128. Puede accederse a estos desde otros métodos en etapas posteriores, en las que "posterior" se entiende en el sentido temporal - estas etapas pueden ser "etapas precedentes" en el contexto de la estructura del sistema 100, con el efecto de que este acceso constituya un bucle de retroalimentación.

- El método 108 de síntesis de imágenes interpoladas usa la información del fotograma actual 128,
- el método 104 de seguimiento preferiblemente usa la información 130 de posición del objeto 3D extrapolada de uno o más fotogramas previos y
- 40 • el método 105 de identificación de objetos preferiblemente usa dicha información 130 para asociar la información 123 de posición y forma del objeto 2D con la correspondiente información 132 del objeto del mundo real.

El **método 108 de síntesis de imagen** interpolada genera la vista sintética de la escena. Las entradas a este método son las posiciones 3D de los objetos 128, los datos 122 de calibrado de la cámara, la textura del objeto y las máscaras alfa 126 y los datos 125 de textura de fondo llenos y marcados. Opcionalmente, las entradas comprenden datos 135 de reconstrucción 3D del entorno, proporcionados por el módulo 113 de datos del entorno y/o formas de los objetos.

El método de síntesis de imagen comprende las siguientes etapas:

1. Reproducir los objetos a partir de una vista virtual usando una representación 3D particular de la escena y usando texturas 126 del objeto y cualquier valor alfa fijo (a partir de la etapa 106 de recortado) o valor alfa dependiente de la vista, teniendo en cuenta la similitud angular, la resolución y el campo de vista. Preferiblemente, el mapeo de textura se consigue usando texturado proyectivo usando los datos 122 de calibrado. La similitud angular penaliza vistas
- 50

cuyo eje óptico está más alejado del eje óptico de la vista virtual. La similitud de resolución penaliza cámaras que, por ejemplo, están más alejadas de la geometría diana o tienen menor resolución en general. El campo de visión penaliza similarmente datos de las vistas de cámara que no ven la geometría diana de la vista virtual.

- 5 2. Reproducir la geometría del fondo mediante combinación alfa y almacenamiento temporal de la profundidad con los objetos ya reproducidos por combinación de una o más texturas 125 de fondo con huecos llenos, y teniendo en cuenta las muestras de datos marcadas durante la combinación. Preferiblemente, se consigue el mapeo de textura usando texturado proyectivo usando los datos 122 de calibrado.

10 En la etapa 1, la representación 3D particular preferiblemente son una o más vallas publicitarias, vallas publicitarias potenciadas para campo de altura, vallas publicitarias potenciadas para mapa de desplazamiento o modelos de 3D completos de objetos.

Para la etapa 2, la geometría del fondo se modela ya sea mediante superficies que representan únicamente el campo de juego (o cancha) o que también incluyen una representación del entorno, tal como un modelo de estadio de mayor o menor complejidad. De esta manera, pueden usarse vallas publicitarias, vallas publicitarias potenciadas para mapa de profundidad o modelos 3D completos para la reproducción.

- 15 La **Figura 11** muestra esquemáticamente un modelo de entorno 3D que comprende una superficie 901 plana que representa una cancha, y otras superficies 902 (inclinadas) que representan un estadio que rodea la cancha. La **Figura 12** muestra una imagen reconstruida generada de acuerdo con la invención. Las vallas publicitarias 903, que normalmente no se muestran, se destacan para ilustración.

20 Teniendo en cuenta las muestras de datos marcadas durante la combinación significa lo siguiente: en lugares donde existen varias vistas de la misma superficie de fondo, y donde en al menos una vista un parche de superficie se marca como imagen restaurada (es decir, datos sintéticos), y en al menos otra vista el mismo parche se marca como imagen no restaurada (es decir, datos reales), los datos sintéticos se ignoran y solo se usan los datos reales. Si están disponibles datos reales de varias fuentes, pueden combinarse, o pueden usarse datos de una única fuente.

25 Además, superficies arbitrarias de la escena 3D pueden colorearse reproduciendo imágenes estáticas o dinámicas predeterminadas sobre los mismos. La síntesis 108 de imágenes genera automáticamente una vista que muestra dichas imágenes como si se hubieran pintado sobre las superficies modeladas correspondientes. De esta manera, los datos determinados por el usuario tales como logos, publicidad, patrocinios, anotaciones interactivas, estadísticas de jugadores, etc. se insertan en la escena 3D y se proyectan en la vista 2D sintética.

30 Finalmente, la salida del método 108 de síntesis de imagen interpolada, en concreto la emisión 129 del video interpolado sintetizado, se hace pasar al destino o consumidor 109 después de haberla transformado a un formato conveniente para el consumidor, por ejemplo un formato SDI convencional como se usa para las entradas de emisiones 120 de video.

35 La **Figura 7** muestra una vista global esquemática del sistema inventivo 100 con dispositivos periféricos, es decir, al menos dos cámaras reales 702 (dibujadas en negro) dispuestas en un campo de juego 701, una unidad de control 704 de un productor de contenidos para recoger, gestionar y dirigir múltiples emisiones de video a una unidad 705 de distribución de medios, conectada a una unidad de difusión, tal como una antena 706. El sistema 100 se dispone para recibir al menos dos emisiones de video de entrada y generar al menos una emisión de video de salida. La emisión de video de salida puede proporcionarse a la unidad 705 de distribución, como se muestra, o a la unidad 704 de control. El sistema 100 genera vistas correspondientes a cámaras virtuales 703 (dibujadas en línea discontinua) localizadas en posiciones y orientaciones donde no existen cámaras reales.

Llenado de huecos/completado de imágenes

45 La **Figura 8** muestra esquemáticamente imágenes con un hueco que se tiene que llenar y un hueco con una marca que pasa a través del mismo. Las imágenes son de un fondo y se han generado segmentando la imagen original en una parte 801 de fondo, con una textura conocida también denominada "área conocida", y un "área desconocida" o huecos 802 en áreas donde los objetos se han identificado y recortado del fondo. En una realización preferida de la invención, la etapa de llenar los huecos 802 en el fondo preferiblemente comprende la aplicación repetida de las etapas de

- Seleccionar un "parche de destino" 803a, 803b,... cuya área es parcialmente conocida y parcialmente desconocida;
- 50 - Seleccionar un "parche de origen" 804a, 804b,... que comprende al menos un pixel con un valor de color conocido, es decir, a partir del área 801 conocida;
- Copiar los datos de imagen (o textura) del parche de origen a las localizaciones correspondientes en el parche de destino para aquellas áreas del parche de destino que son desconocidas, es decir, en el hueco 802. Como resultado, el área del hueco 802 se reduce.

