(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 619 934**

(51) Int. CI.:

H04N 5/262 (2006.01)
G06T 7/00 (2007.01)
G06T 15/20 (2011.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **15.06.2005 E 05105244 (7)**

(97) Fecha y número de publicación de la concesión europea: **04.01.2017 EP 1612732**

(54) Título: **Sistema de video de punto de vista interactivo y procedimiento**

(30) Prioridad:

28.06.2004 US 880774

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.06.2017

(73) Titular/es:

MICROSOFT TECHNOLOGY LICENSING, LLC (100.0%)
One Microsoft Way
Redmond, WA 98052, US

(72) Inventor/es:

ZITNICK III, CHARLES;
UYTTENDAELE, MATTHEW;
SZELISKI, RICHARD;
WINDER, SIMON y
KANG, SING BING

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

DESCRIPCIÓN

Sistema de video de punto de vista interactivo y procedimiento

Antecedentes

Campo técnico

La invención se refiere a la generación y renderización de video y más particularmente a un sistema y procedimiento de generación y renderización de un video de punto de vista interactivo en el que un usuario puede ver una escena dinámica mientras cambia el punto de vista a voluntad.

Antecedentes de la técnica

Desde hace varios años, los espectadores de anuncios de televisión y largometrajes han visto el efecto de "fotograma congelado" usado para crear la ilusión de detener el tiempo y cambiar el punto de vista de cámara. Los primeros anuncios se produjeron usando un sistema basado en película, que rápidamente saltó entre diferentes cámaras fijas dispuestas a lo largo de un rail para dar la impresión de movimiento a través de una fracción de tiempo congelada.

Cuando apareció por primera vez, el efecto fue novedoso y se veía espectacular y pronto se emuló en muchas producciones, el más famoso de los cuales es probablemente los efectos de "tiempo bala" vistos en la película titulada "The Matrix". Desafortunadamente, este efecto es un asunto de una única vez y preplanificado. La trayectoria de punto de vista se planifica con antelación y se gastan muchas horas de trabajo para producir las vistas interpoladas deseadas. Sistemas más nuevos son a base de series de cámaras de video, pero aún dependen de tener muchas cámaras para evitar interpolación de vista de software.

Por lo tanto, los sistemas existentes no permiten que un usuario cambie interactivamente a un punto de vista deseado mientras ve una escena dinámica basada en imágenes. La mayoría del trabajo en renderización basada en imágenes (IBR) en el pasado implica renderizar escenas estáticas, siendo Renderización de Campo de Luz [12] y Lumigraph [8] dos de las técnicas mejor conocidas. Su éxito en la renderización de alta calidad se deriva del uso de un gran número de imágenes muestreadas y ha inspirado una gran masa de trabajo en el campo. Una extensión potencial emocionante de este innovador trabajo implica controlar interactivamente el punto de vista mientras se ve un video. La capacidad de un usuario para controlar interactivamente el punto de vista de un video intensifica la experiencia de visionado considerablemente, permitiendo varias aplicaciones tales como repeticiones instantáneas desde nuevos puntos vista, cambiar el punto de vista en dramatizaciones y crear efectos visuales de "fotograma congelado" a voluntad.

Sin embargo, extender la IBR a escenas dinámicas no es trivial debido a la dificultad (y coste) de la sincronización de tantas cámaras así como la adquisición y almacenamiento de imágenes. No existen únicamente obstáculos significantes que superar en la captura, representación y renderización de escenas dinámicas desde múltiples puntos de vista, sino que ser capaz de hacer esto interactivamente proporciona una complicación adicional significante. Hasta la fecha, los intentos para realizar este objetivo no han sido muy satisfactorios.

Con respecto a los aspectos de renderización basada en video de un sistema de video de punto de vista interactivo, uno de los primeros intentos en capturar y renderizar escenas dinámicas fue el Sistema de Realidad Virtualizada de Kanade y col. [11], que implicaba 51 cámaras dispuestas alrededor de una cúpula geodésica de 5 metros. La resolución de cada cámara es 512×512 y la tasa de captura es 30 fps. Extrajeron una representación de superficie global en cada periodo de tiempo, usando una forma de coloreado de vóxel [15] a base de la ecuación de flujo de escena [18]. Desafortunadamente, los resultados no parecían realistas debido a la baja resolución, errores de coincidencia y manejos impropios de los límites de objeto.

Carranza y col. [3] utilizaron siete cámaras sincronizadas distribuidas alrededor de una habitación mirando hacia su centro para capturar movimiento humano en 3D. Cada cámara es de resolución CIF (320×240) y captura a 15 fps. Usaron un modelo humano en 3D como previo para calcular la forma en 3D en cada periodo de tiempo.

Yang y col. [21] diseñaron una red de cámaras de 8×8 (cada una de 320×240) para la captura de una escena dinámica. En lugar de almacenar y renderizar los datos, transmitían únicamente los rayos necesarios para componer la imagen virtual deseada. En su sistema, las cámaras no se sincronizan; en su lugar, dependen en relojes internos a través de seis PC. La tasa de captura de cámara es 15 fps y la tasa de visionado interactiva es de 18 fps.

Como una prueba de concepto para el almacenamiento de campos de luz dinámicos, Wilburn y col. [20] demostraron que es posible sincronizar seis cámaras (640×480 a 30 fps) y comprimir y almacenar toda la información de imágenes en tiempo real. Desde entonces han conectado 128 cámaras.

Se requieren muchas imágenes para la renderización realista si la geometría de la escena o bien se desconoce o se conoce únicamente en aproximación aproximada. Si la geometría se conoce con precisión, es posible reducir el requisito de imágenes sustancialmente [8]. Una forma práctica de extraer la geometría de la escena es a través de

estéreo y se han propuesto muchos algoritmos de estéreo para escenas estáticas [14]. Sin embargo, ha habido pocos intentos de emplear técnicas de estéreo con escenas dinámicas. Como parte del trabajo de Realidad Virtualizada [11], Vedula y col. [18] propusieron un algoritmo para la extracción de movimiento en 3D (es decir, correspondencia entre formas de escena a través del tiempo) usando flujo óptico en 2D y forma de escena en 3D. En su enfoque, usaron un esquema de elección similar al coloreado de vóxel [15], en el que la medida usada es cómo de bien encaja la ubicación de vóxel hipotética en la ecuación de flujo en 3D.

Zhang y Kambhamettu [22] también integraron en su esquema flujo de imagen y estructura en 3D. Su modelo de movimiento afín en 3D se usa localmente, con regularización espacial y las discontinuidades se preservan usando segmentación de color. Tao y col. [17] asumen que la escena es plana en tramos. También asumen velocidad constante para cada parche plano para restringir la estimación de mapa de profundidad dinámica.

En una tentativa más ambiciosa, Carceroni y Kutulakos [2] recuperan la geometría y reflectancia continua en tramos (modelo Phong) en movimiento no rígido con posiciones de iluminación conocidas. Discretizan el espacio en elementos de superficie ("surfels") y realizan una búsqueda por parámetro de ubicación, orientación y reflectancia para maximizar la concordancia con las imágenes observadas.

En un interesante giro a la coincidencia de ventanas local convencional, Zhang y col. [23] usan ventanas coincidentes que traspasan espacio y tiempo. La ventaja de este procedimiento es que existe menos dependencia en constancia de brillo con el paso del tiempo.

También se han aplicado técnicas de telemetría activa para escenas en movimiento. Hall-Holt y Rusinkiewicz [9] usan patrones de raya codificados en límites proyectados que varían con el paso del tiempo. También existe un sistema comercial en el mercado llamado ZCam™ fabricado por 3DV Systems de Israel, que es un complemento de cámara de video con detector de intervalo usado en conjunción con una cámara de video de teledifusión. Sin embargo, es un sistema caro y proporciona únicamente una única profundidad de punto de vista, que lo hace menos adecuado para video de múltiple punto de vista.

Sin embargo, a pesar de todos los avances en renderización en estéreo y basada en imágenes, todavía es muy difícil renderizar interactivamente vistas de alta calidad y alta resolución de escenas dinámicas. La presente invención aborda este problema de una manera de eficiencia de coste.

Se observa que en los párrafos anteriores, así como en el resto de la memoria descriptiva, la descripción se refiere a diversas publicaciones individuales identificadas mediante una designación numérica contenida entre un par de corchetes. Por ejemplo, tal referencia puede identificarse recitando "referencia [1]" o simplemente "[1]". Múltiples referencias se identificaran mediante un par de corchetes conteniendo más de una designación, por ejemplo, [2, 3]. Un listado de referencias que incluye las publicaciones que corresponden a cada designación puede encontrarse al final de la sección de Descripción Detallada.

B. Chai, "Depth Map Compression for Real-time view-based Rendering", escritos 25 de reconocimiento de patrones (2004), páginas 755 - 766 se refiere a captura, transmisión y exploración de escenas en 3D dinámicas que se capturan usando una colección de cámaras de video sincronizadas. La secuencia de mapas de profundidad se calcula en tiempo real desde el punto de vista de cada cámara usando retransmisiones de video de color múltiple. Cada retransmisión de mapa de profundidad se comprime junto con la correspondiente retransmisión de video para producir una retransmisión de video en 3D que puede retransmitirse por internet. Para la separación de primer plano/fondo, se procesa el fondo separado y completo. A continuación, el mapa de profundidad en tiempo real se pre-procesa de tal forma que los valores de profundidad de primer plano permanecen sin cambios y a las regiones de fondo se asignan valores de profundidad infinitos.

R. Krishnamurthy et al "Compression and Transmission of Depth Maps for Image-Based Rendering", actas de la Conferencia Internacional de 2001 sobre Procesamiento de Imágenes, páginas 828 - 831 considera la codificación de regiones de interés, en la que se identifican esas regiones de la imagen donde la profundidad precisa es más crucial. El intervalo dinámico del mapa de profundidad se remodela.

Sumario

La invención proporciona un procedimiento de acuerdo con las reivindicaciones 1 y 13 y un sistema de acuerdo con las reivindicaciones 21 y 35.

La presente invención se dirige hacia un sistema y procedimiento para primero generar, y a continuación renderizar y presentar un video de punto de vista interactivo en el que un usuario puede ver una escena dinámica mientras manipula (congelar, ralentizar o dar marcha atrás) el tiempo y cambia el punto de vista a voluntad. Ya que pueden tomarse diferentes trayectorias a través del tiempo-espacio, dos experiencias de visionado no necesitan ser iguales. En general, esto se consigue con un sistema y procedimiento que usa un pequeño número de cámaras para capturar múltiples retransmisiones de video en alta calidad de una escena. No solo esto reduce el coste asociado con la captura de las retransmisiones de video necesarias para renderizar la escena desde una diversidad de puntos de vista sobre procedimientos existentes, sino que también es ventajoso ya que hace el sistema de captura portátil. A continuación se emplea una técnica de acoplamiento y reconstrucción en 3D multivista para crear una

representación en capas de los fotogramas de video que permite tanto la compresión efectiva como reproducción interactiva de la escena dinámica capturada, mientras que permite al mismo tiempo la renderización en tiempo real.

Más particularmente, el sistema generador de video de punto de vista interactivo tiene un subsistema de captura de video que incluye un banco de cámaras de video para la captura de retransmisiones de video y equipo de sincronización para la sincronización de las retransmisiones de video. Colectivamente, las retransmisiones de video se pueden caracterizar por ser grupos secuenciales de fotogramas de video capturados contemporáneamente cada uno representando un diferente punto de vista de una escena. El subsistema de captura también puede incluir equipo de almacenamiento para el almacenamiento de las retransmisiones de video antes del procesamiento.

Las cámaras de video se disponen en una manera de una al lado de la otra de tal forma que cada cámara ve la escena desde un punto de vista diferente. El campo de vista de cada cámara superpone el campo de vista de cualquier cámara adyacente por una cantidad prescrita (por ejemplo, campo de vista horizontal de 30 grados por cámara con rotación relativa de 4-8 grados entre dos cámaras adyacentes). La distancia, altura y colocación horizontal del punto de vista de cada cámara en relación con un objeto o área objetivo de la escena se establece para formar una trayectoria deseada que conecta los puntos de vista de las cámaras. Esta trayectoria puede ser, por ejemplo, un arco horizontal o vertical o un arco que recorre hacia arriba o hacia fuera desde un extremo al otro. La importancia de esta trayectoria es que un usuario que ve el video de punto de vista interactivo puede seleccionar un punto de vista para ver la escena capturada desde cualquier sitio a lo largo de la trayectoria, incluso si no coincide con una de las posiciones de cámara.

Las cámaras de video pueden ser cámaras de alta resolución ya que esto mejorará la calidad del video de punto de vista interactivo que se genera. Además, las cámaras pueden tener una característica de sincronizador para facilitar la sincronización de las retransmisiones de video. Además, sería ventajoso para facilitar el procesamiento de las retransmisiones de video si las cámaras pueden añadir metadatos a cada fotograma de video generado. Estos metadatos incluirían, por ejemplo, los ajustes actuales y nivel de exposición de la cámara y una marca de tiempo.

Además de las cámaras y equipo de sincronización, el subsistema de captura incluye un dispositivo informático, que ejecuta un programa de calibrado de cámara para el cálculo de parámetros geométricos y fotométricos asociados con cada retransmisión de video. Además, las cámaras de video pueden ser del tipo que se pueden controlar por ordenador. Si es así, el anteriormente mencionado dispositivo informático también puede usarse para ejecutar un programa de captura de video que apaga o enciende las cámaras simultáneamente y ajusta sus ajustes, a base de un menú preestablecido o en respuesta a entradas de usuario.

