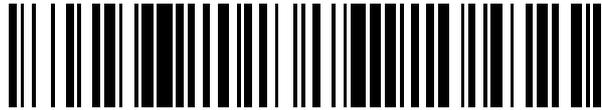


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 561 078**

51 Int. Cl.:

H04N 19/597 (2014.01)
H04N 19/105 (2014.01)
H04N 19/46 (2014.01)
H04N 19/513 (2014.01)
H04N 19/61 (2014.01)
H04N 19/11 (2014.01)
H04N 19/109 (2014.01)
H04N 19/17 (2014.01)
H04N 13/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.07.2010 E 10734970 (6)**

97 Fecha y número de publicación de la concesión europea: **04.11.2015 EP 2594077**

54 Título: **Codificación de vídeo híbrido que soporta síntesis de vistas intermedias**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
24.02.2016

73 Titular/es:

GE VIDEO COMPRESSION, LLC (100.0%)
8 Southwoods Boulevard
Albany, NY 12211, US

72 Inventor/es:

WIEGAND, THOMAS;
MÜLLER, KARSTEN y
MERKLE, PHILIPP

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 561 078 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Codificación de vídeo híbrido que soporta síntesis de vistas intermedias

- 5 **[0001]** La presente invención tiene relación con la codificación de vídeo híbrido que soporta síntesis de vistas intermedias.
- [0002]** Las aplicaciones de vídeo 3D tales como las visualizaciones estéreo y multivista, las aplicaciones de vídeo de punto de vista libre, etc. representan actualmente mercados en auge. Para el contenido de vídeo estéreo y multivista, se ha especificado el estándar MVC. Se hace referencia a la norma ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding", Doc. N9978, Hannover, Alemania, julio de 2008, la ITU-T y la norma ISO/IEC JTC1, "Advanced video coding for generic audiovisual services", la recomendación ITU-T H.264 y la norma ISO/IEC 14496-10(MPEG-4-AVC), Versión 1: mayo de 2003, Versión 2: mayo de 2004, Versión 3: marzo de 2005 (incluyendo la ampliación FExt), Versión 4: septiembre de 2005, Versión 5 y Versión 6: junio de 2006, 10 Versión 7: abril de 2007, Versión 8: julio de 2007 (incluyendo la ampliación SVC), Versión 9: julio de 2009 (incluyendo la ampliación MVC).
- [0003]** Esta norma comprime las secuencias de vídeo procedentes de un número de cámaras adyacentes. El procedimiento de decodificación MVC sólo reproduce estas vistas de cámara en sus posiciones de cámara originales. Para diferentes visualizaciones multivista, sin embargo, se requieren diferentes números de vistas con diferentes posiciones espaciales, de manera que se requieren vistas adicionales, por ejemplo entre las posiciones de cámara originales. De este modo, con el fin de resultar adecuado para todas las visualizaciones multivista diferentes, el contenido de vídeo multivista según la norma MVC tendría que transportar una enorme cantidad de vistas de cámara lo cual, necesariamente, disminuiría la tasa de compresión en relación con la tasa de compresión más baja 25 posible para visualizaciones multivista simplemente aprovechándose de un subconjunto apropiado de las vistas de cámara transportadas. Otras técnicas para transportar datos multivista proporcionan a cada muestra de las tramas de las vistas de cámara no sólo el valor de color correspondiente, sino también un valor de profundidad o disparidad correspondiente basándose en el cual un sintetizador de vista intermedia en la fase de decodificación puede reproducir vistas intermedias proyectando y fusionando vistas de cámara vecinas en la vista intermedia en cuestión. 30 Obviamente, la capacidad de sintetizar vistas intermedias en la fase de decodificación reduce el número de vistas de cámara que han de ser transportadas a través de los datos multivista. Desventajosamente, sin embargo, proporcionar a cada muestra un valor de profundidad o disparidad asociado aumenta la cantidad de datos que han de ser transportados por vista de cámara. Además, los datos de profundidad o disparidad añadidos a los datos de color tienen o bien que ser tratados como un componente de un cuarto color para poder usar un códec de vídeo apropiado para comprimir los datos, o tiene que usarse una técnica de compresión apropiada con el fin de comprimir 35 los datos de color más los de profundidad/disparidad. La primera alternativa no logra la máxima tasa de compresión posible ya que las diferentes estadísticas de los valores de color y profundidad no se consideran correctamente, y la segunda alternativa es engorrosa ya que tiene que diseñarse una solución patentada, y el grado de carga de cálculo en el lado de sintetización es relativamente elevado. 40
- [0004]** En general, sería favorable si, por una parte, la cantidad de datos multivista pudiera mantenerse razonablemente baja, mientras que por otra parte, el número de vistas disponibles en el lado de decodificación sea de una calidad razonablemente elevada.
- 45 **[0005]** El documento US2008/0247462A1 desvela compensación de movimiento para compresión de vídeo usando un "campo de flujo" que comprende un campo por píxel de vectores de movimiento y valores de confianza. Los campos de flujo pueden ser codificados mediante transformación cuantificada para compensación de movimiento de compresión. Los campos de flujo sólo de codificación concuerdan con una o más tramas anteriores y posteriores para determinar tanto la modulación para capas de mejora de resolución, así como el filtrado 50 agudo/suave para una imagen original, una capa base, y para capas de mejora de resolución. Los campos de flujo sólo de codificación pueden usarse con diversos tipos de códec usando la longitud del vector de movimiento de campo de flujo y la confianza para activar los filtros agudo/suave para mejorar la eficiencia por medio de reducción de ruido in situ. Los píxeles pueden ser desplazados a las tramas cercanas usando campos de flujo sólo de codificación, y ponderados para una reducción de ruido eficiente. Los campos de flujo sólo de codificación se desechan después de su uso en la decodificación, y por lo tanto no requieren bits codificados. Los campos de flujo 55 sólo de codificación pueden aplicarse a todos los tipos de trama, incluyendo tramas "F" con predicción intra, con predicción de campo de flujo hacia delante, y tipos de trama "M" de predicción múltiple, y mejoran la eficiencia de intracodificación.