Véase por ejemplo "Fragment-based image completion", Drori, I., Cohen-Or, D. y Yeshurun, H., 2003. ACM Trans. Graph. 22, 3 (jul. 2003), págs. 303-312 o "Image completion with structure propagation", Sun, J. Yuan, L., Jia, J. y Shum, H., 2005, En ACM SIGGRAPH 2005 Papers. J. Marks., Ed. SIGGRAPH'05. ACM Press, Nueva York, NY, págs. 861-868. Tales parches preferiblemente son de forma rectangular o circular. Preferiblemente, múltiples parches de destino 803a y 803b solapan entre sí. En una realización preferida de la invención, los parches de destino solapantes se combinan en la región común 805.

En una realización preferida de las etapas de llenado de agujeros en la imagen de fondo, el parche de destino 803a, 803b,... tiene el mismo tamaño y orientación que el parche de origen 804a. Después, la información de píxel desde el parche de origen puede copiarse directamente. En otra realización preferida, el parche de origen y de destino tienen diferentes tamaños, forma y orientación que las del parche 804b. En tales casos, puede generarse fácilmente una función que mapea las posiciones del parche de destino a las posiciones de los parches de origen, donde una persona debe consultar los valores de color correspondientes. Después, el copiado puede implicar rotación y/o cambio de escala de la imagen de parche de origen (tal como 804b).

En una realización preferida de la invención, la etapa de seleccionar un parche de origen 804a, 804b,... comprende una comparación de parches de origen potencialmente posibles con el parche de destino 803a, 803b,... para encontrar un parche de origen cuyos valores del píxel correspondan a los valores de píxel conocidos en el parche de destino. Atravesando las regiones conocidas del parche de destino, puede realizarse una comparación por píxeles (dando como resultado la diferencia de los valores de píxel) de todos los parches de origen potencialmente coincidentes con la región conocida para encontrar un buen parche de origen coincidente. Preferiblemente, se construye la suma de estas diferencias para encontrar un valor de comparación global de los parches. La diferencia de los valores de píxel puede basarse en (aunque no está restringida a) RGB, HSV u otros valores del sistema de color del píxel.

En una realización preferida de la invención, las etapas de llenado de los huecos 802 en el fondo 801 conocido tiene en cuenta el conocimiento previo sobre las marcas 806a, 806b,... en el fondo. Tales marcas, por ejemplo, son las líneas y círculos en una cancha de fútbol y son importantes para la percepción correcta y el reconocimiento de la escena. Puesto que los parámetros de cámara son conocidos de la etapa de calibrado de la cámara y la posición de las marcas normalmente está bien definida en las normas del juego, es posible proyectar las marcas desde el espacio 3D de vuelta a la imagen de video. Como resultado, se conocen las posiciones 2D de las marcas (por ejemplo, las líneas o círculos proyectados) en la imagen de video. Usando este conocimiento, pueden detectarse regiones desconocidas (por ejemplo huecos) en las marcas proyectadas atravesando toda las posiciones (es decir, píxeles) relacionados con una marca y comprobando la existencia de un hueco 802. Para marcas con regiones 807 desconocidas identificadas, la recreación de las partes que faltan de las marcas puede realizarse por separado antes de llenar los huecos restantes. Para ello, se seleccionan los parches de destino 803c, 803d a lo largo de la marca que se va a completar. Preferiblemente, el punto medio de cada parche siempre está situado en las posiciones proyectadas hacia atrás de la marca dentro de la imagen. Preferiblemente, los parches de origen se buscan después a lo largo de las regiones conocidas de la marca 804c, 804d. Puesto que los parches de origen correspondientes más probablemente muestran una estructura similar, la coloración y orientación de la marca que contienen, la marca dentro del hueco puede recuperarse más rápido y con una mejor calidad.

La aplicación repetida de tales parches de origen recreará después una imagen de la marca dentro del hueco. En esta realización, el tamaño de los parches preferiblemente se elige para cubrir al menos la anchura máxima de la marca.

La **Figura 9** muestra esquemáticamente un círculo proyectado y la rotación de parches: en una realización preferida de la invención, las etapas de llenado de huecos en el fondo, usando el conocimiento previo sobre las marcas, determina un ángulo de rotación entre el parche de origen 804e y el parche de destino 803e cuando se trata con marcas 806b no lineales (es decir, que no son en línea recta), por ejemplo círculos o similares. Este ángulo de rotación puede calcularse a partir del ángulo 809 entre las normales 808a, 808b (es decir, la línea perpendicular a la marca en una cierta posición) en las posiciones de los parches en la marca proyectada. Antes de comparar el parche de origen con el parche de destino, el ángulo de rotación se integra en la función de mapeo. De esta manera, un parche de origen de una parte distante del círculo puede compararse con, y aplicarse, al parche de destino con una orientación correcta, reduciendo la aparición de defectos visibles en el resultado.

En una realización preferida de la invención, la búsqueda de parches de origen comienza en la proximidad cercana del parche de destino antes de considerar parches de origen a una mayor distancia. En esta realización, puede definirse un valor umbral que detiene la búsqueda cuando se encuentra un parche de origen cuyo valor coincidente es menor que el umbral - si "menor" corresponde a una mejor coincidencia. De esta manera, el algoritmo buscará únicamente hasta que se encuentra un buen parche de origen (pero puede que no sea el mejor).

En otra realización preferida de la invención, la selección de los parches de destino cambia alternativamente entre ambos lados de un hueco, si el hueco está entre los extremos de la marca - por ejemplo en ambos lados del hueco, parte de las líneas son visibles. En esta realización, los parches de origen encontrados se aplican alternativamente a parches de destino en ambos lados 803c y 803d del hueco, llenándolo o creciendo hacia dentro desde el límite del hueco, a lo largo de la longitud 10 de la marca, hasta que se alcanza la mitad del hueco. De esta manera, los píxeles

desconocidos en ambos lados del hueco proporcionan una mejor aproximación a los píxeles circundantes conocidos que lo que conseguiría un llenado unidireccional en el lado distante del hueco. El parche de origen de los parches aplicados alternativamente puede diferir para cada etapa de aplicación, o puede permanecer igual para varias etapas alternas.

5 En otra realización preferida de la invención, los píxeles del parche de origen seleccionado actualmente se combinan (por ejemplo, se mezclan) con los píxeles de al menos uno de los últimos parches de origen aplicados antes de aplicarse a un parche de destino actual. El factor de combinación depende de la distancia 801 desde el parche de destino 803c real a la posición donde estaba el último parche 803d de destino. Por ejemplo, si la distancia 801 es grande, entonces el parche de origen no se altera antes de aplicarlo. Sin embargo, cuando la distancia es pequeña, el parche de origen se altera de manera que parece similar al último parche de destino cercano. De esta manera, el parche de destino actual tendrá una diferencia en disminución por píxeles de otros parches de destino con el aumento de la distancia, lo que conduce a un solapamiento sin costuras en el medio del hueco.