El sistema generador de video de punto de vista interactivo también emplea el anteriormente mencionado dispositivo informático o un dispositivo informático diferente para ejecutar un programa de generación de video. En general, este programa de generación se usa para primero producir una reconstrucción en 3D de la escena representada en cada grupo de fotogramas contemporáneos de las retransmisiones de video sincronizadas. En realizaciones ensayadas del programa de generación, se empleó una técnica de reconstrucción basada en segmentación para este fin. Para cada grupo de fotogramas contemporáneos, la reconstrucción resultante se usa para calcular un mapa de disparidades para cada fotograma en el grupo. La reconstrucción también puede usarse para calcular correspondencias a través de todas las vistas de cámara en el grupo fotogramas. Estas correspondencias pueden usarse a continuación, entre otras cosas, para equilibrar los parámetros fotométricos entre los fotogramas en el grupo. Además, para cada fotograma en el grupo en consideración, se definen áreas de discontinuidades de profundidad significantes a base de su mapa de disparidades. Dentro de estas áreas, se calcula información de primer plano y fondo. Usando esta información, se establece una capa principal separada y capa límite para cada fotograma. La capa principal se compone de píxeles que tienen valores de disparidad que no muestran discontinuidades de profundidad que exceden un umbral prescrito y la información de fondo de las áreas que rodean discontinuidades de profundidad. La capa límite se compone de la información de primer plano en áreas cerca de o que tienen discontinuidades de profundidad que exceden el umbral. Por lo tanto, se produce una representación en capas de cada fotograma. La capa principal se caracteriza por un color y profundidad de cada píxel en la capa. Sin embargo, la capa límite se caracteriza no solo por un color y profundidad de cada píxel, sino también un valor de opacidad para cada píxel en la capa. Se observa que en realizaciones ensayadas del programa de generación de video, la capa límite se dilató para incluir un número prescrito de píxeles adyacentes a los píxeles que muestran discontinuidades de profundidad que exceden el umbral. Haciéndose de esta forma para evitar la aparición de "fracturas" durante el procedimiento de renderización que se describirá en breve.

El programa de generación de video de punto de vista interactivo puede adicionalmente incluir una provisión para la compresión de las anteriormente mencionadas representaciones en capas para facilitar transferencia y/o almacenamiento del video. Esta compresión puede emplear técnicas de compresión temporales o espaciales o como en realizaciones ensayadas un enfoque de compresión temporal y espacial combinado. Aún más, el programa de generación puede tener un módulo de programa de generación de archivo para la creación de un archivo que incluye las representaciones en capas de los fotogramas de video de punto de vista interactivo y los datos de calibración anteriormente calculados.

Se observa que el programa de generación de video también puede emplearse para crear un video de punto de vista interactivo de retransmisiones de video generadas por ordenador, en vez de fotogramas capturados por cámaras de

video reales. En este caso, los datos de calibración serían proporcionados directamente por el programa de generación.

El anteriormente mencionado sistema de renderización y presentación que se usa para reproducir el video de punto de vista interactivo tiene un subsistema de interfaz de usuario para la introducción de selecciones de punto de vista de usuario y la presentación de fotogramas de video de punto de vista interactivo renderizados al usuario. Para este fin, el sistema incluye un dispositivo de entrada de algún tipo que se emplea por el usuario para introducir selecciones de punto de vista de forma continua y un dispositivo de presentación usado para presentar los fotogramas de video de punto de vista interactivo renderizados al usuario. Además, el subsistema de interfaz de usuario puede incluir una interfaz de usuario gráfica que permite al usuario indicar gráficamente el punto de vista a través del dispositivo de entrada.

El sistema de renderización y presentación incluye adicionalmente un dispositivo informático en el que se ejecuta un programa de renderización. Este programa de renderización se usa para renderizar cada fotograma del video de punto de vista interactivo. En general, para cada fotograma renderizado, esto incluye primero identificar el punto vista actual especificado por el usuario. A continuación se identifica el fotograma o fotogramas que se necesitan, del grupo de fotogramas contemporáneos asociado con la porción temporal actual del video que se reproduce, para renderizar el fotograma actual del video de punto de vista interactivo desde el punto de vista identificado. Esto generalmente implica usar los datos de calibración para determinar los puntos de vista asociados con cada fotograma en el grupo actual y a continuación determinar si el punto de vista actual seleccionado por el usuario coincide con uno de los puntos de vista de fotograma o se sitúa **entre dos de los fotogramas.** Cuando el punto de vista identificado coincide con un punto de vista de uno de los fotogramas de video, ese fotograma se identifica como el fotograma que se necesita para renderizar la escena. Sin embargo, cuando el punto de vista identificado se sitúa entre los puntos de vista **de dos de los fotogramas de video, ambos de estos fotogramas** se definen como los fotogramas que se necesita para renderizar la escena.

Una vez que se define el fotograma o fotogramas requeridos, se obtienen las representaciones en capas que corresponden a estos fotogramas. Esto como mínimo implica la extracción de solo los datos de fotograma necesarios de los datos de punto de vista interactivo. Si los datos de video se contienen en un fichero, como se ha descrito anteriormente, habitualmente necesitarán ser decodificados. Un módulo de decodificación selectivo del programa de renderización puede emplearse para este fin. Además, si los datos de fotograma en capas se han comprimido, el módulo decodificador es responsable de la descompresión de la porción de los datos de video necesarios para recuperar los fotogramas particulares requeridos para renderizar la escena desde el punto de vista deseado.

Los datos de fotograma decodificados se usan para renderizar el siguiente fotograma del video de punto de vista interactivo desde el punto de vista especificado en la actualidad por el usuario. Este es un procedimiento sencillo si el punto de vista especificado coincide con el punto de vista asociado con un fotograma decodificado. **Sin embargo, si el punto de vista deseado se sitúa entre dos fotogramas, el procedimiento de renderización es más complicado. En una realización del procedimiento de renderización, esto implica, para cada uno de los dos fotogramas de entrada sucesivamente, primero proyectar la capa principal del fotograma de entrada en consideración en una vista virtual que corresponde al punto vista actual especificado por el usuario y a continuación también proyectar la capa límite del fotograma de entrada en consideración en la vista virtual.** Las capas de límite proyectadas y capas principales proyectadas se combinan para crear un fotograma finalizado del video de punto de vista interactivo. Se observa que la combinación de las capas proyectadas implica ponderar cada capa en proporción directa a cómo de cerca está el punto de vista asociado con esa capa al punto vista actual especificado por el usuario.

Se observa que el procedimiento de renderización puede adicionalmente incluir insertar un objeto no encontrado en las representaciones de fotograma en capas introducidas en el fotograma que se está renderizando. Estos objetos pueden generarse por ordenador o basarse en imágenes.

Además de los beneficios recién descritos, otras ventajas de la presente invención resultarán evidentes de la descripción detallada que sigue en lo sucesivo cuando se toma en conjunción con las figuras de los dibujos que acompañan a la misma.

Descripción de los dibujos

Las características específicas, aspectos y ventajas de la presente invención se entenderán mejor con respecto a la siguiente descripción, reivindicaciones adjuntas y dibujos adjuntos en los que:

La Figura 1 es un diagrama que representa un dispositivo informático de fin general que constituye un sistema ilustrativo de implementación de la presente invención.

La Figura 2 es una imagen que representa una realización de ejemplo de la configuración de hardware del sistema de captura de video de punto de vista interactivo de acuerdo con la presente invención.

La Figura 3 es un diagrama de bloques que muestra la arquitectura de módulo de programa informático general para una realización del programa de generación de video de punto de vista interactivo de acuerdo con la presente invención.

Las Figuras 4A y 4B son un diagrama de flujo que representan una realización en forma de diagrama del procedimiento de generación de video de punto de vista interactivo de acuerdo con la presente invención.

La Figura 5 es un diagrama de bloques que muestra la arquitectura de módulo de programa informático general para una realización del programa de renderización de video de punto de vista interactivo de acuerdo con la presente invención.

Las Figuras 6(a)-(c) son una serie de imágenes que muestran un ejemplo de los resultados del procedimiento de renderización de video de punto de vista interactivo de acuerdo con la presente invención. La Figura 6(a) y (c) representan fotogramas capturados contemporáneamente desde dos cámaras de video adyacentes ubicadas en diferentes puntos de vista. La Figura 6(b) es un ejemplo del fotograma que se renderiza cuando el punto de vista especificado por el usuario está entre los puntos de vista asociados con las imágenes de la Figura 6(a) y (c).

La Figura 7 es una imagen que muestra un fotograma de un video de punto de vista interactivo creado de acuerdo con la presente invención, en la que se ha insertado una copia extra de un bailarín de *breakdance*.

La Figura 8 es un diagrama de flujo que representa una realización en forma de diagrama del procedimiento de renderización de video de punto de vista interactivo de acuerdo con la presente invención.

Descripción detallada de las realizaciones preferidas

En la siguiente descripción de las realizaciones preferidas de la presente invención, se hace referencia a los dibujos adjuntos que forman una parte de la misma y en la que se muestran por medio de ilustración realizaciones específicas en las que la invención puede practicarse. Se entiende que pueden utilizarse otras realizaciones y pueden hacerse cambios estructurales sin alejarse del ámbito de la presente invención.

1.0 Video de punto de vista interactivo

La capacidad de controlar interactivamente el punto de vista mientras se ve un video es una emocionante nueva aplicación de renderización basada en imágenes. El presente sistema y procedimiento proporciona esta capacidad renderizando escenas dinámicas con control de punto de vista interactivo usando múltiples retransmisiones de video sincronizadas combinadas con nuevas técnicas de modelado basado en imágenes y de renderización. Porque cualquier vista intermedia puede sintetizarse en cualquier momento, con el potencial para la manipulación del espacio-tiempo, el presente enfoque se ha denominado *video de punto de vista interactivo*.

Una ventaja clave del presente sistema de video de punto de vista interactivo y procedimiento es que se proporcionan fotogramas de vista interpolada de la calidad más alta posible para mejorar la experiencia de visionado, incluso aunque se usen relativamente pocas cámaras para capturar la escena vista. Esto no se logra fácilmente. Un enfoque, como se sugiere en el artículo Renderización de Campo de Luz [12], es simplemente muestrear de nuevo rayos basado únicamente en las posiciones relativas de las cámaras de entrada y virtuales. Como se demuestra en el Lumigraph [8] y posterior trabajo, sin embargo, usar un impostor o proxy 3D para la geometría de la escena puede mejorar enormemente la calidad de las vistas interpoladas. Otro enfoque es para crear un modelo 3D de textura mapeada [11], pero esto generalmente produce resultados inferiores al uso de múltiples vistas de referencia.

El presente sistema y procedimiento emplea el enfoque de renderización asistida por geometría basada en imágenes, que requiere un proxy 3D. Una posibilidad es usar un único modelo poliédrico global, como en los artículos Lumigraph y Lumigraph No Estructurado [1]. Otra posibilidad es usar profundidad por píxel, como en Imágenes de Profundidad en capas [16], mapas de profundidad de compensación en Fachada [6] o conjuntos de imágenes agrupadas en una sola imagen con profundidad [16]. En general, usar diferentes proxys geométricos locales para cada vista de referencia [13, 7, 10] produce resultados de mayor calidad, así que se adopta ese enfoque.

Para obtener la calidad más alta posible para un número fijo de imágenes de entrada, se usan mapas de profundidad por píxel. Estos mapas de profundidad se generan por la nueva técnica de reconstrucción en 3D que se describirá en breve. Sin embargo, incluso múltiples mapas de profundidad todavía muestran impurezas de renderización cuando generan vistas nuevas, es decir, distorsión de pixelado (dentado) debido a la naturaleza abrupta de la transición de primer plano a fondo y colores contaminados debido a píxeles mezclados, que se hacen visibles cuando se compone en nuevos segundos planos u objetos.

Estos problemas se tratan usando una representación de dos capas única. Esta representación se genera localizando primero las discontinuidades de profundidad en un mapa de profundidad d_i, y en una realización de la invención, creando una tira de límite alrededor de los píxeles. A continuación se usa una variante de acoplamiento bayesiano [5] para estimar los colores del límite y capa principal, profundidades y opacidades (valores alfa). Para

reducir el tamaño de datos, las imágenes de profundidad de acoplamiento alfa múltiples pueden comprimirse como también se describirá en breve.

En el momento del renderizado, se eligen las dos vistas de referencia más cercanas a la vista virtual. Cada vista se renderiza independientemente deformando cada una de sus dos capas usando el mapa de profundidad asociado y componiendo la capa límite sobre la capa principal. Las dos vistas deformadas son a continuación combinadas a base de su proximidad a la vista nueva. Como alternativa, las capas principales y capas de límite de cada vista pueden renderizarse independientemente y a continuación combinarse juntas. Una descripción más detallada de este procedimiento también se proporcionará más adelante.

Las siguientes secciones presentarán detalles del presente sistema de video de punto de vista interactivo y procedimiento. Se presenta primero un entorno informático adecuado para la implementación de las porciones programáticas del presente sistema. A esto le sigue una descripción del subsistema de captura de video de punto de vista interactivo usado para capturar y sincronizar múltiples videos. A continuación se presentan descripciones de los programas de sistema de video de punto de vista interactivo.

1.1 El entorno informático

Antes de proporcionar una descripción de las realizaciones preferidas de la presente invención, se describirá una descripción breve y general de un entorno informático adecuado en el que pueden implementarse las porciones de la invención. La Figura 1 ilustra un ejemplo de un entorno 100 de sistema informático adecuado. El entorno 100 de sistema informático es únicamente un ejemplo de un entorno informático adecuado y no pretende sugerir ninguna limitación al ámbito de uso o funcionalidad de la invención. El entorno 100 informático tampoco debe interpretarse como que tiene alguna dependencia o requisito relativo a alguno o combinación de componentes ilustrados en el entorno 100 operativo ilustrativo.

La invención es operacional con numerosos otros entornos o configuraciones de sistema informático de fin general o fin especial. Ejemplos de sistemas informáticos bien conocidos, entornos y/o configuraciones que pueden ser adecuados para su uso con la invención incluyen, pero sin limitación, ordenadores personales, ordenadores de servidores, dispositivos de mano o portátiles, sistemas con multiprocesador, sistemas basados en microprocesadores, decodificadores de salón, electrónica de consumo programable, PC de red, mini ordenadores, ordenadores centrales, entornos informáticos distribuidos que incluyen cualquier de los sistemas o dispositivos anteriores y similares.