[0006] En su artículo “Multi-View Image Compression and Intermediate View Synthesis for Stereoscopic Applications”, ISCAS 2000, IEEE International Symposium on Circuits and Systems, 28-31 de mayo, Ginebra, Suiza, Gan-Chieh Chang y Wen-Nung Lie describen un sistema de compresión de imagen multivista que es capaz de proporcionar información de disparidad suficiente y fiable para síntesis de vistas intermedias para aplicaciones de realidad virtual 3D. La primera parte del artículo presenta un códec que ofrece acceso aleatorio a cualquier vista decodificada y tiene un rendimiento comparable al de la norma MPEG. En la segunda parte, las disparidades son interpoladas, y se hace referencia a imágenes decodificadas apropiadas para la síntesis de cualquier vista intermedia. El diseño de dicho trabajo proporciona suficiente información de manera que cada síntesis puede lograrse al cabo de 0,04 s para una imagen de 320x240 (excluyendo el tiempo de decodificación). También se adopta un visor ocular con sincronización infrarroja para ver los resultados estéreo.

[0007] En su artículo “Intermediate View Synthesis From Binocular Images for Stereoscopic Applications”, Department of Electrical Engineering, National Chung Cheng University, Taiwan, 2001, Wen-Nung Lie y Bo-Er Wei proponen un nuevo procedimiento para sintetizar vistas intermedias a partir de imágenes binoculares captadas por cámaras paralelas. Las técnicas principales incluyen: (1) procedimiento de coincidencia de bloques predictivo y multifase para estimación de disparidad, (2) interpolación de disparidad para mala correspondencia y regiones de oclusión, y (3) síntesis de imagen mediante reproducción de imagen de bloque hipotetizado y comprobado. El procedimiento ofrece una estimación rápida (3 veces) y exacta del mapa de disparidad, así como un compromiso entre la calidad de síntesis (disparidades basadas en píxeles) y la velocidad de síntesis (reproducción basada en bloques de 4x4). En el documento “A Layered Method of Visibility Resolving in Depth Image-based Rendering”, Yu Huang y Chao Zhang, Thomson Corporate Research, Princeton, NJ 08540, US, State Key Lab of Machine Perception, Peking University, Pekín 100871, China, 2008, presentan un procedimiento por capas para resolver el problema de visibilidad en la reproducción basada en imágenes profundas. Se propone una representación novedosa de tres capas para cada vista de referencia, es decir, la capa principal, la capa de fondo y la capa límite. Un procedimiento espacial-temporal está diseñado para generar la capa límite para reproducción basada en píxeles (salpicadura). Mientras tanto, se construye un modelo de fondo temporal para cada trama buscando hacia atrás y hacia delante información de fondo sin cubrir basándose en la variación de profundidad en el vídeo de referencia.

[0008] En el documento “Overview of Multi-view Video Coding”, Yo-Sung Ho y Kwan-Jung Oh, Gwangju Institute of Science and Technology (GIST), Gwangju, Corea, 2007, proponen un sistema de vídeo multivista que puede proporcionar un realismo aumentado a través de una experiencia de visión selectiva. El vídeo multivista es una colección de múltiples vídeos que captan la misma escena 3D en diferentes puntos de vista. Puesto que el tamaño de los datos del vídeo multivista aumenta proporcionalmente al número de cámaras, es necesario comprimir los datos de vídeo multivista para un almacenamiento y una transmisión eficientes. Este trabajo proporciona una perspectiva general de la codificación de vídeo multivista (MVC) y describe sus aplicaciones, requisitos, y el modelo de software de referencia para MVC .

[0009] Por lo tanto, un objeto de la presente invención es proporcionar un esquema para transmisión y recepción multivista que, por una parte, permite transportar un elevado número de vistas disponibles y, por otro lado, simplemente plantea una demanda razonable tanto de tasa de transmisión como de tara de reproducción.

[0010] Este objeto se logra mediante un decodificador de vídeo híbrido según la reivindicación 1, un procedimiento según la reivindicación 7, y un programa informático según la reivindicación 8.

[0011] La presente invención está basada, entre otras cosas, en la conclusión de que los códec de vídeo híbrido según los cuales los vídeos de múltiples vistas son codificados predictivamente en una señal de datos multivista, estando las tramas de un vídeo de una cierta vista subdivididas espacialmente en subregiones y teniendo la señal de datos multivista un modo de predicción de un conjunto de posibles modos de predicción, asociado con cada una de las subregiones, comprendiendo el conjunto de posibles modos de predicción al menos un modo de predicción entrevista y un modo de predicción intravista, ya transportan suficiente información con el fin de permitir una síntesis de vistas intermedias en el lado de decodificación de vídeo híbrido. Es decir, no es necesario ningún formato de datos multivista patentado según el cual los datos de color están acompañados por datos adicionales de profundidad por píxel y/o de disparidad. En otras palabras, los inventores de la presente solicitud averiguaron que incluso cuando al codificador de vídeo híbrido se le da la libertad de seleccionar libremente el modo de predicción preferente de los posibles modos de predicción para cada subregión – según algún esquema de optimización para optimizar una medida de tasa/distorsión, o similar -, los vectores de disparidad transportados realmente dentro de la señal de datos multivista resultante para la subregiones para las cuales se ha escogido el modo de predicción entrevista, son suficientes con el fin de permitir una síntesis de vistas intermedias en la fase de decodificación de vídeo híbrido. Es decir, aunque un reconstructor predictivo reconstruye subregiones de tramas de un vídeo de una cierta vista de la