10 La **Figura 10** muestra esquemáticamente una proyección en perspectiva de una marca en línea recta y la asignación por píxeles para una marca: otra realización preferida de la invención usa parámetros de cámara conocidos para determinar un factor de escala integrado en la función de mapeo de los píxeles de parche de origen a los píxeles de parche de destino (y viceversa). Una marca que empieza en un punto 806c, respectivamente cercano a la cámara, y que termina en un punto distante 806d, conduce a una proyección de la marca que requiere parches de origen más grandes correspondientes a posiciones cerca de la cámara 804f que los parches de origen correspondientes a las posiciones alejadas 804g debido a la proyección de perspectiva. Usando cálculos de geometría proyectiva, el factor de escala entre un parche de origen 804f, 804g y un parche de destino 803f se determina y aplica durante la comparación y aplicación de los parches. Preferiblemente, este factor de escala se usa también cuando se elige el tamaño de los parches de origen y de destino, respectivamente.

15 En una realización preferida adicional de la invención, las marcas no se consideran como líneas o elementos unidimensionales, por ejemplo, puntos, líneas rectas o círculos, sino que se asocian también con una dimensión adicional perpendicular a la orientación principal, tal como la anchura de una línea. La anchura de la marca real puede medirse en el sitio o suponerse fácilmente, basándose en la experiencia. Esto da como resultado una marca que se describe por un área delimitada por dos bordes 811 en lugar de solo por una línea o círculo 806c. En la proyección de la marca en la imagen, los píxeles o, más en general, las partes de la imagen, pueden clasificarse entonces como pertenecientes a la marca 812, no pertenecientes a la marca 813 o pertenecientes parcialmente a la marca 814. Preferiblemente, la función de comparación integra este conocimiento para proporcionar una comparación mejorada de los parches. Asimismo, la aplicación del parche de origen sobre el parche de destino integra preferiblemente este conocimiento.

20 En una realización preferida de la invención, las etapas de recrear las marcas implican un tratamiento especial de las áreas donde al menos dos marcas se intersecan o unen entre sí en un hueco. Tales regiones deberían regenerarse de una manera diferente, puesto que la aplicación de parches como se ha descrito anteriormente puede conducir a efectos visuales. Preferiblemente, la regeneración de estas áreas se realiza basándose en la clasificación de píxeles como pertenecientes a la marca 812, no pertenecientes a la marca 813 o parcialmente pertenecientes a la marca 814. Después, los píxeles que pertenecen a la marca en el parche de origen se comparan y aplican únicamente si se mapean sobre un pixel que pertenece a la marca en el parche de destino y viceversa. En otras palabras, para aquellas áreas de hueco en las que se espera una marca, el algoritmo de coincidencia de parches solo considera los píxeles que son parte de la marca. Preferiblemente, tal realización tiene un tratamiento especial para combinar píxeles clasificados como parcialmente pertenecientes a la marca, por ejemplo por combinación o selección del más probable.

25 Aunque la invención se ha descrito en las presentes realizaciones de la invención, se entiende claramente que la invención no está limitada a estas, sino que de lo contrario puede realizarse y practicarse variadamente dentro del alcance de las reivindicaciones.

Listado de designaciones

- 101 El origen (productor) de al menos dos emisiones de video (120)
- 102 Adquisición de al menos dos emisiones de video (120) en la memoria interna
- 103 Método de calibrado
- 104 Método de seguimiento
- 105 Método de identificación de objetos
- 106 Método de recortado de objetos
- 107 Método de cálculo de la posición del objeto 3D

ES 2 585 903 T3

- 108 Método de síntesis de imagen interpolada
- 109 El destino (consumidor) de la emisión o emisiones de video resultantes
- 110 Un recurso (archivo, base de datos, etc.) que contiene información (equipo, nombre, número, etc.) sobre todos los objetos del mundo real (jugadores, porteros, árbitros, pelota)
- 113 Un recurso (archivo, base de datos, modelo, etc.) que proporcione información sobre el entorno (3D, CAD, etc.)
- 120 Al menos dos emisiones de video aisladas
- 121 Datos de textura de color
- 122 Datos de calibrado de cámara
- 123 Posición y forma 2D del objeto
- 124 Posición y forma 2D del objeto e identificación del objeto del mundo real
- 125 Datos de textura de fondo llenos y marcados
- 126 Textura de objeto y máscara alfa por cámara y objeto del mundo real, por ejemplo información de segmentación
- 127 Posición y forma 2D del objeto refinado con identificación del objeto del mundo real
- 128 Posición 3D del objeto
- 129 Emisión de video interpolada sintetizada
- 130 La posición 3D del objeto extrapolada
- 131 Datos de calibrado de cámara reales o extrapolados
- 132 Información sobre un objeto del mundo real según se almacena en (110)
- 135 Modelo 3D, datos o información similar sobre el entorno

- 201 La primera imagen de la emisión de video que muestra algunos elementos en perspectiva
- 202 Una vista esquemática de la escena con sus elementos
- 203a-d Elementos marcados en la imagen de video
- 204a-d Elementos correspondientes en la vista esquemática
- 301 Un fotograma de una emisión de video en el tiempo $t-n$ ($n>0$)
- 302 Un fotograma de una emisión de video en el tiempo t
- 303 Una vista esquemática de la situación 3D de las escenas en diversos momentos
- 310a,b La posición 2D de los objetos A, B en el fotograma 301
- 311a,b Las trayectorias real y extrapolada de los objetos A, B
- 312a,b La posición 2D extrapolada de los objetos A, B en el fotograma 301 (espacio de imagen)
- 320a,b Las posiciones 2D reales de los objetos A, B en el fotograma 302
- 321a,b Las mismas posiciones 2D que en 312a,b en el fotograma 302 (la misma localización espacial de la imagen)
- 330a,b La posición 3D conocida de los objetos A, B en el tiempo $t-2n$
- 331a,b La posición 3D conocida de los objetos A, B en el tiempo $t-n$
- 332a,b La posición 3D estimada de los objetos A, B en el tiempo t

ES 2 585 903 T3

- 401 Una vista de cámara de una escena donde dos objetos A, B colisionan en 2D
- 402 La misma escena que en 401 vista desde otra cámara
- 403 Una vista esquemática de la escena en 3D (desde arriba)
- 410a,b Posiciones 2D de los objetos A, B en el tiempo t0 en 401
- 411a,b Posiciones 2D de los objetos A, B en el tiempo t1 en 401
- 420a,b Posiciones 2D de los objetos A, B en el tiempo t0 en 402
- 421a,b Posiciones 2D de los objetos A, B en el tiempo t1 en 402
- 430a,b Las posiciones 3D de los objetos A, B en el tiempo t0
- 431a,b Las posiciones 3D de los objetos A, B en el tiempo t1