La invención puede describirse el contexto general de instrucciones ejecutables por ordenador, tales como módulos de programa, que se ejecutan por ordenador. En general, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc. que realizan tareas particulares o implementan tipos abstractos de datos particulares. La invención también puede practicarse en entornos informáticos distribuidos en los que las tareas se realizan mediante dispositivos de procesamiento remoto que se enlazan a través de una red de comunicaciones. En un entorno informático distribuido, módulos de programa pueden ubicarse tanto en medio de almacenamiento informático local como remoto incluyendo dispositivos de almacenamiento de memoria.

Con referencia a la Figura 1, un sistema ilustrativo de implementación de la invención incluye un dispositivo informático de fin general en la forma de un ordenador 110. Componentes del ordenador 110 pueden incluir, pero sin limitación, una unidad 120 de procesamiento, una memoria 130 de sistema y un bus 121 de sistema que acopla diversos componentes de sistema incluyendo la memoria de sistema a la unidad 120 de procesamiento. El bus 121 de sistema puede ser cualquiera de diversos tipos de estructuras de bus incluyendo un bus de memoria o controlador de memoria, un bus periférico y un bus local usando cualquiera de una diversidad de arquitecturas de bus. A modo de ejemplo, y no como limitación, tales arquitecturas incluyen bus de Arquitectura Estándar de la Industria (ISA), bus de Arquitectura Micro Canal (MCA), bus de ISA Mejorada (EISA), bus local de Asociación de Normalización en la Electrónica de Video (VESA) y bus de Interconexión de Componentes Periféricos (PCI) también conocido como bus de entresuelo.

El ordenador 110 habitualmente incluye una diversidad de medios legibles por ordenador. Medio legible por ordenador puede ser cualquier medio disponible que puede accederse mediante el ordenador 110 e incluye tanto medios volátiles como no volátiles, medios extraíbles y no extraíbles. A modo de ejemplo, y no como limitación, medio legible por ordenador puede comprender medio de almacenamiento informático y medio de comunicación. Medio de almacenamiento informático incluye tanto volátiles como no volátiles, medio extraíble y no extraíble implementados en cualquier procedimiento o tecnología para almacenamiento de información tales como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Medio de almacenamiento informático incluye, pero sin limitación, RAM, ROM, EEPROM, memoria flash u otra tecnología de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento de disco óptico, casetes magnéticos, cinta magnética, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético o cualquier otro medio que puede usarse para almacenar la información deseada y que puede accederse mediante el ordenador 110. Medio de comunicación habitualmente contiene instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos en una señal de datos modulada tales como una onda portadora u otro mecanismo de transporte e incluye cualquier medio de distribución de información. La expresión "señal de datos

modulada" significa una señal que tiene una o más de sus características ajustadas o cambiadas de tal manera como para codificar información en la señal. A modo de ejemplo, y no como limitación, medio de comunicación incluye medios cableados tales como una red de cable o conexión directa por cable y medios inalámbricos tales como medios inalámbricos acústicos, RF, infrarrojos u otro. También deben incluirse combinaciones de cualquiera de los anteriores dentro del ámbito de medio legible por ordenador.

La memoria 130 de sistema incluye medio de almacenamiento informático en la forma de una memoria volátil y/o no volátil tales como memoria 131 de sólo lectura (ROM) y memoria 132 de acceso aleatorio (RAM). Un sistema 133 de entrada/salida básico (BIOS), que contiene las rutinas básicas que ayudan a transferir información entre elementos dentro del ordenador 110, tales como durante el arranque, habitualmente se almacena en ROM 131. RAM 132 habitualmente contiene datos y/o módulos de programa que son inmediatamente accesibles y/o siendo actualmente operados mediante unidad 120 de procesamiento. A modo de ejemplo, y no como limitación, la Figura 1 ilustra sistema 134 operativo, programas 135 de aplicación, otros módulos 136 de programa y datos 137 de programa.

El ordenador 110 también puede incluir otro medio de almacenamiento informático extraíble/no extraíble, volátil/no volátil. Únicamente a modo de ejemplo, la Figura 1 ilustra una unidad 141 de disco duro que lee de o escribe a medio magnético no extraíble y no volátil, una unidad 151 de disco magnético que lee de o escribe a un disco 152 magnético extraíble y no volátil y una unidad 155 de disco óptico que lee de o escribe a un disco 156 óptico extraíble y no volátil tal como un CD ROM u otro medio óptico. Otro medio de almacenamiento informático extraíble/no extraíble, volátil/no volátil que puede usarse en el entorno operativo ilustrativo incluye, pero sin limitación, casetes de cinta magnética, tarjetas de memoria flash, discos versátiles digitales, cinta de vídeo digital, RAM de estado sólido, ROM de estado sólido y similares. La unidad 141 de disco duro habitualmente se conecta al bus 121 de sistema a través de una interfaz de memoria no extraíble tal como la interfaz 140, y la unidad 151 de disco magnético y unidad 155 de disco óptico habitualmente se conectan al bus 121 de sistema mediante una interfaz de memoria extraíble, tal como la interfaz 150.

Las unidades y sus medios de almacenamiento informático asociados analizados anteriormente e ilustrados en la Figura1, proporcionan almacenamiento de instrucciones legibles por ordenador, estructuras de datos, módulos de programa y otros datos para el ordenador 110. En la Figura 1, por ejemplo, la unidad 141 de disco duro se ilustra como un sistema 144 operativo de almacenamiento, programas 145 de aplicación, otros módulos 146 de programa, y datos 147 de programa. Obsérvese que estos componentes pueden ser tanto el mismo como diferente del sistema 134 operativo, programas 135 de aplicación, otros módulos 136 de programa y datos 137 de programa. En este punto se dan diferentes números al sistema 144 operativo, programas 145 de aplicación, otros módulos 146 de programa y datos 147 de programa para ilustrar que, como mínimo, existen diferentes copias. Un usuario puede introducir órdenes e información en el ordenador 110 a través de dispositivos de entrada tales como un teclado 162 y dispositivo 161 apuntador, comúnmente denominado ratón, bola de mando o panel táctil. Otros dispositivos de entrada (no mostrados) pueden incluir un micrófono, palanca de mandos, control de juegos, disco de satélite, escáner o similares. Estos y otros dispositivos de entrada a menudo se conectan a la unidad 120 de procesamiento a través de una interfaz 160 de entrada de usuario que se acopla al bus 121 de sistema, pero puede conectarse mediante otras estructuras de interfaz o bus, tales como un puerto paralelo, puerto de juego o un bus de serie universal (USB). Un monitor 191 u otro tipo de dispositivo de presentación también se conecta al bus 121 de sistema a través de una interfaz, tal como una interfaz 190 de video. Además del monitor, los ordenadores también pueden incluir otros dispositivos de salida periféricos tales como altavoces 197 e impresora 196, que pueden conectarse a través de una interfaz 195 periférica de salida. También puede incluirse una cámara 192 (tales como cámaras de video o fijas digitales/electrónicas o escáner de película/fotográfico) capaz de capturar una secuencia de imágenes 193 como un dispositivo de entrada al ordenador 110 personal. Además, mientras solo se representa una cámara, múltiples cámaras podrían incluirse como dispositivos de entrada al ordenador 110 personal. Las imágenes 193 desde la una o más cámaras se introducen en el ordenador 110 a través de una interfaz 194 de cámara adecuada. Esta interfaz 194 se conecta al bus 121 de sistema, permitiendo de este modo que las imágenes se encaminen y almacenen en la RAM 132 o uno de los otros dispositivos de almacenamiento asociados con el ordenador 110. Sin embargo, se observa que también pueden introducirse datos de imagen en el ordenador 110 desde cualquiera de los anteriormente mencionados medios legible por ordenador, sin requerir el uso de la cámara 192.

El ordenador 110 puede operar en un entorno de interconexión usando conexiones lógicas a uno o más ordenadores remotos, tal como un ordenador 180 remoto. El ordenador 180 remoto puede ser un ordenador personal, un servidor, un encaminador, un PC de red, un dispositivo homólogo u otro nodo de red común y habitualmente incluye muchos o todos de los elementos descritos anteriormente relativos al ordenador 110, aunque solo un dispositivo 181 de almacenamiento de memoria se ha ilustrado en la Figura1. Las conexiones lógicas representadas en la Figura1 incluyen una red 171 de área local (LAN) y una red 173 de área extensa (WAN), pero también pueden incluir otras redes. Tales entornos de red son habituales en oficinas, redes informáticas de toda la empresa, intranets e Internet.

Cuando se usa en un entorno de interconexión LAN, el ordenador 110 se conecta a la LAN 171 a través de una interfaz de red o adaptador 170. Cuando se usa en un entorno de interconexión WAN, el ordenador 110 habitualmente incluye un módem 172 u otros medios para el establecimiento de comunicaciones en la WAN 173, tales como Internet. El módem 172, que puede ser interno o externo, puede conectarse al bus 121 de sistema a través del interfaz 160 de entrada de usuario u otro mecanismo apropiado. En un entorno de red, módulos de programa representados relativos al ordenador 110, o porciones del mismo, pueden almacenarse en el dispositivo

de almacenamiento de memoria remoto. A modo de ejemplo, y no como limitación, la Figura 1 ilustra programas 185 de aplicación residiendo en dispositivo 181 de memoria. Se apreciará que las conexiones de red mostradas son ilustrativas y pueden usarse otros medios de establecimiento de enlaces de comunicación entre los ordenadores.

El entorno operativo ilustrativo que se ha analizado ahora, las partes restantes de esta sección de descripción se dedicará a una descripción del hardware de sistema de video de punto de vista interactivo, arquitectura de módulo de programa y los propios módulos.

1.2 Sistema de captura de video de punto de vista interactivo

En general el sistema de captura de video de punto de vista interactivo incluye el hardware y software de soporte necesario para capturar las anteriormente mencionadas múltiples retransmisiones de video. El hardware de sistema de captura incluye un banco de cámaras de video, dispuestas en una manera de una al lado de otra. Además, las cámaras tienen una configuración convergente de tal forma que cada una apunta generalmente hacia el mismo objeto o área objetivo en una escena con cada campo de vista de la cámara superponiéndose al campo de vista de la cámara o cámaras adyacentes por una cantidad prescrita (por ejemplo, 30 grados). La orientación de las cámaras en relación al objeto o área objetivo puede variar dependiendo de los resultados deseados. En otras palabras, la distancia que cada cámara se aleja del objeto o área objetivo y su altura y/o colocación horizontal en relación al objetivo puede variar. Por ejemplo, una disposición útil es colocar las cámaras en un arco horizontal, como se muestra en la Figura 2. Esto resultaría en el usuario pudiendo ver la escena desde cualquier punto panorámico a lo largo del arco horizontal que conecta las cámaras. Por lo tanto, a los usuarios les parecería que pueden moverse horizontalmente en un arco alrededor del objetivo. Otra configuración que puede tomar el banco de cámaras es un arco vertical. Esto resultaría en el usuario pudiendo ver la escena desde puntos panorámicos que harían parecer como si el usuario se elevase por encima del objeto o área objetivo. Las cámaras tampoco necesitan estar todas alineadas en un plano horizontal o vertical. Por ejemplo, las cámaras podrían situarse diagonalmente a través de, es decir, en una configuración generalmente horizontal, pero con un recorrido hacia arriba desde un extremo al otro. Esto daría al espectador la impresión de que él o ella puede ver el objeto como si se movieran alrededor del mismo así como elevándose por encina del mismo al mismo tiempo. En general, cualquier trayectoria de vista puede crearse alineando las cámaras a lo largo de esa trayectoria deseada.

Las cámaras pueden ser cualquier cámara de video apropiada, sin embargo se prefiere una cámara de video digital. Si no se emplea una cámara digital, cada fotograma de video tendrá que digitalizarse antes de realizar el procesamiento adicional que se describirá es secciones posteriores. Además, mientras las cámaras de video no necesitan ser cámaras de alta resolución, la calidad del video de punto de vista interactivo resultante se mejora si se emplean tales cámaras. Aún más, para ayudar en la sincronización de las retransmisiones de video, es ventajoso si las cámaras son del tipo que pueden ser controladas remotamente a través de un ordenador para iniciar y parar y para ajustar los ajustes de cámara. Por lo tanto, un usuario puede controlar todas las cámaras simultáneamente a través del ordenador. Además, cámaras con capacidad de sincronizador serían deseables para facilitar la sincronización de sus transmisiones de video. También es ventajoso para procesamiento futuro si cada cámara añade metadatos a cada fotograma de video generado indicando los ajustes de cámara y exposición actuales, así como una marca de tiempo.

Haciendo referencia de nuevo a la Figura 2, se muestra una configuración ilustrativa del presente hardware de sistema de captura de video. En este ejemplo, ocho cámaras 200 se disponen a lo largo de un arco horizontal. Se usan cámaras a color de alta resolución (por ejemplo, 1024 × 768) para capturar video a 15 fps, con lentes de 8 mm, produciendo un campo de vista horizontal de aproximadamente 30 grados.

Otra característica clave del presente sistema de captura de video es la adquisición en tiempo real de retransmisiones de video sincronizadas desde las cámaras. Para realizar esta tarea es ventajoso equipo capaz de recibir y sincronizar las transmisiones individuales desde las cámaras, como es equipo para el almacenamiento de los datos de retransmisión de video sincronizada. En la configuración de ejemplo mostrada en la Figura2, la sincronización en tiempo real y almacenamiento de todos los videos de entrada se maneja mediante dos unidades 202 de concentrador y un banco de discos 204 duros. Cada concentrador 202 sincroniza la transmisión desde cuatro cámaras y canaliza las cuatro retransmisiones de video sin comprimir en el banco de discos 204 duros a través de un cable de fibra óptica. Los dos concentradores 202 se sincronizan a través de un cable FireWire para garantizar que todas las ocho transmisiones de video son síncronas. En una realización alternativa del sistema, cada cámara podría tener su propio dispositivo de grabación tales como cinta DV, cinta VHS, etc. El video entonces puede transferirse a disco duro después de la grabación.