señal de datos multivista, con la cual está asociado el modo de predicción intervista, generando una predicción a partir de una versión reconstruida de una porción de tramas de un vídeo de otra vista de la señal de datos multivista usando los vectores de disparidad extraídos de la señal de datos multivista para las subregiones respectivas, y también se extrae un residuo de predicción para las subregiones respectivas a partir de la señal de datos multivista, un sintetizador de vista intermedia puede reconstruir porciones de un vídeo de vista intermedia usando la versión reconstruida de las porciones de las tramas del vídeo de la cierta vista, y los vectores de disparidad extraídos de la señal de datos multivista. Las porciones restantes del vídeo de vista intermedia que no son reconstruidos usando los vectores de disparidad extraídos de la señal de datos multivista – puesto que el codificador de vídeo híbrido decidió usar el código de predicción intravista para otras subregiones -, pueden ser rellenadas posteriormente por medio de intra/extrapolación en el tiempo y/o espacialmente, o estimando vectores de disparidad adicionales interpolando vectores de disparidad extraídos de la señal de datos multivista, temporalmente y/o espacialmente.

[0012] Las realizaciones preferentes de la presente solicitud se describen a continuación con respecto a las figuras. En particular:

15

La fig. 1 muestra un diagrama esquemático que ilustra la coexistencia de vectores de movimiento y disparidad en un esquema de codificación multivista respecto a un ejemplo que incluye dos vistas y dos instancias de tiempo, subyacente al esquema híbrido de codificación y decodificación descrito con respecto a las figs. 5 a 8;

20 la fig. 2 muestra un diagrama esquemático que ilustra las correspondencias puntuales definidas por un vector de disparidad entre vistas adyacentes de acuerdo con una realización de la presente invención;

la fig. 3 muestra un diagrama esquemático que ilustra una síntesis de vistas intermedias mediante proyección de contenido de escena a partir de dos vistas vecinas usando vectores de disparidad a escala de acuerdo con una realización de la presente invención;

la fig. 4 muestra un diagrama de bloques de un codificador de vídeo híbrido según una realización de la presente invención;

30 la fig. 5 muestra un diagrama de bloques de un decodificador de vídeo híbrido según una realización de la presente invención;

la fig. 6 muestra un diagrama esquemático que ilustra el modo de funcionamiento del decodificador de vídeo híbrido de la fig. 5 con respecto a la síntesis de vistas intermedias de acuerdo con una realización de la presente invención;

35

la fig. 7 muestra un diagrama esquemático que ilustra un esquema de decodificación de N vistas de un tren de datos de 2 vistas para una visualización de 9 vistas en la cual puede emplearse el decodificador de vídeo híbrido de la fig. 5.

40 **[0013]** Antes de describir diversas realizaciones de un decodificador de vídeo híbrido o un procedimiento de decodificación de vídeo híbrido así como un codificador de vídeo híbrido o un procedimiento de codificación de vídeo híbrido correspondiente, estas realizaciones están motivadas en primer lugar explicando el uso de vectores de disparidad en la codificación predictiva de datos de vistas múltiples.

45 **[0014]** Si el contenido de escena es captado con múltiples cámaras, una percepción 3D de este contenido puede presentarse a un espectador. Con este fin, tienen que proporcionarse recorridos estéreo con una dirección de visualización diferente para el ojo izquierdo y el derecho. La desviación del mismo contenido en ambas vistas para instancias de tiempo iguales se representa por la denominada paralaje. En otras palabras, la paralaje describe una desviación de muestras dentro de una vista en relación con las posiciones correspondientes dentro de otra vista.

50 Puesto que ambas vistas muestran el mismo contenido de escena, ambas vistas son muy similares dentro de las porciones relacionadas entre sí por medio de la paralaje. Igualmente, las tramas consecutivas de un vídeo que corresponden a una vista individual comprenden similitudes entre sí. Por ejemplo, en caso de una cámara no móvil, las muestras que corresponden a un fondo estático debería aparecer constantemente dentro de tramas consecutivas del vídeo en posiciones cúbicadas espacialmente. Los objetos móviles dentro del contenido de escena cambian sus posiciones dentro de tramas consecutivas del vídeo. En las técnicas de comprensión de vídeo híbrido, las similitudes entre tramas temporalmente consecutivas se aprovecha por medio de la predicción de movimiento compensado según la cual se usan vectores de movimiento con el fin de obtener predicciones para ciertas subregiones de una trama basándose en porciones previamente codificadas y reconstruidas de otras tramas, principalmente correlacionando porciones de los mismos en la subregión en cuestión.

55

[0015] Igualmente, con el fin de comprimir datos multivista, la similitud entre las tramas del mismo instante de tiempo de direcciones de vista espacialmente distintas pero similares puede aprovecharse con el fin de comprimir predictivamente el contenido de vídeo de estas vistas. La desviación del mismo contenido en ambas vistas para 5 instancias de tiempo iguales puede representarse mediante vectores de disparidad. Esta desviación es comparable a la desviación de contenido dentro de una secuencia de tramas entre diferentes instancias de tiempo representadas por los vectores de movimiento anteriormente mencionados. La fig. 1 ilustra la utilización conjunta de vectores de disparidad y vectores de movimiento con el fin de reducir la redundancia de datos multivista para un caso ilustrativo de dos vistas en dos instancias de tiempo.