- 501 El cuadro de delimitación
- 502 El objeto está completamente dentro del cuadro de delimitación
- 503 Cierta parte del cuadro de delimitación pertenece al fondo
- 504 Cierta parte del cuadro de delimitación pertenece al primer plano (pertenece al objeto)
- 505 Un aumento de una parte de límite entre el primer plano y el fondo consiste en píxeles
- 506 Algunos de estos píxeles pueden clasificarse como 100% fondo
- 507 Algunos de estos píxeles no pueden clasificarse como 100% fondo o 100% primer plano, pero por ejemplo sí como 57% primer plano
- 508 Algunos de estos píxeles pueden clasificarse como 100% primer plano
- 509 El cuadro de delimitación se define mediante su posición en la esquina inferior izquierda
- 510 ...y su anchura
- 511 ... y su altura

- 601 Un cuadro de delimitación proporcionado por (123) y (124) es demasiado grande comparado con los límites del objeto
- 602 El cuadro de delimitación refinado después del método de recortado del objeto (106)
- 603 Un cuadro de delimitación proporcionado por (123) y (124) interseca los límites de los objetos
- 604 El punto de anclaje del cuadro de delimitación para el método (107) de cálculo de la posición 3D

- 801 El área conocida de la imagen
- 802 El área desconocida de la imagen, es decir, el hueco
- 803a,b Parches de destino solapantes
- 803c,d Parches de destino en una marca (806a) en ambos lados del hueco
- 804a Un parche de origen
- 804b Un parche de origen con diferente tamaño y orientación
- 804c,d Parches de origen en una marca (806a)
- 805 La región solapantes de los parches de destino

ES 2 585 903 T3

- 806a Una marca de línea proyectada
- 806b Una marca de círculo proyectado
- 806c Una marca con una anchura perpendicular a la dirección de la marca
- 807 La parte de la marca dentro del hueco
- 808a,b Normales de la marca circular proyectada
- 809 Ángulo entre dos normales correspondientes a parches
- 810 La longitud de la parte de la marca dentro del hueco
- 811 Límites de una marca (806c) definida por la anchura
- 812 Píxeles clasificados como pertenecientes a la marca (806c)
- 813 Píxeles clasificados como no pertenecientes a la marca (806c)
- 814 Píxeles clasificados como parcialmente pertenecientes a la marca (806c)
- 901 Cancha, campo de juego
- 902 Modelo de estadio 3D
- 903 Valla publicitaria

REIVINDICACIONES

1. Un método para generar una representación 3D de una escena 3D dinámicamente cambiante, que comprende las etapas de
- 5 a) adquirir (102) al menos dos emisiones (120) de video de al menos dos cámaras (702) localizadas en diferentes localizaciones y que observan la misma escena 3D (701);
- b) determinar (103) parámetros (122) de cámara que comprenden la posición, la orientación y parámetros internos para dichas al menos dos cámaras (702);
- 10 c) seguir el movimiento de los objetos (310a,b, 312a,b; 330a,b, 331a,b, 332a,b; 410a,b, 411a,b; 430a,b, 431a,b; 420a,b, 421a,b) en cada una de las al menos dos emisiones (104) de video, en las que un objeto es un jugador o un árbitro o una pelota en una escena deportiva;
- d) determinar la posición 3D de los objetos combinando la información de las al menos dos emisiones (107) de video;
- caracterizado por que** la etapa de seguimiento (104) del movimiento de objetos (310a,b, 312a,b; 330a,b, 331a,b, 332a,b; 410a,b, 411a,b; 430a,b, 431a,b; 420a,b, 421a,b) en cada una de las al menos dos emisiones (12) de video usa la posición 3D de los objetos (130) como se deduce en la etapa de determinación de la posición 3D de los objetos (107); y
- 15 el método comprende la etapa de
- e) determinar la identidad de dichos objetos en las al menos dos emisiones (105) de video asociando objetos con identificadores únicos y asociando así los objetos con información (132) almacenada, previamente conocida sobre objetos del mundo real;
- 20 f) determinar la posición y orientación de los objetos (903) de reproducción 3D correspondientes a los objetos en la emisión de video, en los que los objetos (903) de reproducción 3D sirven para reproducir información de imagen de una o más emisiones de video cuando generan la representación 3D de la escena.
- 25 2. El método de la reivindicación 1, en el que la posición 3D de al menos un objeto se extrapola de al menos dos posiciones 3D previas del objeto (332a,b).
3. El método de una de las reivindicaciones anteriores, en donde la etapa de determinar la identidad de los objetos (105) en las al menos dos emisiones (120) de video usa información que se deduce de una de las etapas (107) posteriores de las al menos dos emisiones (120) de video y surge a partir de uno o más instantes anteriores en el tiempo.
- 30 4. El método de la reivindicación 3, en donde la información deducida del uno o más instantes anteriores en el tiempo es la posición 3D del objeto (130).
5. El método de una de las reivindicaciones anteriores, en donde la etapa de seguir el movimiento de los objetos (104) comprende la etapa de incorporar parámetros (131) de cámara dinámicamente cambiantes en la función (104) de seguimiento, de manera que la función (104) de seguimiento compensa los cambios en los parámetros (131) de cámara.
- 35 6. El método de la reivindicación 5, en donde los parámetros de cámara tenidos en cuenta en la función de seguimiento son parámetros (131) de cámara determinados por una etapa (103) de calibrado de cámara, realizada para el mismo fotograma de video para el que se realiza el seguimiento.
- 40 7. El método de la reivindicación 5, en donde los parámetros de cámara tenidos en cuenta en la función de seguimiento son parámetros (131) de cámara determinados por una etapa (103) de calibrado de cámara, realizada para uno o más fotogramas de video previos y que se extrapolan opcionalmente.
8. El método de una de las reivindicaciones anteriores, que comprende una etapa (106) de segmentación en la cual los objetos se separan del fondo, que comprende la etapa de
- 45 • usar coincidencia de canal alfa para asignar, a cada elemento de la imagen, un valor que expresa la probabilidad con la cual el pixel es parte de un objeto o parte del fondo.
9. El método de una de las reivindicaciones anteriores, que comprende proporcionar (108) una vista sintetizada a partir de un punto de vista virtual que es distinto de las posiciones de la cámara mediante las etapas de:
- proporcionar parámetros de cámara de una cámara virtual (703);
 - determinar una imagen de fondo según se ve por la cámara (703) virtual en un modelo de fondo (901, 902);