El anteriormente mencionado ordenador 206 ejecuta un programa de captura de video diseñado para controlar las múltiples cámaras. En esencia, el programa de captura de video puede ser cualquier programa convencional que es capaz de encender y apagar simultáneamente múltiples cámaras de video, así como ajustar los ajustes de cámara (por ejemplo, exposición, balance de blancos, enfoque, entre otros) de cada una de las cámaras. En práctica, ajustes de cámara apropiados se determinarían usando procedimientos estándar antes de la sesión de captura y el programa de captura de video se usaría para ajustar todas las cámaras a estos ajustes. El programa de captura también inicia simultáneamente todas las cámaras en un momento preestablecido o por la entrada de una orden de usuario para hacerlo. Análogamente, el programa de captura simultáneamente para todas las cámaras en un

momento preestablecido o por la entrada de una orden de usuario.

Además de la captura y almacenamiento de retransmisiones de video, el sistema de captura de video de punto de vista interactivo también incluye programa de calibrado de cámara, que puede ejecutarse en el mismo ordenador usado para controlar las cámaras o un ordenador diferente. Las cámaras se calibran antes de cada sesión de captura para obtener todos los atributos de cámara necesarios para la reconstrucción en 3D. Incluyendo estos atributos tanto parámetros geométricos (por ejemplo, parámetros de cámara intrínsecos y extrínsecos) y parámetros fotométricos (por ejemplo, exposición, balance de blancos, degradado). En realizaciones ensayadas, los parámetros de cámara geométricos se obtuvieron usando la técnica de calibración de Zhang [24]. Este procedimiento generalmente implica el movimiento de un patrón de calibración que se ha montado en una superficie plana en frente de cada cámara. La retransmisión de video generada por cada cámara que representa el patrón de calibración se analiza a continuación para recuperar los atributos de cámara anteriormente mencionados. Los parámetros de cámara se almacenan y proporcionan, junto con las retransmisiones de video, al programa de generación de video de punto de vista interactivo que se describirá en breve.

1.3 Programas de sistema de video de punto de vista interactivo

El sistema de video de punto de vista interactivo también incluye programas informáticos tanto para la generación del video de punto de vista interactivo como para su renderización para reproducción a un usuario. La arquitectura y módulos de programa que componen cada uno de estos programas se describirán ahora.

1.3.1 Arquitectura de programa de generación de video de punto de vista interactivo

Haciendo referencia a la Figura 3, las retransmisiones 312 de video generadas mediante el anteriormente mencionado sistema de captura de video y datos 314 de calibración de cámara se transmiten primero en un módulo 300 de reconstrucción en 3D para procesamiento. El fin del módulo 300 de reconstrucción en 3D es generar correspondencias consistentes con fotos de alta calidad a través de todas las vistas de cámara y mapas de disparidades, para cada fotograma en cada grupo de fotogramas de video capturados contemporáneamente. Además, el módulo 300 de reconstrucción puede equilibrar los parámetros fotométricos de cada grupo de fotogramas una vez se conocen las correspondencias.

Cada mapa de disparidades generado es procesado por el módulo 302 de acoplamiento. En general, el módulo 302 de acoplamiento es responsable de la identificación de áreas de discontinuidades de profundidad significantes en un fotograma a base de su mapa de disparidades. Esta información se proporciona a continuación al módulo 304 de representación en capas, que en una realización del presente sistema, genera una capa principal compuesta de píxeles asociada con áreas en un fotograma que no muestran discontinuidades significantes de profundidad e información de fondo en áreas alrededor de discontinuidades de profundidad y una capa límite compuesta de información de primer plano de píxeles asociados con áreas que tienen discontinuidades de profundidad significantes. Por lo tanto, se crea una representación de dos capas para cada fotograma de las retransmisiones de video de cada una de las cámaras.

Las representaciones de fotograma de video de dos capas se proporcionan opcionalmente a continuación a un módulo 306 de compresión. Mientras la compresión de los datos es opcional, se observa que las dos capas para cada fotograma generado mediante cada cámara en el anteriormente mencionado banco de cámaras representarán una significante cantidad de datos (por ejemplo, del orden de 800 MB sin compresión para 8 cámaras a 15 fps grabando durante 1 segundo). Por lo tanto, cualquier compresión de estos datos ayudará en su transmisión y/o almacenamiento. La naturaleza opcional de este módulo se indica en la Figura 3 mediante el uso de una caja de línea discontinua.

Las representaciones de fotograma de video de dos capas, ya sea comprimida o no, se pasan a continuación a un módulo 308 de generación de archivo. Adicionalmente, los anteriormente obtenidos datos 314 de calibración de cámara se proporcionan al módulo 308 de generación de archivo. En esencia, el módulo 308 de generación de archivo codifica las representaciones de fotograma de video de dos capas y datos de calibración para transmisión en directo al anteriormente mencionado programa de renderización de video de punto de vista interactivo para procesamiento o para almacenamiento para transmisión futura al programa de renderización. En una realización de la presente invención los datos 314 de calibración de cámara se sitúan en la cabecera del archivo de video.

Se observa que mientras el modo de operación preferido del anterior programa de generación de video de punto de vista interactivo es emplear representaciones basadas en imágenes de una escena capturada, todavía es posible implementar el presente sistema y procedimiento usando imágenes en 3D generadas por ordenador en su lugar. En esta realización alternativa, se elimina el módulo de reconstrucción en 3D y en su lugar se introducen fotogramas de video generados por ordenador en el módulo 302 de acoplamiento a través de un módulo 310 de imágenes en 3D generadas por ordenador. Por lo tanto, el anteriormente descrito sistema de captura tampoco se necesita. En este punto de nuevo la naturaleza opcional del módulo 310 de imágenes se indica en la Figura 3 mediante el uso de una caja de líneas discontinuas.

La entrada de fotogramas sintéticos en lugar de los fotogramas capturados con cámara todavía mostraría todos los mismos atributos descritos anteriormente en conexión con la descripción de las retransmisiones de video reales.

Además, se introduciría información de parámetros de cámara virtual en el módulo de generación de archivo para cada retransmisión de video sintetizada en lugar de los datos de calibración de cámara reales. El fotograma sintetizado y datos de parámetros de cámara se procesarían a continuación de la misma manera que los datos basados en imágenes. Como tal, para los fines de la restante descripción de la invención, no se hará distinción entre si los datos de fotograma proporcionados al módulo de representación en capas se basan en imágenes o se sintetiza. Análogamente, no se hará distinción entre si las cámaras son reales o virtuales y si los parámetros de cámara se calcularon o sintetizaron.

1.3.1.1 Módulo de reconstrucción en 3D

Cuando se desarrolla un procedimiento de visión en estéreo para su uso en interpolación de vista, los requisitos de precisión varían de los algoritmos de estéreo estándar usando para la reconstrucción en 3D. Específicamente, errores en disparidad no son tan preocupantes como errores en valores de intensidad para la imagen interpolada. Por ejemplo, un error de disparidad multipíxel en un área de baja textura, tal como una pared blanca, resultará en error de intensidad significantemente menor que el mismo error de disparidad en un área de alta textura. En particular, bordes o líneas rectas en la escena necesitan renderizarse correctamente.

Los algoritmos de estéreo tienden a producir resultados erróneos alrededor de discontinuidades de disparidad. Desafortunadamente, tales errores producen algunas de las impurezas más perceptibles en escenas interpoladas, ya que las discontinuidades de disparidad habitualmente coinciden con bordes de intensidad. Por esta razón, el algoritmo de estéreo para la interpolación de vista debe coincidir correctamente con los píxeles alrededor de los bordes de intensidad que incluyen discontinuidades de disparidad.

Recientemente, se ha propuesto un nuevo enfoque para visión en estéreo llamado estéreo basado en segmentación. Estos procedimientos segmentan la imagen en regiones con probabilidad de tener disparidades similares o suaves antes del cálculo de estéreo. A continuación se aplica una limitación de suavidad a cada segmento. Tao y col. [17] usaron una limitación plana, mientras Zhang y Kambhamettu [22] usaron los segmentos para soporte local. Estos procedimientos han mostrado resultados muy prometedores en el manejo preciso de discontinuidades de disparidad.

Mientras las anteriores técnicas de estéreo basadas en segmentación pueden emplearse para conseguir la tarea de reconstrucción en 3D, las realizaciones de la presente invención ensayadas emplearon un nuevo enfoque basado en segmentación. Este nuevo enfoque es el tema de una solicitud en tramitación con la presente titulada "Color Segmentation-Based Stereo Reconstruction System And Process" por los inventores de esta solicitud y transferida al cesionario común. La solicitud en tramitación con la presente se presentó el 28 de junio de 2004 y lleva el número de publicación US 2005/0286757.

1.3.1.2 Módulo de acoplamiento

Durante el cálculo de estéreo, se asume que cada píxel tiene una disparidad única. En general este no es el caso, ya que algunos píxeles a lo largo del límite de los objetos recibirán contribuciones tanto desde regiones de primer plano como de fondo. Sin embargo, si los colores de píxel mezclados originales se usan durante la renderización basada en imágenes, resultarán impurezas visibles.

Para resolver este problema, se definen pequeñas áreas en la proximidad de las discontinuidades de profundidad, que se definen como cualquier salto de disparidad mayor de λ píxeles (por ejemplo, 4 píxeles). Más particularmente, se usa acoplamiento para encontrar información de primer plano y fondo para cada píxel dentro de estas áreas. La información de primer plano se almacena dentro de la capa de límite, mientras la información de fondo y la información de píxeles fuera de los λ píxeles de una discontinuidad de profundidad se almacenan en la capa principal. Para evitar la aparición de fracturas durante el procedimiento de renderización que se describirá más tarde, el acoplamiento de límite se dilata (por ejemplo, por un píxel hacia el interior de la región de píxel de primer plano). Esta información de etiqueta de píxel se pasa a continuación en el módulo de representación en capas.

Mientras la anterior tarea de acoplamiento puede conseguirse usando cualquier técnica de acoplamiento convencional apropiada, las realizaciones de la presente invención ensayadas emplearon un nuevo enfoque. Este nuevo enfoque es el tema de una solicitud en tramitación con la presente titulada "A System And Process For Generating A Two-Layer, 3D Representation Of A Scene" por los inventores de esta solicitud y transferida al cesionario común. La solicitud en tramitación con la presente se presentó el 28 de junio de 2004 y lleva el número de publicación US 2005/0286757.

1.3.1.3 Módulo de representación en capas

El módulo de representación en capas toma los datos asociados con cada fotograma, así como la información de etiqueta de píxel generada mediante el módulo de acoplamiento y estima los colores, profundidades y opacidades (es decir, valores alfa) para el acoplamiento de límite. Esto puede conseguirse usando una técnica de acoplamiento, tal como, por ejemplo, acoplamiento de imagen bayesiana [5]. Obsérvese que [5] no estima profundidades, únicamente colores y opacidades. Las profundidades pueden estimarse usando medias ponderadas por alfa de profundidades cercanas en las regiones de píxel de primer plano y fondo. Los datos de píxel de primer plano

resultantes se designan como la capa límite para el fotograma que se está procesando. A continuación, los datos de píxel de fondo junto con los datos de píxel de los restantes píxeles fuera del acoplamiento límite se usan para componer la capa principal del fotograma.

Por consiguiente, la salida del módulo de representación en capas es una capa límite para cada fotograma de la retransmisión de video de cada una de las anteriormente mencionadas cámaras, que identifica para cada píxel en la capa, el color B_C del píxel, profundidad B_D y opacidad **α**. Además, una capa principal se extrae para cada fotograma, que identifica para cada píxel en esa capa, el color M_C del píxel, profundidad M_D

Mientras la tarea anterior de deposición de capas puede conseguirse usando cualquier técnica apropiada convencional de deposición de capas, las realizaciones de la presente invención ensayadas emplearon un nuevo enfoque. Este nuevo enfoque es el tema de la anteriormente mencionada solicitud en tramitación con la presente titulada "A System And Process For Generating A Two-Layer, 3D Representation Of A Scene", que se presentó el 28 de junio de 2004 y lleva el número de publicación US 2005/0286757.

1.3.1.4 Módulo de compresión

La compresión puede usarse opcionalmente para reducir los grandes conjuntos de datos asociados con la presente invención a un tamaño manejable y para soportar reproducción más rápida. En la presente invención puede adoptarse con ventaja cualquier esquema de compresión basado en tiempo, por ejemplo MPEG-4 (ISO/IEC 14496). Sin embargo, porque cada cámara está capturando parte de la misma escena, existe una oportunidad para comprimir los datos aprovechando las redundancias entre cámaras (es decir, espacial). La predicción temporal usa estimaciones compensadas de movimiento del fotograma anterior, mientras que la predicción espacial usa una textura de cámara de referencia y mapas de disparidades transformados en el punto de vista de una cámara adyacente espacialmente como una base para la compresión. Por ejemplo MPEG-4 contiene un estándar para la compresión de datos estereoscópicos que aprovecha la similitud de puntos de vista adyacentes. Por lo tanto, en general un códec que aprovecha la redundancia temporal o espacial es apropiado para esta tarea. Sin embargo, la máxima compresión puede lograrse combinando los dos procedimientos de compresión.

Mientras el anterior esquema de compresión combinado puede implementarse usando técnicas de compresión temporales y espaciales existentes, las realizaciones de la presente invención ensayadas emplearon un nuevo enfoque integrado. Este nuevo enfoque es el tema de una solicitud en tramitación con la presente titulada "A System And Process For Compressing And Decompressing Multiple, Layered, Video Streams Employing Spatial And Temporal Encoding" por los inventores de esta solicitud y transferida al cesionario común. La solicitud en tramitación con la presente se presentó el 3 de agosto de 2004 y lleva el número de publicación US 2006/0028473.

1.3.2 Procedimiento de generación de video de punto de vista interactivo

La anterior arquitectura de programa puede emplearse para realizar el siguiente procedimiento para generar un video de punto de vista interactivo en una realización de la presente invención como se muestra en las Figuras 4A-B. Primero, las retransmisiones de video sincronizadas se introducen desde el anteriormente descrito sistema de captura de video (acción 400 de procedimiento). A continuación se realiza una reconstrucción en 3D para calcular las correspondencias a través de todas las vistas de cámara y mapas de disparidades para cada fotograma en cada grupo de fotogramas de video capturados contemporáneamente desde las retransmisiones de video introducidas (acción 402 de procedimiento). Además, los parámetros fotométricos de cada grupo de fotogramas se equilibran una vez que se conocen las correspondencias (acción 404 de procedimiento).