10

[0016] En particular, la fig. 1 muestra una trama 10 de una primera vista que corresponde a un instante de tiempo t y una segunda trama 12 de la misma vista 1 que corresponde al instante de tiempo $t-1$, y además, se muestra una trama 14 de una segunda vista que corresponde al instante de tiempo t y una trama adicional 16 de la vista 2 en el instante de tiempo $t-1$. Un vector de movimiento 18 ilustra el desplazamiento espacial de contenido de escena similar 15 dentro de las tramas consecutivas 10 y 12 de la primera vista, con un vector de movimiento 20 que ilustra igualmente el desplazamiento espacial de contenido de escena mutuamente correspondiente en las tramas consecutivas 14 y 16 de la segunda vista. Tal como se explicó anteriormente, el movimiento de contenido de escena mutuamente correspondiente dentro de tramas consecutivas dentro de una vista individual varía espacialmente, dependiendo del contenido de escena, y por lo tanto, en la codificación de vídeo híbrido con la cual están relacionadas las siguientes 20 realizaciones, los vectores de movimiento son asignados individualmente para diferentes subregiones de las tramas 10 a 16 con el fin de indicar, para la subregión respectiva, cómo ha de desplazarse la trama de referencia a la cual apunta o se refiere el vector de movimiento respectivo 18 y 20 con el fin de servir como predicción en la subregión respectiva de la trama actual. Hasta ahora, en la fig. 1, las tramas 12 y 16 representan las tramas de referencia para predecir porciones de las tramas 10 y 14, respectivamente, usando vectores de movimiento 18 y 20, 25 respectivamente. Un codificador de vídeo híbrido puede estar configurado para establecer los vectores de movimiento 18 y 20 de manera que se minimice una cierta medida de tasa/distorsión considerando que representar los vectores de movimiento 18 y 20 a una resolución final aumenta la velocidad binaria necesaria para transportar la información de movimiento en tanto que, por otra parte, aumentando la calidad de predicción y por lo tanto, reduciendo el error de predicción y la velocidad binaria necesaria para codificar el error de predicción. Con el fin de 30 determinar el vector de movimiento para una cierta subregión, el codificador de vídeo híbrido puede, por ejemplo, determinar la similitud de porciones de la trama de referencia 12 y 16, respectivamente, desplazadas en relación con la subregión en cuestión dentro de la trama actual 10 y 14, respectivamente, mediante posibles vectores de movimiento diferentes escogiendo, como candidatos a vectores de movimiento, aquellos vectores de movimiento que conducen a un error de predicción local bajo o mínimo tal como se mide mediante el error cuadrático medio.

35

[0017] En un sentido similar, los vectores de disparidad 22 y 24, respectivamente, muestran un desplazamiento espacial de contenidos de escena mutuamente correspondientes dentro de las tramas 12, 16 y 10, 14 en el mismo instante de tiempo de las diferentes vistas 1 y 2, y el vídeo híbrido y el codificador pueden establecer estos vectores de disparidad 22 de una manera que corresponde a la determinación de los vectores de movimiento 18 y 20 40 esbozados anteriormente con, por ejemplo, las tramas 10 y 12 de la vista 1 representando las tramas de referencia para los vectores de disparidad 22 y 24, que a su vez ayudan a las tramas 12 y 10 que han de ser desplazadas espacialmente con el fin de servir como predicción para las subregiones de las tramas 14 y 16 a las cuales corresponden los vectores de disparidad 22 y 24. Por lo tanto, la estimación de movimiento tal como se realiza mediante un codificador de vídeo híbrido, es aplicable no sólo a la dirección temporal, sino también a una dirección 45 entrevista. En otras palabras, si se codifican juntas múltiples vistas, las direcciones temporal e entrevista pueden tratarse de manera similar, de manera que la estimación de movimiento se lleva a cabo en la dirección temporal así como en la entrevista durante la codificación. Los vectores de movimiento estimados en la dirección entrevista son los vectores de disparidad 22 y 24. Como los vectores de disparidad corresponden al desplazamiento espacial de contenido de escena mutuamente correspondiente dentro de diferentes vistas, tales codificadores de vídeo híbrido 50 también llevan a cabo la estimación de disparidad implícitamente y los vectores de disparidad 22 y 24 tal como se incluyen en el tren de bits codificado, pueden aprovecharse para la síntesis entrevista como se esbozará con más detalle más adelante. Estos vectores 22 y 24 pueden usarse para síntesis de vistas intermedias adicionales en el decodificador.

55 **[0018]** Con el fin de ilustrar esto con más detalle, se hace referencia a la fig. 2. Considérese un píxel $p_1(x_1, y_1)$ en la vista 1 en la posición (x_1, y_1) y un píxel $p_2(x_2, y_2)$ en la vista 2 en la posición (x_2, y_2) , los cuales tienen valores de luminancia idénticos o, en otras palabras, representan muestras de escena mutuamente correspondientes. Entonces, considérese un píxel $p_1(x_1, y_1)$ en la vista 1 en la posición (x_1, y_1) y un píxel $p_2(x_2, y_2)$ en la vista 2 en la posición (x_2, y_2) , los cuales tienen valores de luminancia idénticos. Entonces:

$$p_1(x_1, y_1) = p_2(x_2, y_2). \quad (1)$$

[0019] Sus posiciones (x_1, y_1) y (x_2, y_2) están conectadas por el vector de disparidad 2D 26, por ejemplo desde la vista 2 hasta la vista 1, el cual es $d_{21}(x_2, y_2)$ con las componentes $d_{x,21}(x_2, y_2)$ y $d_{y,21}(x_2, y_2)$. Por lo tanto, es válida la siguiente ecuación:

$$(x_1, y_1) = (x_2 + d_{x,21}(x_2, y_2), y_2 + d_{y,21}(x_2, y_2)). \quad (2)$$

combinando (1) y (2),

$$p_1(x_2 + d_{x,21}(x_2, y_2), y_2 + d_{y,21}(x_2, y_2)) = p_2(x_2, y_2). \quad (3)$$

Tal como se muestra en la figura 2, abajo a la derecha, dos puntos con contenido idéntico pueden estar conectados con un vector de disparidad 26: Sumar este vector a las coordenadas de p_2 da la posición de p_1 en coordenadas de imagen. Si ahora se cambia la escala del vector de disparidad $d_{21}(x_2, y_2)$ por un factor $\alpha \in [0, 1]$, puede abordarse cualquier posición intermedia entre (x_1, y_1) y (x_2, y_2) . Por lo tanto, pueden generarse vistas intermedias desplazando el contenido de imagen de la vista 1 y/o la vista 2 mediante vectores de disparidad a escala. En la figura 3 se muestra un ejemplo para una vista intermedia.