- determinar una proyección de cada uno de los objetos en la cámara (703) virtual y superponerla sobre la imagen de fondo; y
 - producir la imagen combinada para almacenamiento o para procesamiento adicional,
- 5 y que comprende además una etapa (106) de segmentación, en la cual los objetos se separan del fondo, que comprende la etapa de
- después de retirar los objetos, llenar los huecos correspondientes o partes desconocidas que quedan en el fondo mediante restauración de imagen y marcar tales elementos de imagen restaurada como datos de imagen sintéticos.
10. El método de la reivindicación 9, en donde la etapa de determinar una imagen de fondo según se ve por la cámara (703) virtual comprende las etapas de
- 10
- combinar, para cada elemento de la imagen de fondo, información de imagen desde las diferentes emisiones de video que corresponden a la misma localización de fondo;
 - dar prioridad a la información de imagen que no está marcada como datos de imagen sintéticos; y
 - hacer que la información de imagen en un modelo de fondo comprenda una o más superficies (901, 902) que representan el fondo.
- 15 11. El método de la reivindicación 10, en el que, en el modelo de fondo, la superficie que representa el fondo es una superficie (901) que representa el campo de juego (701) y opcionalmente también comprende superficies (902) que representan un modelo de entorno 3D.
12. El método de una de las reivindicaciones 9 a 11, en donde la etapa de determinar una imagen de fondo según se ve por la cámara (703) virtual comprende además:
- 20
- reproducir datos de imagen predeterminados en el modelo de fondo (901, 902), superponiéndolos sobre o reemplazando la información de imagen proporcionada por las emisiones de video.
13. El método de una de las reivindicaciones 9 a 12, en donde la etapa de determinar una proyección de cada uno de los objetos en la cámara (703) virtual comprende la etapa de
- 25
- reproducir la información de imagen a partir de una o más emisiones de video en objetos de reproducción 3D (903) situados en el modelo de fondo 3D (901, 902).
14. El método de la reivindicación 13, en donde los objetos de reproducción son vallas publicitarias (903) y se realiza una superposición en la cámara (703) virtual de la proyección de los objetos y la imagen de fondo mediante valores ponderados en canales alfa de las vallas publicitarias (903) basándose en similitud de ángulo, resolución o campo de vista entre las características ópticas de las cámaras (702) reales y la cámara (703) virtual.
- 30 15. Un sistema de procesamiento de datos para generar una representación 3D de una escena 3D dinámicamente cambiante, que comprende
- a) un módulo (102) de adquisición de datos que adquiere al menos dos emisiones de video de al menos dos cámaras localizadas en diferentes localizaciones y que observa la misma escena 3D;
- 35 b) un módulo (103) de calibrado de cámara para determinar parámetros de cámara, que comprende la posición, la orientación y parámetros internos, para dichas al menos dos cámaras;
- c) un módulo (104) de seguimiento 2D para seguir el movimiento de objetos en las al menos dos emisiones de video, en donde un objeto es un jugador o un árbitro o una pelota en una escena deportiva; y
- d) un módulo (107) de combinación 3D y cálculo de posición de objetos 3D para determinar la posición 3D (128) de los objetos combinando la información determinada a partir de las al menos dos emisiones de video;
- 40 **caracterizado por que**
- el módulo (104) de seguimiento 2D está configurado para depender de la posición 3D (128) de los objetos determinada por el módulo (107) de combinación 3D y cálculo de posición de objetos 3D; y **por que** el sistema comprende además
- 45 e) un módulo (105) de identificación de objetos para determinar la identidad de dichos objetos en las al menos dos emisiones de video asociando objetos con identificadores únicos y, de esta manera, asociando los objetos con información (132) almacenada, previamente conocida, sobre objetos del mundo real; y
- por que** el módulo (107) de combinación 3D y cálculo de posición de objetos 3D está configurado para proporcionar

la función de

f) determinar la posición y orientación de objetos (903) de reproducción 3D que corresponden a los objetos en la emisión de video, en donde los objetos (903) de reproducción 3D sirven para reproducir información de imagen de una o más emisiones de video cuando se genera la representación 3D de la escena.

5 16. El sistema de procesamiento de datos de la reivindicación 15, que comprende un módulo (106) de recortado de objetos para determinar

- datos (125) de textura de fondo llenos que incorporan un banderín que especifica si un parche de imagen particular o pixel se deriva de datos de imagen reales o se ha generado sintéticamente,

- una textura de objeto y máscara alfa (126) para cada emisión de video y cada objeto que se sigue y

10 • para cada objeto que se sigue, una posición y forma 2D del objeto y una identificación (127) del objeto del mundo real.

17. El sistema de procesamiento de datos de la reivindicación 16, que comprende

- un módulo (108) de síntesis de imagen que proporciona, a partir de la posición 3D (128) de los objetos, los datos (125) de textura de fondo llenos y la textura de objeto y máscara alfa (126) de los datos de video a un consumidor (109).