A continuación, cada fotograma se selecciona en un orden prescrito (acción 406 de procedimiento). Más específicamente, esto implica la selección de cada fotograma en cada grupo de fotogramas de video entrantes capturados contemporáneamente (en cualquier orden deseado) y a continuación hacer lo mismo para la entrada del siguiente grupo de fotogramas y así sucesivamente. Para cada fotograma seleccionado, se definen áreas de discontinuidades de profundidad significantes a base de su mapa de disparidades (acción 408 de procedimiento). Esta información se usa para generar una capa límite compuesta de información de primer plano de píxeles asociados con áreas que tienen significantes discontinuidades de profundidad y una capa principal que consiste en la restante información (acción 410 de procedimiento). A continuación se determina si queda algún fotograma previamente no seleccionado por procesar (acción 412 de procedimiento). Si es así, se repiten las acciones 406 hasta 412 de procedimiento hasta que todos los fotogramas se hayan procesado. Por lo tanto, por último se crea una representación de dos capas para cada fotograma. Si no quedan fotogramas para seleccionar, entonces el procedimiento de generación continua con la compresión opcionalmente de los datos de fotograma de video (acción 414 de procedimiento). Esto puede hacerse usando, por ejemplo, tanto técnicas de compresión temporal (es decir, entre grupos de fotogramas capturados contemporáneamente) como espaciales (es decir, entre los fotogramas en el mismo grupo). Se observa que la naturaleza opcional de esta última acción se indica en la Figura 4 usando una caja de líneas discontinuas. Se compriman o no los datos de fotograma, la siguiente acción 416 de procedimiento es generar un archivo de video de punto de vista interactivo que contiene las representaciones de fotograma de video en capas y los datos de calibración de cámara proporcionados desde el sistema de captura de video.

Se observa que la entrada de retransmisiones de video basadas en imágenes desde el sistema de captura de video puede reemplazarse en el anterior procedimiento de generación de video de punto de vista interactivo con datos de video generado por ordenador como se describió anteriormente. En un caso de este tipo, los datos de calibración de cámara también se reemplazarían con datos virtuales del mismo tipo.

1.3.3 Arquitectura de programa de renderización de video de punto de vista interactivo

Haciendo referencia a la Figura 5, un archivo 504 de video de punto de vista interactivo generado mediante el anteriormente mencionado módulo de generación de archivo se transmite primero en un módulo 500 de decodificación selectivo. En general, el módulo 500 de decodificación selectivo decodifica solo esas porciones del archivo entrante que se necesitan para renderizar un fotograma del video actual desde un punto panorámico seleccionado por un usuario que ve el video. Más particularmente, el módulo 500 decodifica las porciones del archivo identificado por el módulo 502 de renderización (que se describirá en breve) para recuperar los datos de fotograma de video en capas asociados con los fotogramas de video particulares que se necesitan para renderizar la escena desde el punto de vista deseado. De esta manera, debe codificarse la mínima cantidad de datos posible, de este modo acelerando el procedimiento y proporcionando capacidad de renderización en tiempo real.

Los datos de fotograma en capas decodificados se proporcionan al módulo 502 de renderización. En general, este módulo 502 toma los datos de fotograma y renderiza una vista de la escena para la porción actual del video que el usuario está viendo desde un punto panorámico especificado por el usuario. Esto implica primero la obtención de la entrada 506 actual de usuario y a continuación la generación de la vista deseada.

1.3.3.1 Módulo de decodificación selectivo

El fin del módulo de decodificación selectivo es decodificar únicamente la información que se necesita para renderizar la escena capturada en el video desde el punto panorámico actual seleccionado por el usuario. En esencia esto implica la decodificación del fotograma o fotogramas desde el grupo de fotogramas capturados contemporáneamente asociados con la ubicación temporal actual en el video que se está renderizando que debe ser decodificado para obtener los datos de fotograma de video en capas que se necesitan para renderizar una vista de la escena representada en la porción actual del video desde un punto panorámico seleccionado por un usuario particular. Este punto panorámico debería coincidir con la vista de la escena capturada por una de las cámaras, a continuación únicamente necesitan decodificarse los datos asociados con ese fotograma. Sin embargo, si el punto de vista deseado se sitúa en algún lugar entre dos de las vistas de cámara, entonces los datos de fotograma asociados con ambas de estas cámaras adyacentes deben decodificarse para renderizar la escena desde el punto de vista deseado.

El fotograma o fotogramas particulares que se necesitan para renderizar la escena desde el punto panorámico deseado se identifica mediante el módulo de renderización (que se describirá a continuación). Una vez identificado, los datos de fotograma de video en capas asociado con el fotograma o fotogramas identificados se decodifica usando la técnica de decodificación apropiada aplicable al tipo de compresión y esquemas de codificación empleados en los anteriormente descritos módulos de compresión y generación de archivo. En casos en los que se emplea el anteriormente mencionado enfoque integrado como en las realizaciones de la presente invención ensayadas, la decodificación se consigue como se describe en la solicitud en tramitación con la presente titulada "A System And Process For Compressing And Decompressing Multiple, Layered, Video Streams Employing Spatial And Temporal Encoding", que se presentó el 3 de agosto de 2004 y lleva el número de publicación US 2006/0028473.

Además de la decodificación de datos de fotograma desde el archivo de video de punto de vista interactivo, el módulo de decodificación también decodifica los anteriormente mencionados datos de calibración de cámara. Como se indicó anteriormente, estos datos podrían encontrarse en la cabecera del archivo o como metadatos.

1.3.3.2 Módulo de renderización

El trabajo del módulo de renderización es primero procesar la entrada de usuario referida al punto de vista que se desea para la escena a renderizar e identificar el fotograma o fotogramas desde el grupo de fotogramas capturados contemporáneamente asociados con la porción temporal actual del video que se está renderizando que se necesitan para renderizar la vista deseada. Para conseguir esta tarea, el módulo de renderización se inicializa con los anteriormente mencionados datos de calibración de cámara contenidos en el archivo de video de punto de vista interactivo. Estos datos de calibración incluyen la ubicación e información de punto de vista para cada una de las cámaras de video usadas para capturar la escena asociada con el video que se está viendo. Dada esta información el módulo de renderización calcula las ubicaciones de los puntos de vista de cámara. Como se describe anteriormente, el usuario puede especificar cualquier punto de vista a lo largo de la trayectoria que conecta los puntos de vista de cámara, con las dos cámaras exteriores representando los puntos finales de las selecciones posibles de puntos de vista. Como se analizó anteriormente, el punto de vista seleccionado puede coincidir con la vista de la escena capturada por una de las cámaras (o como alternativa la vista sintetizada desde una posición de cámara virtual). En un caso de este tipo únicamente el fotograma "actual" asociado con esa cámara se identifica como el que se necesita para renderizar la vista deseada. Sin embargo, el caso habitual será que el punto de vista se sitúe entre los puntos de vista de dos cámaras adyacentes. En este último caso, el módulo de renderización

identifica fotogramas actuales asociados con ambas de estas cámaras adyacentes.

Como para la entrada de usuario, esta información puede obtenerse de cualquier manera convencional apropiada, tal como a través de una interfaz de usuario de algún tipo usada para introducir y procesar las selecciones de punto de vista de usuario. Por ejemplo, esta interfaz puede incluir una interfaz gráfica de usuario (GUI) que se presenta al usuario en un dispositivo de presentación (por ejemplo, monitor de ordenador, pantalla de presentación, gafas 3D, entre otras). Esta GUI incluiría alguna disposición gráfica que permita al usuario indicar el punto de vista, entre los posibles puntos de vista, desde el que él o ella quiere ver la escena capturada en el video para la porción actual del video que se está renderizando. El usuario también puede cambiar el punto de vista deseado mientras se reproduce el video. Estas selecciones podrían hacerse por el usuario interactuando con la GUI usando cualquier dispositivo de entrada estándar (por ejemplo, ratón, palanca de mandos, dispositivo de seguimiento ocular, entre otros).

Una vez que el fotograma o fotogramas que se necesitan para renderizar la vista deseada se han identificado, el módulo de renderización dirige el módulo de decodificación selectivo para decodificar los datos de fotograma necesarios. La salida de datos de fotograma del módulo de decodificación selectivo que consiste en 5 planos de datos para cada fotograma proporcionado: el color de capa principal, profundidad de capa principal, acoplamiento alfa de capa límite, color de capa límite y profundidad de capa límite. En el caso en el que el punto de vista deseado coincida con uno de los puntos de vista de cámara, únicamente se usan los planos de datos de capa principal y capa límite de esa cámara para reconstruir la escena. Sin embargo, en el caso en el que el punto de vista deseado se sitúe entre dos de los puntos de vista de cámara, el procedimiento de renderización es más complicado. En una realización del presente módulo de renderización en la que se requieren datos de dos puntos de vista de cámara para renderizar una vista de la escena desde el punto de vista especificado por el usuario, los datos de capa principal y límite de cada cámara se proyectan al punto de vista deseado. Esto puede conseguirse usando procedimientos de renderización convencionales y los datos de calibración de cámara proporcionados en el archivo de video de punto de vista interactivo. Las capas principal y límite proyectas se combinan a continuación para generar el fotograma final. En este punto de nuevo pueden emplearse procedimientos de combinación convencionales con cada contribución de capa a la vista final que se está viendo ponderada en proporción a cómo de cerca está el punto de vista de la cámara asociada del punto de vista deseado. En otras palabras, si el punto de vista deseado está más cerca de uno de los puntos de vista de la cámara que el otro, la capa proyectada asociada con la primera de estas cámaras se pondera con más peso que la otra.

Mientras pueden emplearse técnicas convencionales de proyección de vistas y renderización para conseguir la anterior tarea de renderización, las realizaciones de la presente invención ensayadas emplearon un nuevo enfoque. Este nuevo enfoque es el tema de una solicitud en tramitación con la presente titulada "An Interactive, Real-Time Rendering System And Process For Virtual Punto de vista Video" por los inventores de esta solicitud y transferida al cesionario común. La solicitud en tramitación con la presente se presentó el 3 de agosto de 2004 y lleva el número de publicación US2006/0028473. Se observa adicionalmente que el anterior procedimiento de renderización puede conseguirse usando unidades de procesamiento gráficas, técnicas de renderización de software o ambas. La Figura 6(a)-(c) muestra un ejemplo de los resultados del anterior procedimiento de renderización. Las Figuras 6(a) y (c) representan fotogramas capturados contemporáneamente desde dos cámaras de video adyacentes ubicadas en diferentes puntos de vista. La Figura 6(b) es un ejemplo del fotograma que se renderiza cuando el punto de vista especificado por el usuario está entre los puntos de vista asociados con las imágenes de las Figuras 6(a) y (c).

Como se indica anteriormente, el modo de operación preferido del anterior programa de generación de video de punto de vista interactivo es emplear representaciones basadas en imágenes de una escena capturada. Sin embargo, no está fuera del ámbito e la presente invención introducir además elementos sintetizados en las escenas renderizadas. Por lo tanto, en una realización del módulo de renderización (como se muestra en la Figura 5), datos 508 de objeto en 3D se introducen en el módulo de renderización para incorporación en el fotograma que se está renderizando en la actualidad. En una realización, está entrada incluiría los datos necesarios para renderizar un objeto u objetos animados desde un punto de vista que corresponde al punto de vista seleccionado e información de ubicación para la incorporación del objeto(s) en una preestablecida posición dentro del fotograma que se está renderizando. El objeto(s) puede cambiar de forma con el paso del tiempo (es decir, como para tener un aspecto diferente en fotogramas diferentes de los fotogramas renderizados) o tener un aspecto estático. Además, la posición en el fotograma renderizado en el que el objeto(s) se incorpora puede cambiar con el tiempo (es decir, como para tener una ubicación diferente en fotogramas diferentes de los fotogramas renderizados) o puede ubicarse en el mismo lugar en cada fotograma sucesivo renderizado.

También pueden insertarse objetos basados en imágenes en la escena durante el procedimiento de renderización. Por ejemplo, la Figura 7 muestra un fotograma de un video de punto de vista interactivo creado de acuerdo con la presente invención, en el que se ha insertado una copia extra de un bailarín de *breakdance*. Este efecto se logró "extrayendo" primero un acoplamiento del bailarín usando un umbral de profundidad y a continuación insertando el conjunto de imágenes agrupadas en una sola imagen extraído en el video original usando *z-buffering*.

1.3.4 Procedimiento de renderización de video de punto de vista interactivo

La anterior arquitectura de programa de renderización puede emplearse para realizar el siguiente procedimiento para renderizar el video de punto de vista interactivo en una realización de la presente invención como se muestra

en la Figura 8. En general, para cada fotograma del video renderizado, se introduce primero el punto vista actual especificado por el usuario (acción 800 de procedimiento). Sin embargo, se observa que en lugar de introducir el punto de vista cada vez que se renderiza un nuevo fotograma del video, únicamente podrían introducirse cambios en el punto de vista especificado. En este caso, salvo que se haya recibido un cambio en punto de vista, se asumirá que el último punto de vista especificado es todavía válido y se usará en la renderización del fotograma actual del video.

Una vez que se establece el punto de vista especificado por el usuario, la siguiente acción 802 de procedimiento es la identificación del fotograma o fotogramas desde el grupo de fotogramas de entrada capturados contemporáneamente asociados con el fotograma actual del video que se está renderizando, que se necesitan para generar la vista deseada. El fotograma o fotogramas identificados con a continuación decodificados (acción de 804 procedimiento).

A continuación, el fotograma actual del video de punto de vista interactivo se renderiza usando los datos de video decodificado (acción 806 de procedimiento). Este fotograma representará la escena asociada con la porción temporal actual del video como se ve desde el punto de vista especificado en la actualidad por el usuario. Esto puede requerir la sintetización del fotograma si el punto de vista deseado se sitúa entre los puntos de vista de dos de las cámaras adyacentes usadas para capturar la escena. Se observa que el anterior procedimiento puede opcionalmente modificarse para también insertar objetos generados por ordenador o basados en imágenes en la escena durante el procedimiento de renderización como se describe anteriormente, aunque esta acción no se muestra en la Figura 8.