Por lo tanto, pueden generarse nuevas vistas intermedias con cualquier posición entre la vista 1 y la vista 2.

Aparte de esto, también puede lograrse una extrapolación de vista usando factores de escala $\alpha < 0$ y $\alpha > 1$ para las disparidades.

Estos procedimientos de escala también pueden aplicarse en la dirección temporal, de manera que pueden extraerse nuevas tramas cambiando la escala de los vectores de movimiento, lo cual conduce a la generación de secuencias de vídeo de frecuencia de tramas más elevada.

Después de haber ilustrado la posibilidad de usar los vectores de disparidad tal como son generados y transmitidos por un codificador multivista híbrido en la síntesis de vistas intermedias, o al menos los principios subyacentes a ello, a continuación se describen realizaciones para un esquema de codificación de vídeo híbrido que soporta síntesis de vistas intermedias. En particular, la fig. 4 muestra un codificador de vídeo híbrido que es adecuado para generar una señal de datos multivista basándose en la cual se permite la decodificación de vídeo híbrido, que soporta síntesis de vistas intermedias tal como se describe con respecto a las figs. 5 y 6 siguientes.

El codificador de vídeo híbrido según la fig. 4 está indicado en general con el signo de referencia 30. El codificador de vídeo híbrido 30 de la fig. 4 es un codificador predictivo que soporta uno o más modos de predicción intermedia, y uno o más modos de intrapredicción. Además, el codificador de vídeo híbrido 30 de la fig. 4 está configurado para seleccionar y establecer el modo de predicción a una granularidad de una subtrama, concretamente en unidades de subregiones de las tramas de las vistas que han de ser codificadas.

En particular, el codificador de vídeo híbrido de la fig. 4 comprende una entrada 32 para un vídeo de primera vista 34, y una entrada 34 para un vídeo de segunda entrada 36. El vídeo de primera vista 34 se considera que es el resultado de una captación de una escena desde una dirección de primera vista, mientras que el vídeo de segunda vista 36 se espera que represente una captación de la misma escena desde una segunda vista que es diferente de la primera vista. La primera y la segunda vistas difieren, por ejemplo, en la posición de vista, es decir, la posición de captación/cámara y/o el ángulo de vista, es decir, la dirección del eje de vista. La primera y la segunda vista pueden diferir simplemente en la posición de vista siendo la dirección del eje de vista la misma. En general, la primera y la segunda vistas pueden estar colocadas una en relación con otra de manera que las mismas ubicaciones de objeto en la escena, colocadas a una distancia media de los objetos de escena captados por la primera y la segunda vistas, están desplazadas dentro de las imágenes de ambas vistas en menos de 5 píxeles, o, aún con más preferencia, menos de 2 píxeles.

Además, el codificador de vídeo híbrido 30 de la fig. 4 comprende una salida 38 para producir como salida la señal de datos multivista. En medio, el codificador de vídeo híbrido 30 comprende dos bucles de estimación de predicción 40 y 42, respectivamente, el primero de los cuales está conectado entre la primera entrada 32 y la salida 38, y el segundo de los cuales está conectado entre la segunda entrada 34 y la salida 38. En particular, el primer bucle de estimación de predicción 40 comprende un sustractor 44 y una fase de

cuantificación/escalado/transformación 46 conectada, en el orden mencionado, entre la entrada 32 y una primera entrada de un generador de señal de datos 48, cuya salida está conectada a la salida 38. Además, el primer bucle de predicción 40 comprende un bloque de reescalado/transformación inversa 15, un filtro de desagrupamiento 52, y un reconstructor predictivo 54, los cuales están conectados en el orden mencionado entre una salida de la fase de cuantificación/escalado/transformación 46 y una entrada de inversión del sustractor 44. De manera similar, el segundo bucle de estimación de predicción se forma conectando en serie un sustractor 56, una fase de cuantificación/escalado/transformación 58, un bloque de reescalado/transformación inversa 60, un filtro de desagrupamiento 62 y el reconstructor predictivo 54. Para ser más exactos, el constructor predictivo 54 está conectado dentro de los dos bucles de estimación de predicción 40 y 42, respectivamente, y comprende un primer par de entrada y salida conectadas dentro del primer bucle de estimación de predicción 40 y un segundo par de entrada y salida conectadas dentro del segundo bucle de estimación de predicción 42. Además el sustractor 56 y la fase de cuantificación/escalado/transformación 58 están conectados en el orden mencionado entre la entrada 34 y otra entrada del generador de señal de datos 48, mientras que el bloque de reescalado/transformación inversa 60 y el filtro de desagrupamiento 62 están conectados en serie en el orden mencionado entre la salida de la fase de cuantificación/escalado/transformación 58 y la entrada correspondiente del reconstructor predictivo 54. Finalmente, otra salida del reconstructor predictivo 54 está conectada a otra entrada del generador de señal de datos 48. Por último, la salida del reconstructor predictivo 54 conectado dentro del primer bucle de estimación de predicción 40 también está conectada a una segunda entrada de un sumador 64 conectado, por su primera entrada, entre el bloque de reescalado/transformación inversa 50 y el filtro de desagrupamiento 52, y de manera similar, la otra salida del reconstructor predictivo 54 también está conectada a una segunda entrada de un sumador 66, que está conectado a través de su primera entrada entre el bloque de reescalado/transformación inversa 60 y el filtro de desagrupamiento 62. Después de haber descrito la estructura general del codificador de vídeo híbrido 30 de la fig. 4, más adelante se describe su modo de funcionamiento.