15

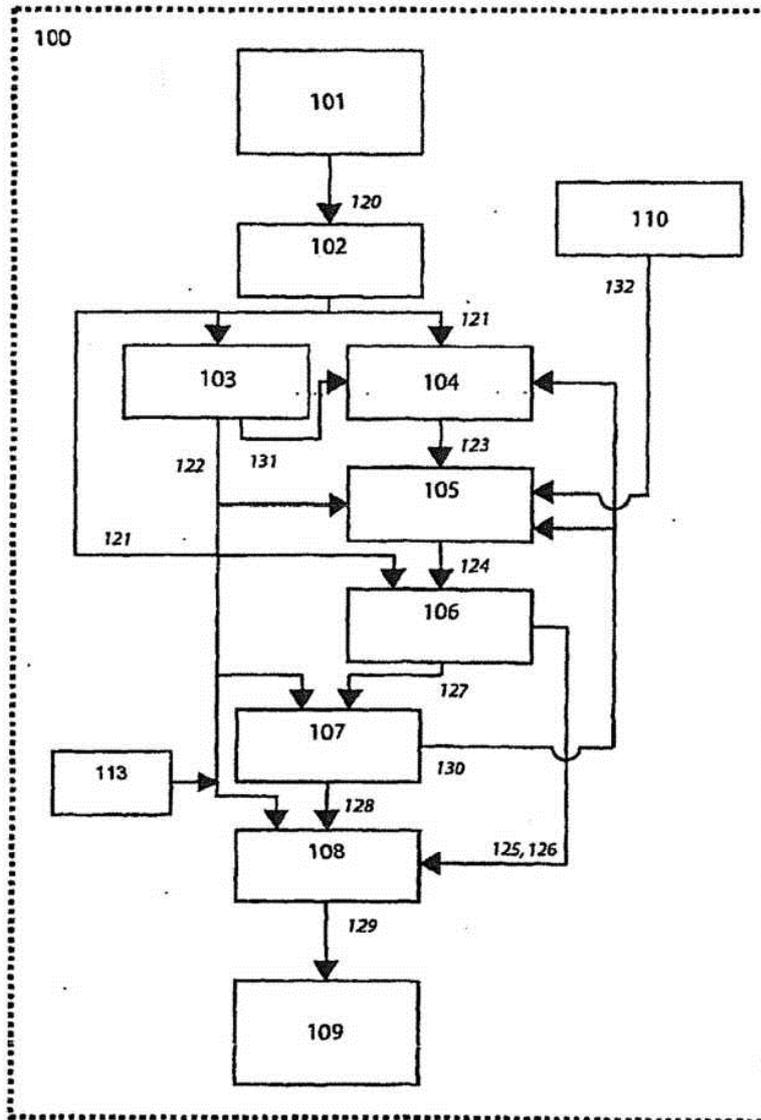


Fig. 1

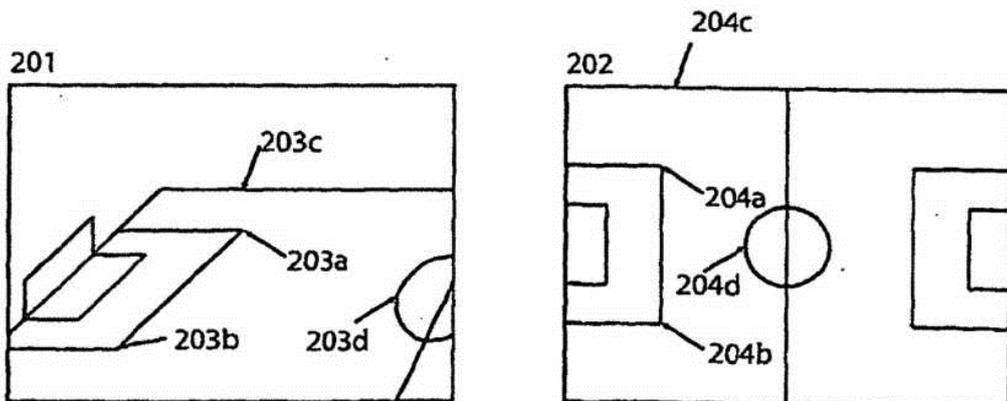


Fig. 2

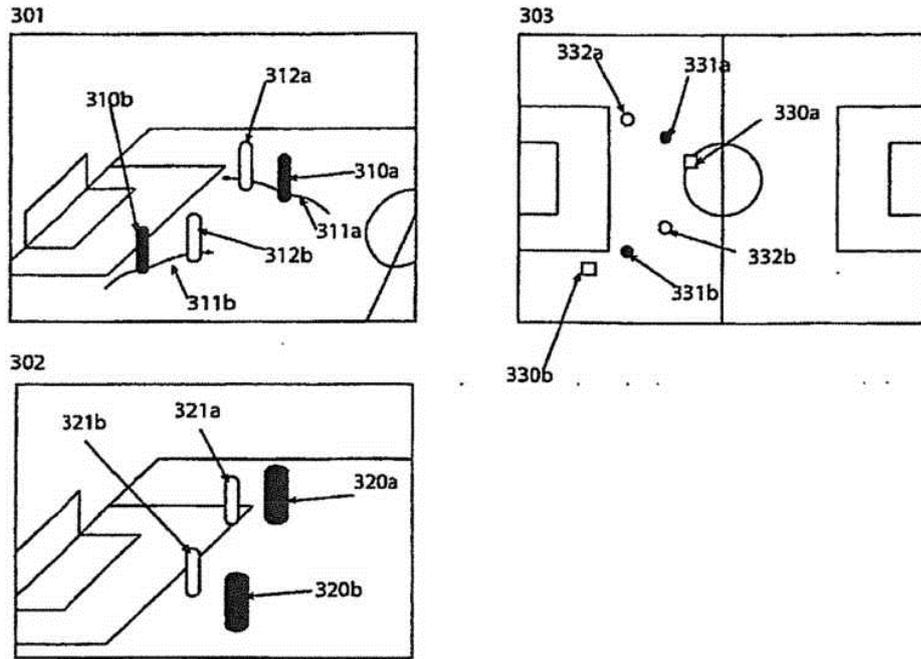


Fig. 3

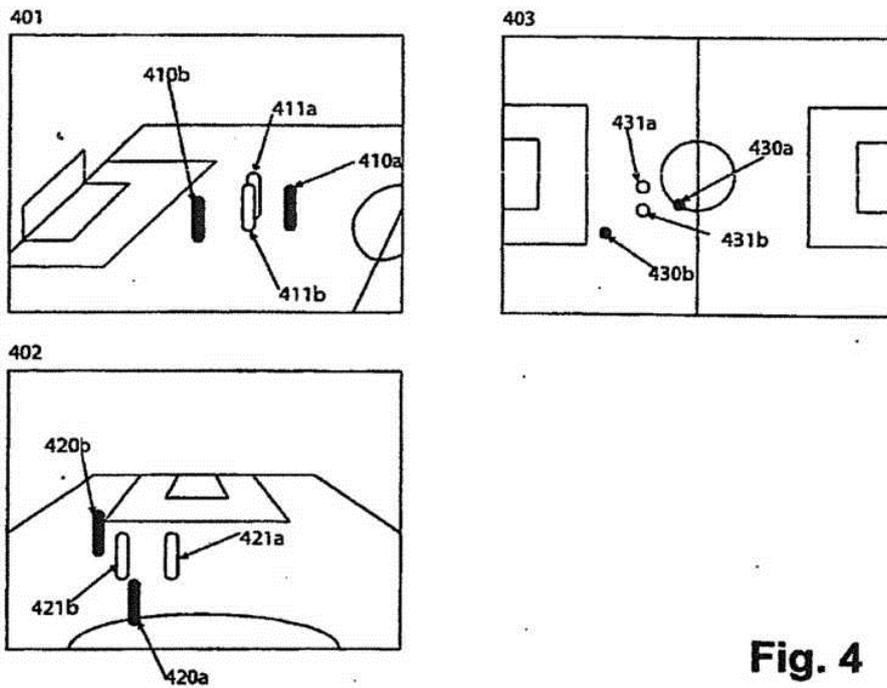