2.0 Aplicaciones potenciales

El sistema de video de punto de vista interactivo y procedimiento puede emplearse en una diversidad de interesantes aplicaciones. En su nivel básico un usuario puede reproducir un video y cambiar continuamente su punto de vista mientras ve. Por lo tanto, el video de punto de vista interactivo permite a los usuarios experimentar el video como un medio en 3D interactivo. Esto tiene un gran potencial para cambiar el modo en que se ven eventos dinámicos e intensifica el realismo de juegos. Ejemplos de eventos dinámicos de interés son eventos deportivos (béisbol, baloncesto, patinaje en monopatín, tenis, etc.), videos instructivos (videos explicativos de golf, artes marciales, etc.) y actuaciones (Circo del Sol, ballet, danza moderna, etc.). Además, si hay disponible suficiente ancho de banda, el video podría difundirse o multidifundirse, por lo tanto proporcionando una experiencia de visionado que podría describirse como televisión en 3D.

Sin embargo, la presente invención no se limita solo a cambiar puntos de vista mientras se ve el video. También puede usarse para producir una diversidad de efectos especiales tal como manipulación del espacio-tiempo. Por ejemplo un usuario puede congelar el video y ver la escena representada desde una diversidad de puntos de vista. Un usuario también puede reproducir el video mientras ve la escena representada desde uno o más puntos de vista y a continuación retroceder el video y ver la escena desde diferentes puntos de vista. Aún más el video puede reproducirse hacia delante o hacia atrás a cualquier velocidad, mientras se cambian los puntos de vista como se desee.

Las anteriores características de punto de vista interactiva no únicamente son interesantes para un espectador casual, sino que sería particularmente útil para la industria televisiva y cinematográfica. En lugar del procedimiento meticuloso de determinar qué parte de una escena se debe capturar y desde qué punto de vista con antelación, con la posibilidad de que se pierda el plano más deseable, puede usarse el sistema y procedimiento de la presente invención. Por ejemplo, una escena se capturaría primero como un video de punto de vista interactivo. A continuación, el cineasta puede ver el video y elegir por cada plano (incluso hasta en una base de fotograma por fotograma) el punto de vista que se desea para la película final. Además, la anteriormente descrita característica de inserción de objeto también es una herramienta que podría ser ventajosa para el cineasta. Por lo tanto, las técnicas presentadas nos traen una etapa más cerca para hacer de la renderización basada en imagen (y basada en video) un componente integral de la autoría y distribución de los medios futuros.

3.0 Referencias

[1] Buehler, C., Bosse, M., McMillan, L., Gortler, S. J., y Cohen, M. F. 2001. Unstructured lumigraph rendering. Proceedings of SIGGRAPH 2001 (Agosto), 425-432.

[2] Carceroni, R. L., and Kutulakos, K. N. 2001. Multiview scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance. En Eighth International Conference on Computer Vision (ICCV 2001), vol. II, 60-67.

[3] Carranza, J., Theobalt, C., Magnor, M. A., y Seidel, H.-P. 2003. Free-viewpoint video of human actors. ACM Transactions on Graphics 22, 3 (Julio), 569-577.

[4] Chang, C.-L., et al. 2003. Inter-view wavelet compression of light fields with disparity-compensated lifting. En Visual Communication and Image Processing (VCIP 2003).

[5] Chuang, Y.-Y., et al. 2001. A Bayesian approach to digital matting. En Conference on Computer Vision and Pattern Recognition (CVPR'2001), vol. II, 264-271.

[6] Debevec, P. E., Taylor, C. J., y Malik, J. 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. Computer Graphics (SIGGRAPH'96) (Agosto), 11-20.

[7] Debevec, P. E., Yu, Y., y Borshukov, G. D. 1998. Efficient view-dependent image-based rendering with projective texture-mapping. Eurographics Rendering Workshop 1998, 105-116.

[8] Gortler, S. J., Grzeszczuk, R., Szeliski, R., y Cohen, M. F. 1996. The Lumigraph. In Computer Graphics (SIGGRAPH'96) Pro-ceedings, ACM SIGGRAPH, 43-54.

[9] Hall-Holt, O., y Rusinkiewicz, S. 2001. Stripe boundary codes for real-time structured-light range scanning of moving objects. En Eighth International Conference on Computer Vision (ICCV 2001), vol. II, 359-366.

[10] Heigl, B., et al. 1999. Plenoptic modeling and rendering from image sequences taken by handheld camera. En DAGM'99, 94-101.

[11] Kanade, T., Rander, P. W., y Narayanan, P. J. 1997. Virtualized reality: constructing virtual worlds from real scenes. IEEE Mul-tiMedia Magazine 1, 1 (Enero-Marzo), 34-47.

[12] Levoy, M., y Hanrahan, P. 1996. Light field rendering. En Computer Graphics (SIGGRAPH'96) Proceedings, ACM SIG-GRAPH, 31-42.

[13] Pulli, K., et al. 1997. View-based rendering: Visualizing real objects from scanned range and color data. En Proceedings of the 8-th Eurographics Workshop on Rendering.

[14] Scharstein, D., y Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47, 1 (Mayo), 7-42.

[15] Seitz, S. M., y Dyer, C. M. 1997. Photorealistic scene reconstruction by voxel coloring. En Conference on Computer Vision and Pattern Recognition (CVPR'97), 1067-1073.

[16] Shade, J., Gortler, S., He, L.-W., y Szeliski, R. 1998. Layered depth images. En Computer Graphics (SIGGRAPH'98) Proceedings, ACM SIGGRAPH, Orlando, 231-242.

[17] Tao, H., Sawhney, H., y Kumar, R. 2001. A global matching framework for stereo computation. En Eighth International Conference on Computer Vision (ICCV 2001), vol. I, 532-539.

[18] Vedula, S., Baker, S., Seitz, S., y Kanade, T. 2000. Shape and motion carving in 6D. En Conference on Computer Vision and Pattern Recognition (CVPR'2000), vol. II, 592-598.

[19] Wexler, Y., Fitzgibbon, A., y Zisserman, A. 2002. Bayesian estimation of layers from multiple images. En Seventh European Conference on Computer Vision (ECCV 2002), vol. III, 487-501.

[20] Wilburn, B., Smulski, M., Lee, H. H. K., y Horowitz, M. 2002. The light field video camera. En SPIE Electonic Imaging: Media Processors, vol. 4674, 29-36.

[21] Yang, J. C., Everett, M., Buehler, C., y McMillan, L. 2002. A real-time distributed light field camera. En Eurographics Workshop on Rendering, P. Debevec and S. Gibson, Eds., 77-85.

[22] Zhang, Y., y Kambhamettu, C. 2001. On 3D scene flow and structure estimation. En Conference on Computer Vision and Pattern Recognition (CVPR'2001), vol. II, 778-785.

[23] Zhang, L., Curless, B., y Seitz, S. M. 2003. Spacetime stereo: Shape recovery for dynamic scenes. En Conference on Computer Vision and Pattern Recognition, 367-374.

[24] Zhang, Z. 2000. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 11, 1330-1334.

REIVINDICACIONES

1. Un procedimiento implementado informáticamente de generación de un video de punto de vista interactivo, que comprende el uso de un ordenador para realizar las siguientes acciones de procedimiento:

introducir una pluralidad de retransmisiones de video sincronizadas cada una representando una porción de la misma escena y datos de calibración definiendo parámetros geométricos y fotométricos asociados con cada retransmisión de video; y
para cada grupo de fotogramas contemporáneos de las retransmisiones de video sincronizadas, generar una reconstrucción en 3D de la escena,
usar la reconstrucción para calcular un mapa de disparidades para cada fotograma en el grupo de fotogramas contemporáneos, y
para cada fotograma en el grupo de fotogramas contemporáneos,
identificar áreas de discontinuidades de profundidad significantes a base de su mapa de disparidades,
generar una capa principal que comprende información de píxeles asociada con áreas en un fotograma que no muestra discontinuidades de profundidad que exceden un umbral prescrito e información de píxeles de fondo de áreas que tienen discontinuidades de profundidad por encima del umbral, y una capa límite que comprende información de píxeles de primer plano asociada con áreas que tienen discontinuidades de profundidad que exceden el umbral, para producir una representación en capas para el fotograma en consideración.

2. El procedimiento de la reivindicación 1, que comprende además las acciones de procedimiento de:

usar la reconstrucción para calcular correspondencias a través de todas las vistas de cámara; y equilibrar los parámetros fotométricos de cada grupo de fotogramas una vez se calculan las correspondencias.

3. El procedimiento de la reivindicación 1, que comprende además una acción de procedimiento de comprimir las representaciones en capas generadas para los fotogramas del video de punto de vista interactivo para facilitar transferencia y/o almacenamiento del video.

4. El procedimiento de la reivindicación 3, en el que la acción de procedimiento de comprimir las representaciones en capas comprende el uso de técnicas de compresión temporales entre los grupos contemporáneos de fotogramas de video de punto de vista interactivo.

5. El procedimiento de la reivindicación 4, en el que la acción de procedimiento de comprimir las representaciones en capas comprende el uso de técnicas de compresión espaciales entre los fotogramas de video de punto de vista interactivo en el mismo grupo contemporáneo de fotogramas.

6. El procedimiento de la reivindicación 3, en el que la acción de procedimiento de comprimir las representaciones en capas comprende el uso de técnicas de compresión espaciales entre los fotogramas de video de punto de vista interactivo en el mismo grupo contemporáneo de fotogramas.

7. El procedimiento de la reivindicación 1, que comprende además una acción de procedimiento de generación de un archivo de video de punto de vista interactivo que comprende las representaciones en capas generadas de fotogramas de las retransmisiones de video introducidas y dichos datos de calibración.

8. El procedimiento de la reivindicación 1, en el que la pluralidad de retransmisiones de video sincronizadas y datos de calibración se deriva de fotogramas de imágenes capturados mediante una pluralidad de cámaras de video.

9. El procedimiento de la reivindicación 1, en el que la pluralidad de retransmisiones de video sincronizadas y datos de calibración se genera por ordenador.

10. El procedimiento de la reivindicación 1, en el que la acción de procedimiento de generación de una reconstrucción en 3D de la escena comprende una acción de empleo de una técnica de reconstrucción basada en segmentación.

11. El procedimiento de la reivindicación 1, en el que la acción de procedimiento de generación de la capa principal comprende una acción de establecimiento del color y profundidad de cada píxel en la capa y en el que la acción de procedimiento de generación de la capa límite comprende una acción de establecimiento del color, profundidad y opacidad de cada píxel en la capa.

12. El procedimiento de la reivindicación 1, en el que la acción de procedimiento de generación de la capa límite comprende una acción de dilatación de la capa para incluir un número prescrito de píxeles adyacentes a los píxeles que muestran discontinuidades de profundidad que exceden el umbral.

13. Un procedimiento implementado informáticamente para la renderización de un video de punto de vista interactivo de datos que comprende representaciones en capas de fotogramas de video generadas de grupos secuenciales de fotogramas de video capturados contemporáneamente representando cada uno una porción de la misma escena, y que comprende datos de calibración que comprenden parámetros geométricos asociados con la captura de cada fotograma de video, siendo dicho video de punto de vista interactivo generado de acuerdo con el procedimiento de la

reivindicación 1 y dicho procedimiento que comprende el uso de un ordenador para realizar las siguientes acciones de procedimiento para cada fotograma del video de punto de vista interactivo a renderizar:

identificar un punto vista actual especificado por el usuario;
identificar el fotograma o fotogramas de un grupo de fotogramas capturados contemporáneamente que corresponden con una porción temporal actual del video que se está renderizando que se necesitan para renderizar la escena representada en los mismos desde el punto de vista identificado;
introducir las representaciones en capas del fotograma o fotogramas de video identificados; y
renderizar el fotograma del video de punto de vista interactivo desde el punto de vista especificado en la actualidad por el usuario usando las representaciones de fotograma en capas introducidas.

14. El procedimiento de la reivindicación 13, en el que los datos de fotograma de video se comprimen y en el que acción de procedimiento de introducir las representaciones en capas del fotograma o fotogramas de video identificados, comprende una acción de decodificación de la porción de los datos de fotograma de video necesaria para obtener las representaciones en capas del fotograma o fotogramas de video identificados.

15. El procedimiento de la reivindicación 13 en el que la acción de procedimiento de identificación del fotograma o fotogramas de un grupo de fotogramas capturados contemporáneamente que corresponden con una porción temporal actual del video que se está renderizando que se necesitan para renderizar la escena representada en los mismos desde el punto de vista identificado, comprende las acciones de:

usar los datos de calibración para determinar los puntos de vista asociados con cada uno de los fotogramas de video de los que se generaron las representaciones de capa;
siempre que el punto de vista identificado coincide con un punto de vista de uno de los fotogramas de video de los que se generaron las representaciones de capa, identificar ese fotograma como el único fotograma que se necesita para renderizar la escena; y
siempre que el punto de vista identificado se sitúa entre los puntos de vista de dos de los fotogramas de video de los que se generaron las representaciones de capa, identificar ambos fotogramas como los fotogramas que se necesitan para renderizar la escena.

16. El procedimiento de la reivindicación 13, en el que la acción de procedimiento de renderización del fotograma del video de punto de vista interactivo, comprende una acción de generación de un fotograma de video de punto de vista interactivo de dos fotogramas de entrada del grupo de fotogramas capturados contemporáneamente que corresponden con una porción temporal actual del video que se está renderizando usando los datos de calibración asociados con los mismos, siempre que el punto de vista identificado se sitúa entre los puntos de vista asociado con dichos dos fotogramas de entrada.