[0028] Cada vídeo 32 y 36 consiste en una secuencia de tramas 68 y 70, respectivamente, siendo cada trama 68 y 70 una matriz de muestras que representan un valor de color de la escena captada por ambos vídeos 34 y 36. Cada trama 68 y 70 está subdividida en subregiones 72, es decir, grupos de muestras inmediatamente adyacentes de las tramas 68 y 70, respectivamente. La subdivisión de las tramas puede ser constante en el tiempo para cada vídeo 34 y 36, y pueden corresponder espacialmente entre sí cuando se comparan el vídeo 34 y el 36. Por ejemplo, la subdivisión espacial de las tramas en subregiones puede ser tal que las subregiones forman una disposición regular de bloques dispuestos en columnas y filas, como se muestra de manera ejemplar con respecto a la trama 68. Alternativamente, la subdivisión espacial de la trama 68 y el 70 en subregiones puede variar en el tiempo tal como basándose en trama por trama. El reconstructor predictivo 54 puede ser responsable de establecer la subdivisión espacial con el objetivo de optimizar alguna medida de tasa/distorsión tal como se esboza con más detalle más adelante. Con este fin, las subregiones 72 pueden ser los bloques de hojas de un multiárbol, tal como un árbol cuaternario, la subdivisión de la trama 68 y el 70 como se ilustra de manera ejemplar con respecto al vídeo 36. En este caso, el reconstructor predictivo 54 puede señalar la subdivisión seleccionada al generador de señal de datos 48 que ha de ser insertada dentro de la señal de datos multivista 38. La subdivisión puede estar diseñada de manera que un límite inferior del tamaño de las subregiones es 4x4 posiciones de muestras de color, o de manera que un promedio del tamaño de posibles tamaños de las subregiones entre las cuales puede escoger el reconstructor predictivo durante la subdivisión, es mayor que 4x4 muestras.

[0029] En general, la subdivisión espacial de las tramas 68 y 70 en subregiones forma la granularidad a la cual el reconstructor predictivo 54 asigna diferentes modos de predicción a diferentes regiones espaciales de las tramas 68 y 70. Tal como se describió anteriormente, el reconstructor predictivo 54 soporta, al menos, uno o más modos de predicción intervista, y uno o más modos de predicción intravista. El modo de predicción intervista puede expresarse tal como se esbozó anteriormente con respecto a las figs. 1 y 2 y un ejemplo de un modo de predicción intravista es el modo de predicción de movimiento compensado también ilustrado anteriormente con respecto a las figs. 1 y 2. Ejemplos adicionales para modos de predicción intravista engloban un modo de intrapredicción según el cual se usan valores de muestra ya codificados y reconstruidos de subregiones vecinas de la trama actual dentro del mismo vídeo o vista para predecir – mediante interpolación o extrapolación, los valores de muestra de una subregión actual 72. Un modo de predicción intravista adicional puede suprimir cualquier predicción de manera que los valores de muestra dentro de esta subregión son codificados dentro de la señal de datos multivista de una manera no predicha.

[0030] Dependiendo del modo de predicción, el reconstructor de predicción 54 asigna diferente información de predicción a una subregión que ha de ser codificada actualmente y señala la misma al generador de señal de datos 48 para ser introducida dentro de la señal de datos multivista en la salida 38. Generalmente, esta información de predicción permite que el decodificador de vídeo híbrido recupere el mismo resultado de predicción que el reconstructor de predicción 54 a partir de tramas codificadas/decodificadas previamente.

[0031] En el sustractor 44, la predicción de la subregión que ha de ser codificada actualmente es sustraída de los valores de muestra de la subregión que ha de ser codificada actualmente, tras lo cual el error de predicción así obtenido es cuantificado y transformado en el bloque 46. En particular, el bloque 46 puede aplicar una transformación de descomposición espectral sobre el error de predicción con una cuantificación posterior de los coeficientes de transformación. Los datos de residuo de predicción así obtenidos se pasan al generador de señal de datos 48 para una incorporación a la señal de datos multivista en la salida 38, así como el bloque 50 para reconstruir el error de predicción que entra en el bloque 46 y desviar de éste simplemente debido a la cuantificación realizada en el bloque 46. El bloque 50 aplica una descuantificación seguida por una transformación inversa sobre los niveles de coeficiente de transformación y produce como salida el residuo de predicción reconstruido para la primera entrada del sumador 64 donde se realiza una suma con una predicción usada previamente con el fin de obtener el residuo de predicción respectivo. Así, en la salida del sumador 64, se produce como salida una reconstrucción de la subregión actual y el filtro de desagrupamiento 52, que es opcional, recibe la reconstrucción de esta subregión junto con la reconstrucción de las otras subregiones de la trama actual para producir como salida una reconstrucción de la antigua trama, es decir, el codificado/decodificado previamente, para que se pase al reconstructor predictivo 54.

[0032] La descripción recién presentada se refería a la codificación de subregiones de tramas en el vídeo de primera vista 34, pero esta descripción puede transferirse fácilmente a la funcionalidad del bucle de estimación de predicción 42 con respecto a la codificación de subregiones de tramas 70 del vídeo de segunda vista 36.