Fig. 4

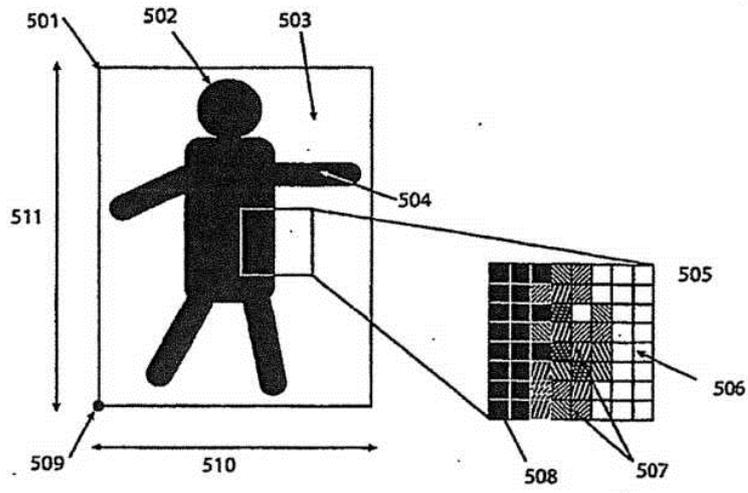


Fig. 5

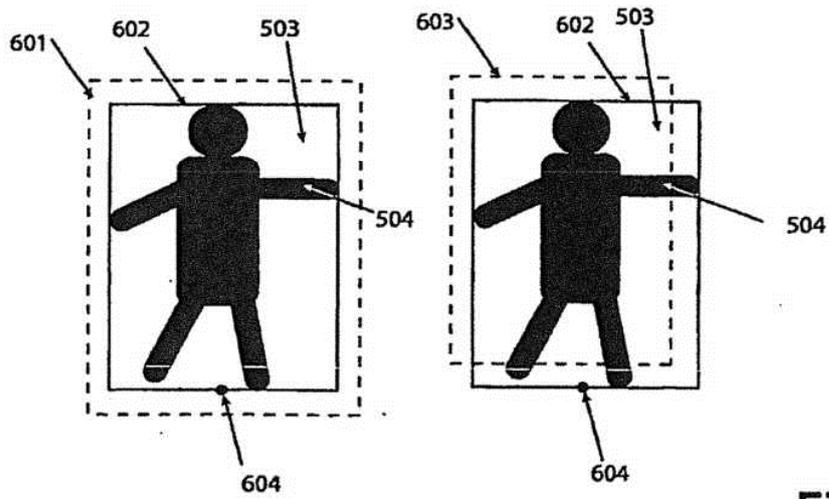


Fig. 6

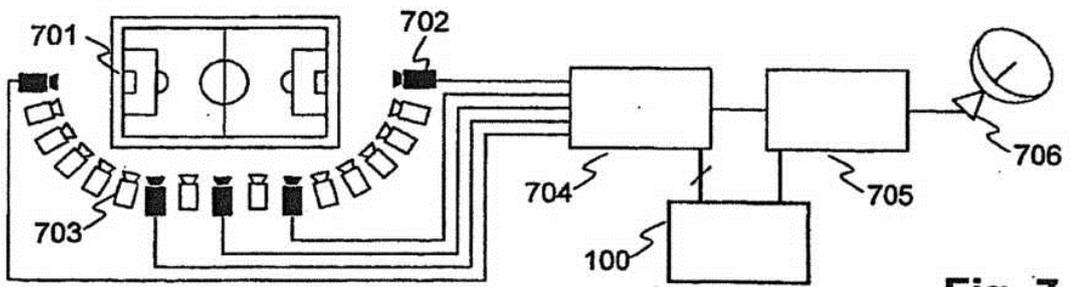
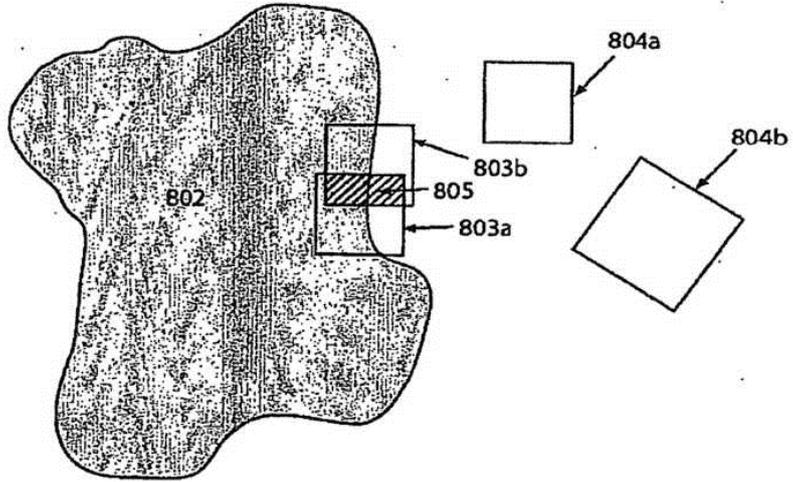