17. El procedimiento de la reivindicación 16, en el que la representación de capa de cada fotograma de entrada comprende una capa principal que comprende información de píxeles asociada con áreas en un fotograma que no muestran discontinuidades de profundidad que exceden un umbral prescrito e información de píxeles de fondo de áreas de discontinuidades de profundidad por encima del umbral y una capa límite que comprende información de píxeles de primer plano asociada con áreas que tienen discontinuidades de profundidad que exceden el umbral, y en el que la acción de procedimiento de generación de un fotograma de video de punto de vista interactivo de dos fotogramas de entrada del grupo de fotogramas capturados contemporáneamente que corresponden con una porción temporal actual del video que se está renderizando, comprende las acciones de:

para cada uno de los dos fotogramas de entrada sucesivamente,
proyectar la capa principal del fotograma de entrada en consideración en una vista correspondiendo con el punto vista actual especificado por el usuario, y
proyectar la capa límite del fotograma de entrada en consideración en la vista correspondiendo con el punto vista actual especificado por el usuario;
mezclar los dos conjuntos resultantes de capas proyectadas para crear un fotograma finalizado del video de punto de vista interactivo.

18. El procedimiento de la reivindicación 17, en el que la acción de procedimiento de mezclar los dos conjuntos resultantes de capas proyectadas comprende una acción de mezclar las capas proyectadas de tal forma que el peso dado a cada una está en proporción directa con qué cercano está el punto de vista asociado con la capa de entrada usada para crear la capa proyectada del punto vista actual especificado por el usuario.

19. El procedimiento de la reivindicación 13, en el que la acción de procedimiento de renderización del fotograma del video de punto de vista interactivo comprende además insertar un objeto no encontrado en las representaciones de fotograma en capas introducidas en el fotograma que se está renderizando.

20. Un medio legible por ordenador que tiene instrucciones ejecutables por ordenador para la realización de las acciones de procedimiento enumeradas en una de las reivindicaciones anteriores.

21. Un sistema de generación de un video de punto de vista interactivo, que comprende:

un subsistema de captura de video que comprende,
una pluralidad de cámaras de video para la captura de retransmisiones de video,
equipo de sincronización para la sincronización de las retransmisiones de video para crear una secuencia de grupos de fotogramas de video capturados contemporáneamente cada uno representando una porción de la misma escena,
uno o más dispositivos informáticos de fin general;
un primer programa informático que tiene módulos de programa ejecutables mediante al menos uno de dicho uno o más dispositivos informáticos de fin general, comprendiendo dichos módulos, un módulo de calibración de cámara para el cálculo de parámetros geométricos y fotométricos asociados con cada retransmisión de video; y
un segundo programa informático que tiene módulos de programa ejecutables mediante al menos uno de dicho uno o más dispositivos informáticos de fin general, comprendiendo dichos módulos,
un módulo de reconstrucción en 3D que genera una reconstrucción en 3D de la escena representada por cada grupo de fotogramas contemporáneos de las retransmisiones de video sincronizadas y que usa la reconstrucción para calcular un mapa de disparidades para cada fotograma en el grupo de fotogramas contemporáneos,
un módulo de acoplamiento que, para cada fotograma en cada grupo de fotogramas contemporáneos, identifica áreas de discontinuidades de profundidad significantes a base del mapa de disparidades del fotograma,
un módulo de representación en capas que, para cada fotograma en cada grupo de fotogramas contemporáneos, genera una capa principal que comprende información de píxeles asociada con áreas en un fotograma que no muestra discontinuidades de profundidad que exceden un umbral prescrito e información de píxeles de fondo de píxeles en áreas que tienen discontinuidades de profundidad que exceden el umbral, y una capa límite que comprende información de píxeles de primer plano asociada con áreas que tienen discontinuidades de profundidad que exceden el umbral, para producir una representación en capas para el fotograma en consideración.

22. El sistema de la reivindicación 21, en el que la pluralidad de cámaras de video se disponen en una manera de una al lado de la otra de tal forma que cada cámara ve una escena desde un punto de vista diferente.

23. El sistema de la reivindicación 22, en el que un campo de vista de cada cámara superpone el campo de vista de cualquier cámara adyacente por una cantidad prescrita.

24. El sistema de la reivindicación 22, en el que la distancia, altura y colocación horizontal del punto de vista de cada cámara en relación con un objeto o área objetivo de la escena se establece para formar una trayectoria prescrita que conecta los puntos de vista de las cámaras.

25. El sistema de la reivindicación 24, en el que la trayectoria prescrita es un arco sustancialmente horizontal.

26. El sistema de la reivindicación 24, en el que la trayectoria prescrita es un arco sustancialmente vertical.

27. El sistema de la reivindicación 24, en el que la trayectoria prescrita es un arco sustancialmente horizontal que recorre hacia arriba desde un extremo al otro.

28. El sistema de la reivindicación 21, en el que una o más de las cámaras de video son cámaras de alta resolución.

29. El sistema de la reivindicación 21, en el que las cámaras de video comprenden una característica de sincronizador.

30. El sistema de la reivindicación 21, en el que cada cámara de video añade metadatos a cada fotograma de video que genera, en el que dichos metadatos comprenden los ajustes de cámara actuales y nivel de exposición de la cámara y una marca de tiempo.

31. El sistema de la reivindicación 21, en el que las cámaras son de un tipo que se controlan a distancia a través de dichos uno o más dispositivos informáticos y en el que el sistema comprende además un tercer ordenador de programa informático que tiene módulos de programa ejecutables mediante al menos uno de dicho uno o más dispositivos informáticos de fin general, en el que dichos módulos comprenden un módulo de programa de captura de video para el control de la pluralidad de cámaras de video para apagarlas o encenderlas simultáneamente y para ajustar sus ajustes de cámara.

32. El sistema de la reivindicación 21, en el que el subsistema de captura de video comprende además equipo de almacenamiento para el almacenamiento de las retransmisiones de video antes del procesamiento.

33. El sistema de la reivindicación 21, en el que el segundo programa informático comprende además un módulo de programa de compresión para la compresión de las representaciones en capas generadas para los fotogramas del video de punto de vista interactivo para facilitar transferencia y/o almacenamiento del video.

34. El sistema de la reivindicación 21, en el que el segundo programa informático comprende además un módulo de programa de generación de archivo de video de punto de vista interactivo para la creación de un archivo que comprende las representaciones en capas generadas de fotogramas de las retransmisiones de video introducidas y la salida del módulo de calibración.

35. Un sistema para la renderización de y presentación de un video de punto de vista interactivo generado mediante el sistema de la reivindicación 21 y el uso de datos que comprende representaciones en capas de fotogramas de video generados de grupos secuenciales de fotogramas de video capturados contemporáneamente cada uno representando una porción de la misma escena, y que comprende datos de calibración que definen parámetros geométricos asociados con la captura de cada fotograma de video, comprendiendo dicho sistema:

un subsistema de interfaz de usuario para la introducción de selecciones de punto de vista de usuario y la presentación de fotogramas de video de punto de vista interactivo renderizados al usuario, que comprende,
un dispositivo de entrada empleado por el usuario para introducir selecciones de punto de vista,
un dispositivo de presentación para la presentación de los fotogramas de video de punto de vista interactivo renderizados al usuario;
un dispositivo informático de fin general;
un programa informático que tiene módulos de programa ejecutables mediante el dispositivo informático de fin general, comprendiendo dichos módulos,
un módulo de decodificación selectivo que decodifica datos especificados asociados con las representaciones en capas de fotogramas de video para cada fotograma del video de punto de vista interactivo a renderizar y presentar,
un módulo de renderización que para cada fotograma del video de punto de vista interactivo que se está renderizando y presentando,
identifica el punto de vista actual seleccionado por el usuario;
especifica al módulo de decodificación selectivo qué fotograma o fotogramas de un grupo de fotogramas capturados contemporáneamente que corresponden con una porción temporal actual del video que se está renderizando y presentando se necesitan para renderizar la escena representada en los mismos desde el punto de vista identificado;
obtiene los datos de fotograma decodificados del módulo de decodificación selectivo; y
renderiza el fotograma del video de punto de vista interactivo desde el punto de vista seleccionado en la actualidad por el usuario usando los datos de fotograma decodificados.

36. El sistema de la reivindicación 35, en el que el subsistema de interfaz de usuario comprende además una interfaz de usuario gráfica que permite al usuario indicar gráficamente el punto de vista, entre los posibles puntos de vista, desde los que se desea ver la escena.

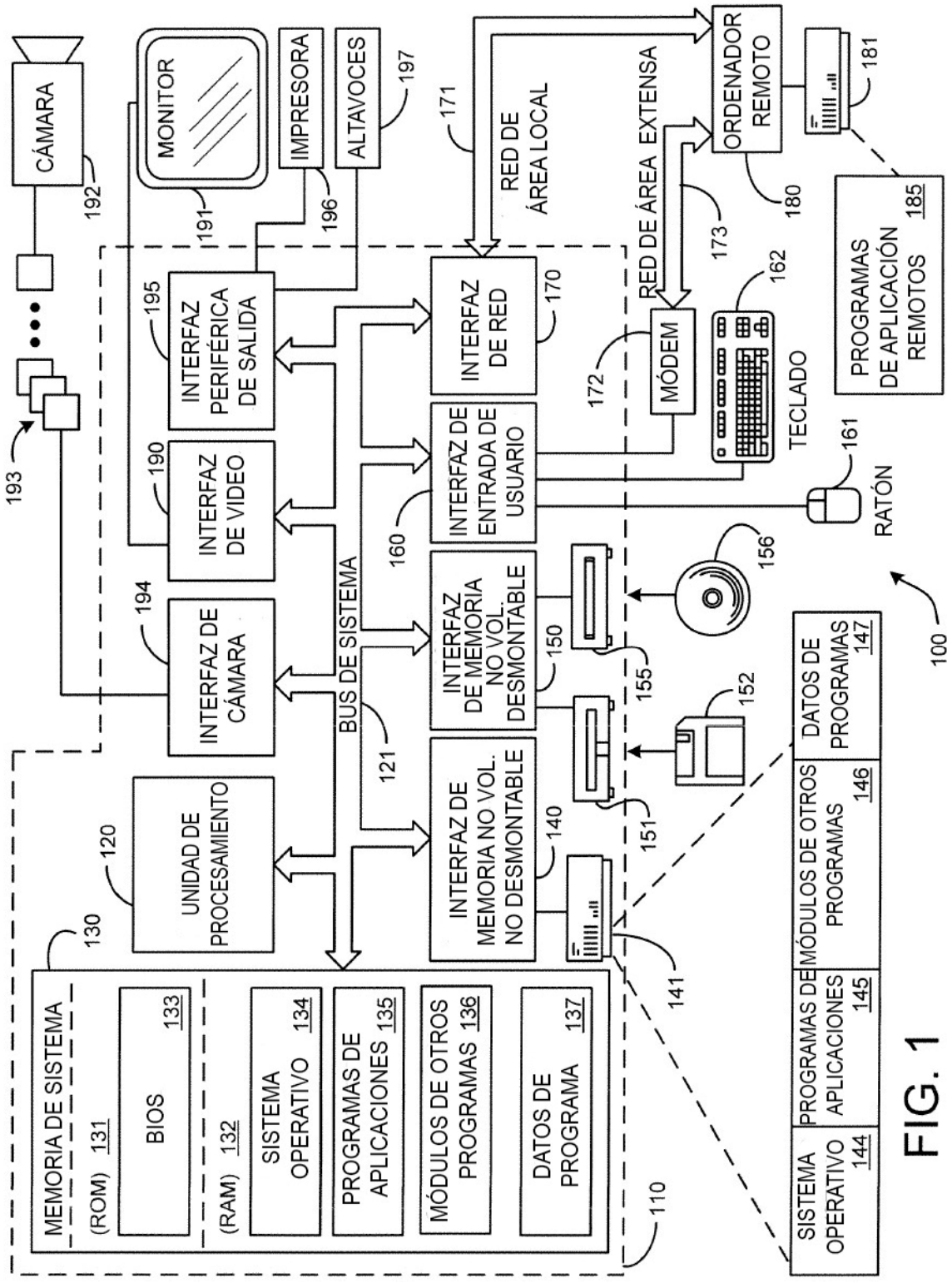
CÁMARA
192
MONITOR
191
IMPRESORA
196
ALTAVOCES
197
171
RED DE ÁREA LOCAL
RED DE ÁREA EXTENSA
ORDENADOR REMOTO
180
181
173
162
TECLADO
161
RATÓN
PROGRAMAS DE APLICACIÓN REMOTOS
185
193
195
INTERFAZ PERIFÉRICA DE SALIDA
190
INTERFAZ DE VIDEO
194
INTERFAZ DE CÁMARA
INTERFAZ DE RED
170
172
MÓDEM
INTERFAZ DE ENTRADA DE USUARIO
160
BUS DE SISTEMA
121
INTERFAZ DE MEMORIA NO VOL. DESMONTABLE
150
155
156
152
INTERFAZ DE MEMORIA NO VOL. NO DESMONTABLE
140
151
141
UNIDAD DE PROCESAMIENTO
120
130
MEMORIA DE SISTEMA
(ROM) 131
BIOS 133
(RAM) 132
SISTEMA OPERATIVO 134
PROGRAMAS DE APLICACIONES 135
MÓDULOS DE OTROS PROGRAMAS 136
DATOS DE PROGRAMA 137
110
SISTEMA OPERATIVO 144
PROGRAMAS DE APLICACIONES 145
MÓDULOS DE OTROS PROGRAMAS 146
DATOS DE PROGRAMAS 147
100