[0033] Como ya se mencionó anteriormente, el reconstructor predictivo 54 tiene que realizar muchas decisiones durante la codificación/compresión de los valores de muestra de las tramas 68 y 70 de los vídeos 34 y 36, concerniendo las decisiones, opcionalmente, a subdivisiones espaciales de las tramas en subregiones 72 y, para cada subregión, la selección de un modo de predicción que ha de usarse para codificar la subregión respectiva junto con los detalles de predicción respectivos concernientes al modo de predicción seleccionado. Por ejemplo, para una subregión que tiene un modo de predicción entrevista asociado con la misma, el reconstructor predictivo 54 también determina el vector de disparidad anteriormente mencionado. En particular, el reconstructor predictivo 54 puede estar configurado para determinar exactamente un vector de disparidad por subregión, mientras que la granularidad a la cual se varía espacialmente el modo de predicción sobre las tramas, puede ser más gruesa, tal como en unidades de grupos de una o más subregiones vecinas.

[0034] Basándose en el vector de disparidad, la predicción para la subregión respectiva se determina correlacionando las posiciones de las muestras de la subregión respectiva según el vector de disparidad para obtener posiciones muestreadas correlacionadas, y adoptando como la predicción la versión reconstruida de la trama correspondiente temporalmente del otro de los vídeos 34 y 36 en las posiciones de muestra correlacionadas. La correlación puede ser una correlación lineal tal como, por ejemplo, un desplazamiento de traslación en una cantidad y dirección determinadas por el vector de disparidad. Con el fin de optimizar los ajustes de predicción, el reconstructor predictivo 54 puede probar vectores de disparidad diferentes dentro de una cierta área de búsqueda alrededor del vector cero, y determinar el error de predicción resultante, así como la velocidad binaria resultante necesaria para representar el error de predicción mediante coeficientes de forma cuantificados para estos vectores de disparidad diferentes. El área de búsqueda, por ejemplo, restringe los posibles vectores de disparidad para una cierta subregión a una cierta longitud máxima de los vectores de disparidad. La dirección de los posibles vectores de disparidad que son sometidos a pruebas respectivas en la determinación del vector de disparidad óptimo, sin embargo, puede ser sin restricciones o restringida a direcciones horizontales teniendo en cuenta que las disparidades entre diferentes vistas por lo general se extienden a lo largo de la dirección horizontal más que la vertical. El área de búsqueda puede extenderse incluso simplemente en una dirección horizontal en relación con el vector cero aprovechando que las disparidades normalmente apuntan en una cierta dirección de las direcciones izquierda y derecha.

[0035] El reconstructor predictivo 54 puede estar configurado para determinar, para cada subregión para la cual se escoge el modo de predicción entrevista, un vector de disparidad. Sin embargo, el reconstructor predictivo 54 también puede analizar el resultado de búsqueda anteriormente mencionado de las otras pruebas de posibles vectores de disparidad dentro del área de búsqueda anteriormente mencionada. Por ejemplo, el reconstructor predictivo 54 puede estar configurado para asignar una fiabilidad al vector de disparidad seleccionado finalmente. Como ya se describió anteriormente, los vectores de disparidad seleccionados no son necesariamente el que conduce al error de predicción más bajo, aunque es muy probable que el error de predicción que resulta del vector de disparidad seleccionado sea relativamente bajo. De acuerdo con una realización, el reconstructor predictivo 54 determina la fiabilidad asignada al vector de disparidad seleccionado finalmente enviado al generador de señal de datos 48 dependiendo del resultado de las pruebas de posibles vectores de disparidad dentro del área de búsqueda

anteriormente mencionada de manera que la fiabilidad se determina dependiendo de una función que:

1. aumenta monótonamente con un error de predicción decreciente, y/o

5 2. aumenta monótonamente con un valor decreciente de una medida de dispersión de la distribución del error de predicción resultante en los vectores de disparidad probados cuando se traza frente a la distancia del vector de disparidad probado respectivo desde el vector escogido realmente.

10 **[0036]** En efecto, la fiabilidad será una medida que indica una probabilidad de que el vector de disparidad insertado dentro de la señal de datos multivista, realmente coincida con la disparidad real, o en cuanto a si el vector de disparidad simplemente corresponde a alguna similitud artificial de las porciones de las tramas sincronizadas en el tiempo de las diferentes vistas. Cabe destacar que la medida de dispersión mantiene su dependencia del error de predicción aun cuando se usen como referencia tramas reconstruidas, y por tanto obtenibles a partir del tren de bits.

15 **[0037]** El reconstructor de predicción 54 puede entonces estar configurado para pasar este valor de fiabilidad junto con el vector de disparidad asociado al generador de señal de datos 48 para ser insertado dentro de la señal de datos multivista.

20 **[0038]** En principio, el reconstructor de predicción 54 puede actuar de la misma manera que se describió anteriormente con respecto al modo de predicción intervista con respecto a subregiones para las cuales ha sido escogido un modo de predicción de compensación de movimiento por el reconstructor predicho 54. Es decir, el reconstructor predictivo 54 puede determinar un vector de movimiento para tales subregiones junto con, opcionalmente, una fiabilidad asociada, con el paso de esta información de predicción al generador de señal de datos 48 para la introducción dentro de la señal de datos multivista.