Fig. 7

801



801

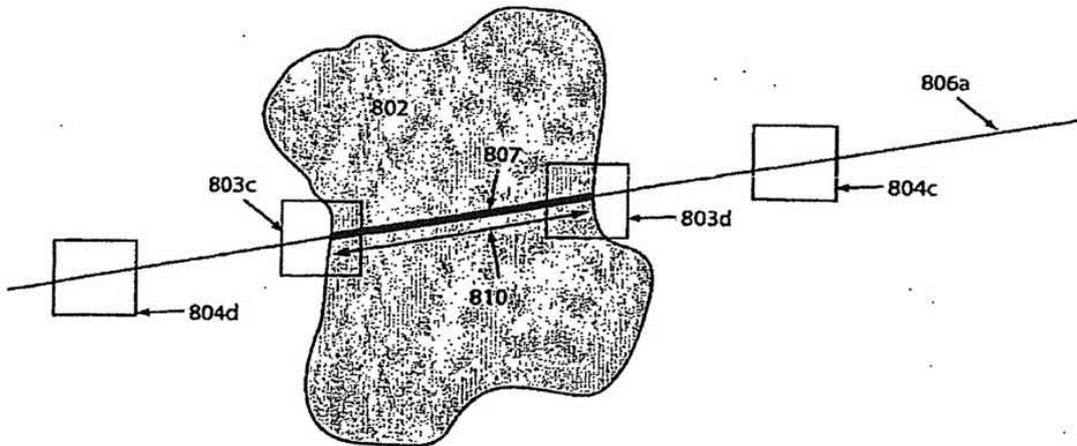


Fig. 8

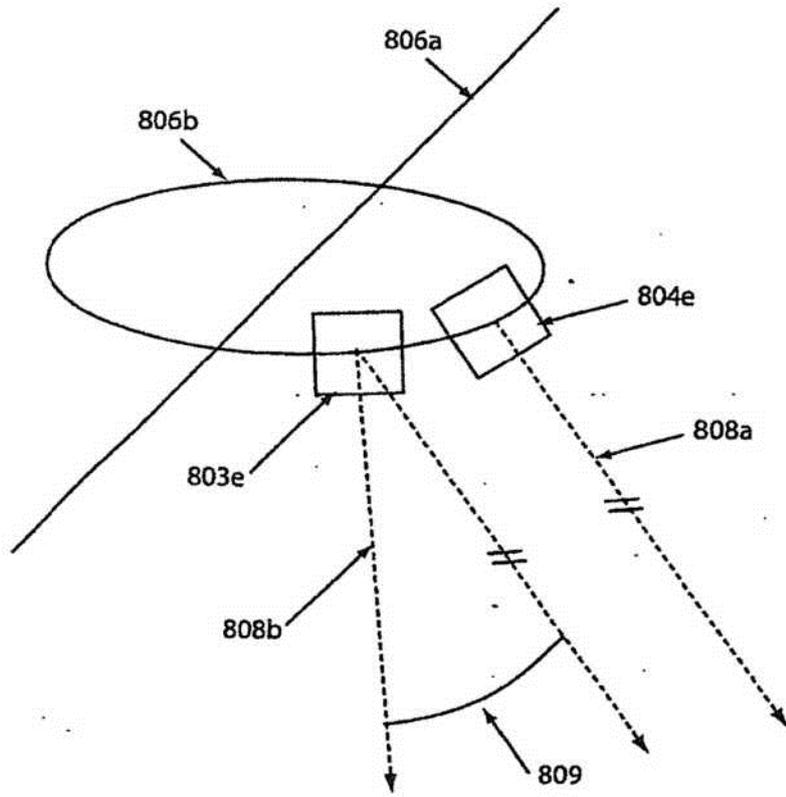


Fig. 9

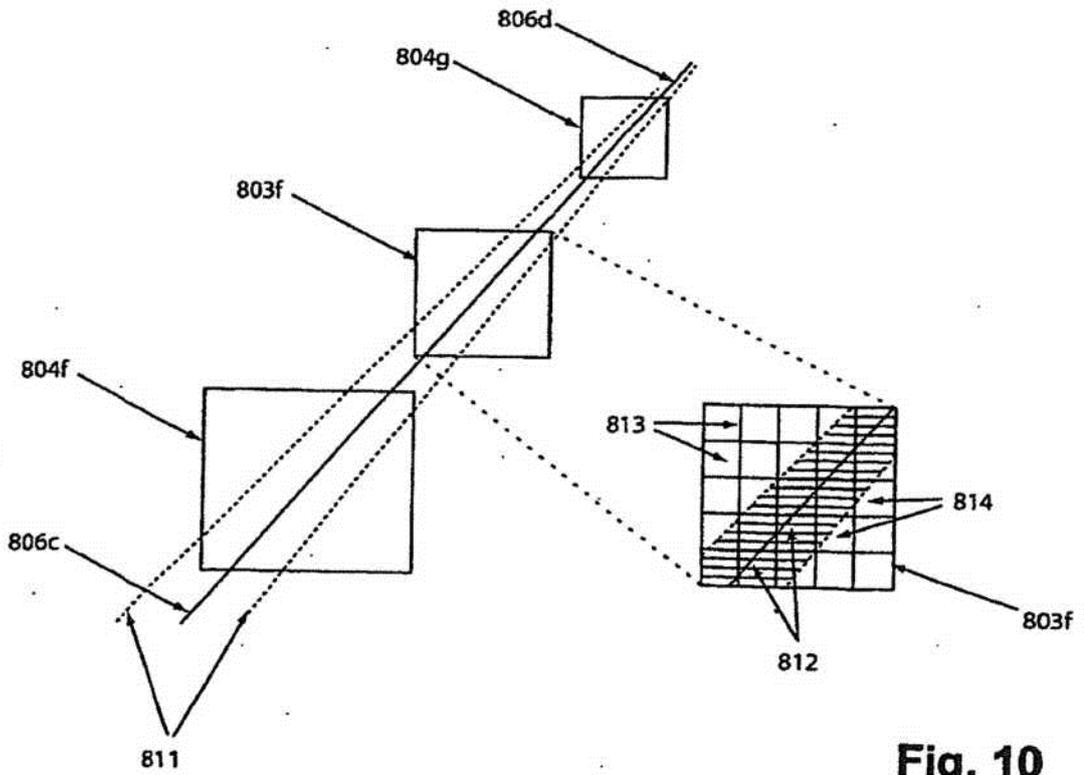


Fig. 10

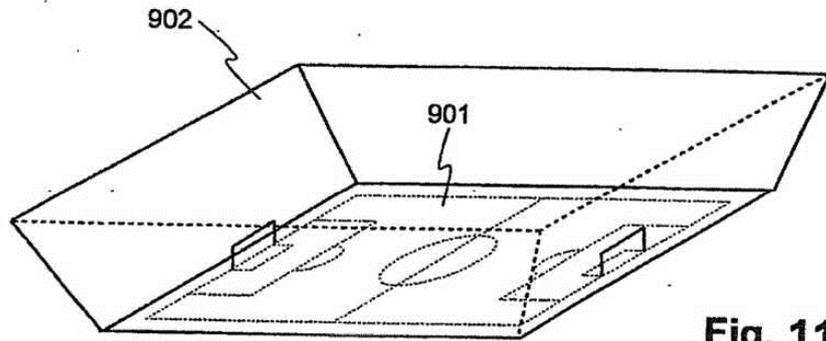


Fig. 11

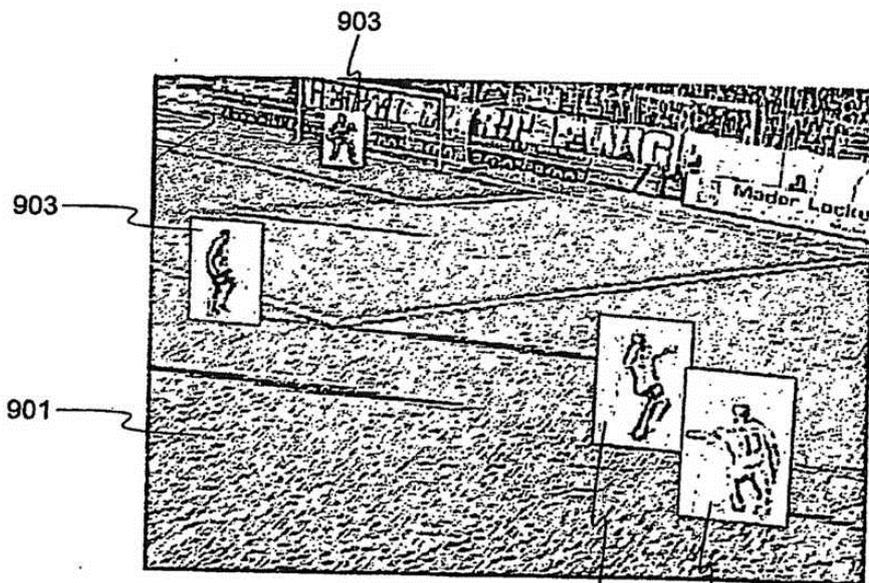


Fig. 12