FIG. 1

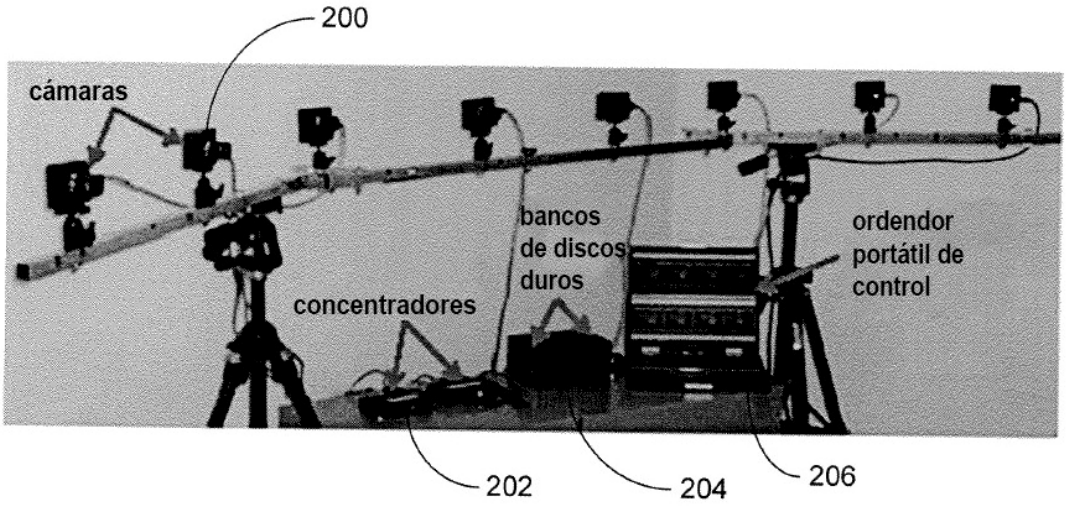
200
cámaras
bancos
de discos
duros
ordendor
portátil de
control
concentradores
202
204
206

FIG. 2

FIG. 7

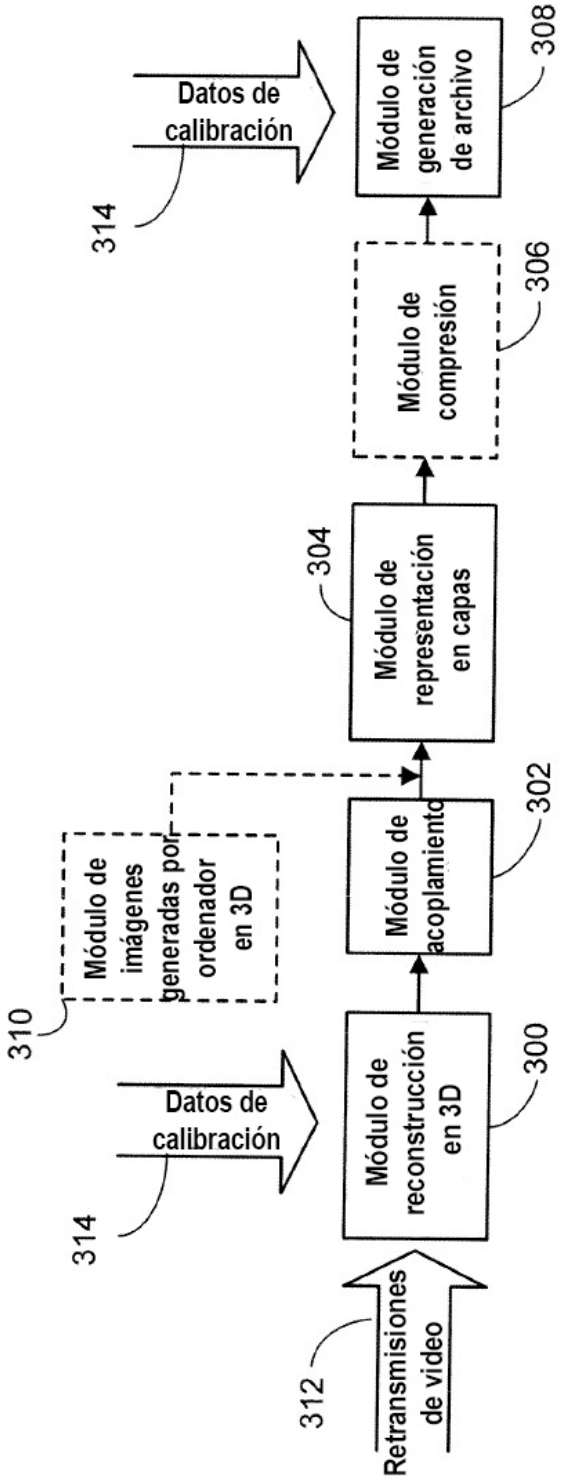
Retransmisiones de video
312
Datos de calibración
314
Módulo de reconstrucción en 3D
300
Módulo de imágenes generadas por ordenador en 3D
310
Módulo de acoplamiento
302
Módulo de representación en capas
304
Módulo de compresión
306
Datos de calibración
314
Módulo de generación de archivo
308

FIG. 3

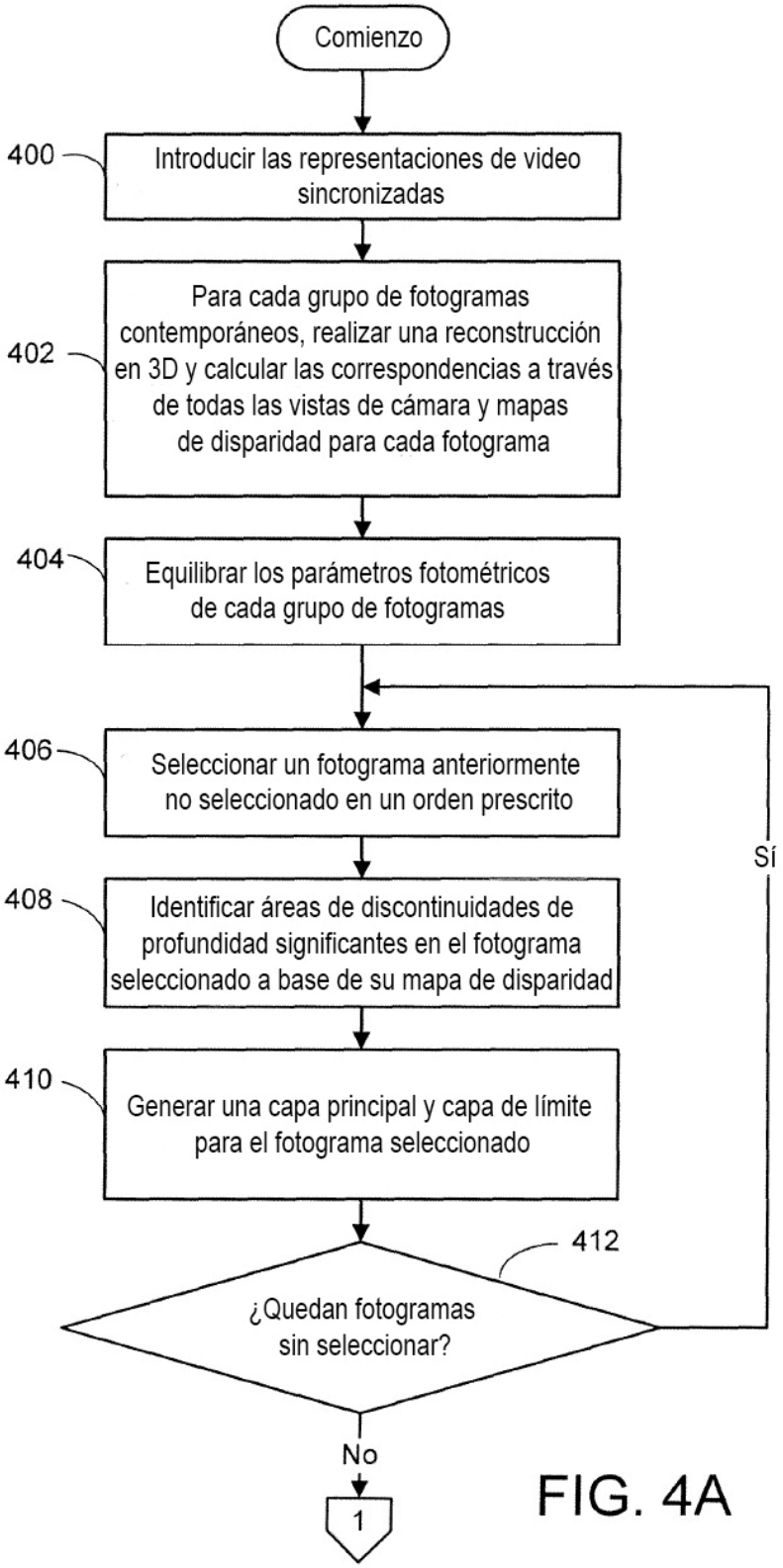
Comienzo
400
Introducir las representaciones de video sincronizadas
402
Para cada grupo de fotogramas contemporáneos, realizar una reconstrucción en 3D y calcular las correspondencias a través de todas las vistas de cámara y mapas de disparidad para cada fotograma
404
Equilibrar los parámetros fotométricos de cada grupo de fotogramas
406
Seleccionar un fotograma anteriormente no seleccionado en un orden prescrito
408
Identificar áreas de discontinuidades de profundidad significantes en el fotograma seleccionado a base de su mapa de disparidad
410
Generar una capa principal y capa de límite para el fotograma seleccionado
412
¿Quedan fotogramas sin seleccionar?
Sí
No
1

FIG. 4A

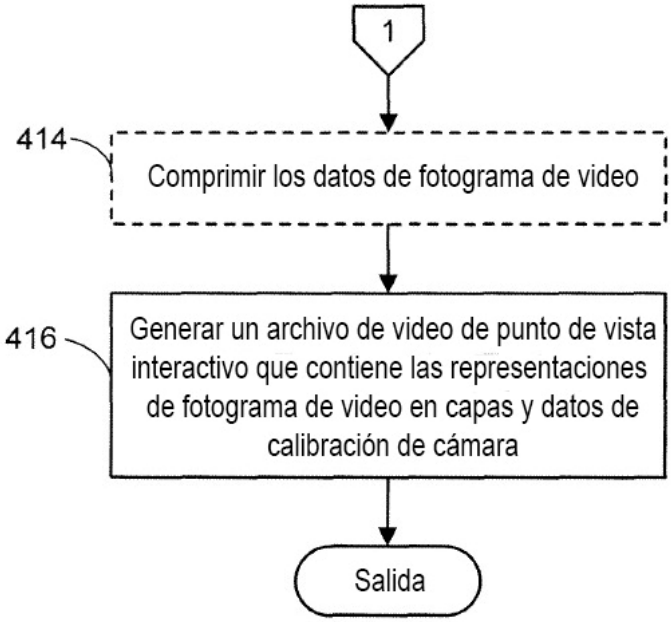
1
414
Comprimir los datos de fotograma de video
416
Generar un archivo de video de punto de vista interactivo que contiene las representaciones de fotograma de video en capas y datos de calibración de cámara
Salida

FIG. 4B

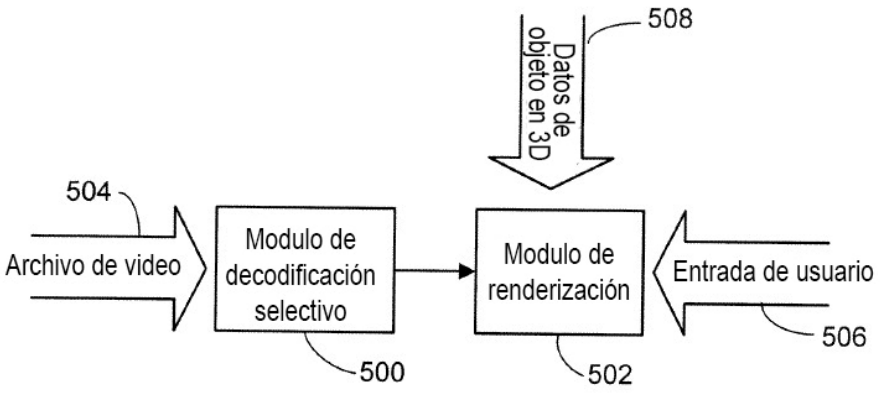

FIG. 5

FIG. 6(a) FIG. 6(b) FIG. 6(c)

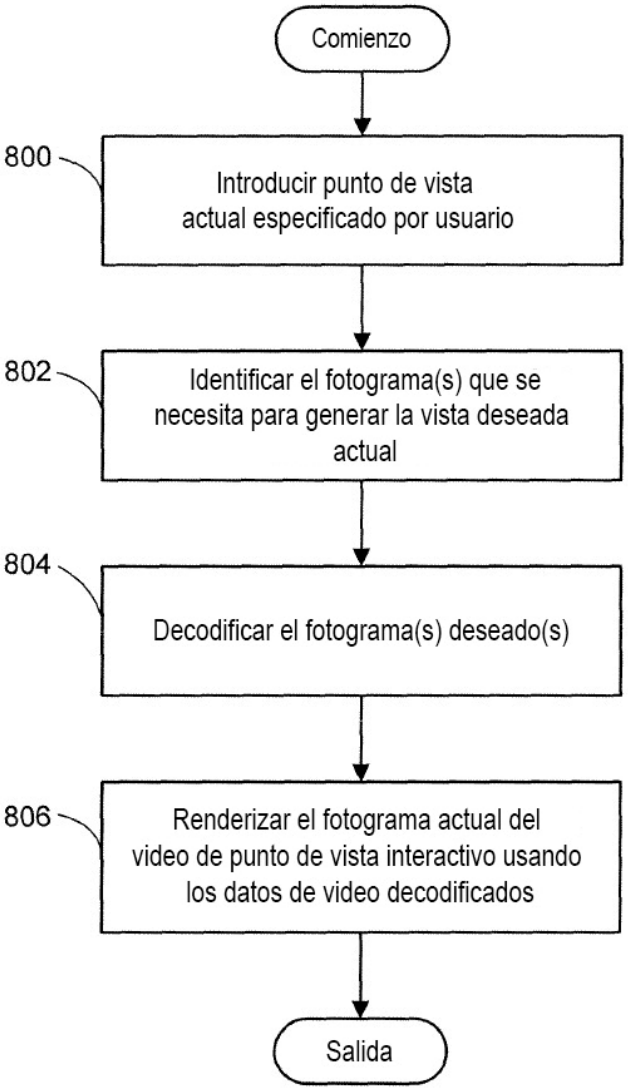
Comienzo
800
Introducir punto de vista actual especificado por usuario
802
Identificar el fotograma(s) que se necesita para generar la vista deseada actual
804
Decodificar el fotograma(s) deseado(s)
806
Renderizar el fotograma actual del video de punto de vista interactivo usando los datos de video decodificados
Salida

FIG. 8