25 **[0039]** Antes de describir realizaciones para un decodificador de vídeo híbrido adecuado para decodificar la señal de datos multivista producida como salida en la salida 38, cabe destacar que varias características descritas anteriormente con respecto a la fig. 4 son opcionales. Por ejemplo, el error de predicción en los sustractores 44 y 56 no necesita necesariamente ser codificado por transformación. Además, en caso de una codificación sin pérdidas, la
30 cuantificación en los bloques 46 y 58 puede omitirse. Además, el codificador de vídeo híbrido en la fig. 4 codifica predictivamente ambos vídeos 34 y 36. Sin embargo, los bloques 56, 58, 60, 62 y 66 pueden sustituirse por otro motor de codificación para codificar de otro modo el vídeo de segunda vista 36. Como ya se mencionó anteriormente, el filtro de desagrupamiento 52 y 62 es opcional, o puede sustituirse por otro filtro, tal como un filtro de mejora adaptativo. Aunque no se mencionó de manera explícita anteriormente, el generador de señal de datos 48
35 puede estar configurado para codificar los datos recibidos desde los bloques 46, 54 y 58 dentro de la señal de datos multivista mediante codificación de entropía tal como la codificación Huffman o aritmética con el fin de comprimir más los datos. Por último, se observa que más de dos vistas o más de dos vídeos 34 y 36 pueden estar presentes y ser codificados por el codificador de vídeo híbrido de la fig. 4. La extensión de la realización de la fig. 4 sobre más de
40 dos vídeos que corresponden a las diferentes vistas del mismo escenario debería resultar suficientemente clara a partir de la descripción anterior.

[0040] En lo que viene a continuación, se describe una realización para un decodificador de vídeo híbrido con respecto a la fig. 5. El decodificador de vídeo híbrido de la fig. 5 soporta síntesis de vistas intermedias de un vídeo de vista intermedia a partir del vídeo de primera y de segunda vista codificado predictivamente dentro de la señal de
45 datos multivista en la salida 38 del codificador de vídeo híbrido de la fig. 4. Se recuerda brevemente que el codificador de vídeo híbrido o, como responsable por lo tanto, el reconstructor predictivo 54 no asocia necesariamente cada subregión con el modo de predicción intervista. En cambio, la asociación se realiza con el objetivo de optimizar alguna medida de tasa/distorsión y hasta ahora, el modo de predicción intervista compite con la predicción de movimiento compensado y modos de predicción intravista adicionales disponibles. No obstante, los
50 inventores de la presente invención averiguaron que el porcentaje de subregiones - ya sea medidas en número o en área de trama - es suficiente con el fin de aprovechar los vectores de disparidad asociados con estas subregiones para sintetizar un vídeo de vista intermedia a partir de las mismas, es decir, un vídeo que muestra la misma escena que los vídeos de primera y de segunda vista, pero desde otra vista, concretamente una vista aparte de la primera y la segunda vistas, que puede estar colocada localmente entre la primera y la segunda vistas, pero puede estar
55 colocada incluso aún más alejada de una de la primera y la segunda vistas, que la otra de las dos.

[0041] El codificador de vídeo híbrido de la fig. 5 está indicado en general con el signo de referencia 80. Comprende una fase de extracción 82, una fase de reconstrucción predictiva 84 y una fase de sintetización 86. La fase de extracción 82 actúa como un extractor configurado para extraer, de la señal de datos multivista aplicada a

una entrada 88 del decodificador de vídeo híbrido 80, para subregiones de las tramas, con las cuales está asociado el modo de predicción entrevista, un vector de disparidad y un residuo de predicción. La fase de reconstrucción predictiva 84, a su vez, está configurada para reconstruir las subregiones de las tramas con las cuales está asociado el modo de predicción entrevista, generando una predicción a partir de la versión reconstruida de una porción de 5 tramas del vídeo de vista de referencia usando los vectores de disparidad extraídos de la señal de datos multivista para las subregiones respectivas, y el residuo de predicción para las subregiones respectivas. Por último, la fase de sintetización 86 actúa como un sintetizador de vista intermedia configurado para reconstruir las primeras porciones del vídeo de vista intermedia usando la versión reconstruida de las porciones de las tramas del vídeo de vista de referencia, y los vectores de disparidad extraídos de la señal de datos multivista.

10

[0042] El vídeo de vista intermedia así obtenido es producido como salida en una salida 90 del decodificador de vídeo híbrido 80, ya sea solo o junto con los vídeos de primera y de segunda vista representados en la señal de datos multivista que entra en la entrada 88.

15 **[0043]** Para ser más exactos, la fase de extracción 82 comprende un extractor de señal de datos 92, un bloque de reescalado/transformación inversa 94 y un bloque de reescalado/transformación inversa 96. La fase de reconstrucción predictiva 84 comprende sumadores 98 y 100, filtros de desagrupamiento 102 y 104, y un reconstructor predictivo 106. La fase de sintetización 86 comprende un constructor de vista intermedia 108.

20 **[0044]** En efecto, el decodificador de vídeo híbrido 80 comprende una primera parte responsable de reconstruir el vídeo de primera vista, que implica los bloques 92, 94, 98, 102 y 106, y una segunda parte responsable de reconstruir el vídeo de segunda vista, que implica los bloques 92, 96, 100, 104 y 106. Es decir, el extractor de señal de datos 92 y el reconstructor predictivo 106 participan en la reconstrucción de ambos vídeos, los vídeos de primera vista y de segunda vista. En efecto, los componentes 94, 96, 98, 100, 102, 104 y 106 cooperan de una manera que 25 emula el modo de funcionamiento de los componentes 50, 60, 64, 66, 52, 62 y 54 del codificador de vídeo híbrido de la fig. 4. Para ser más exactos, el extractor de señal de datos 92 está configurado para extraer de la señal de datos multivista en la entrada 88, los niveles de coeficiente de transformación cuantificados de las subregiones de las tramas de los vídeos de primera vista y de segunda vista y pasar esta información a los bloques de reescalado/transformación inversa 94 y 96, respectivamente, los cuales a su vez actúan para reconstruir el residuo 30 de predicción respectivo de las subregiones de las tramas del vídeo de primera vista y de segunda vista respectivo. Además, el extractor de señal de datos 92 extrae de la señal de datos multivista en la entrada 88 la información de predicción asociada con cada subregión. Es decir, el extractor de señal de datos 92 recupera de la señal de datos multivista el modo de predicción asociado con cada subregión. Para las subregiones que tienen un modo de predicción entrevista asociado con las mismas, el extractor de señal de datos 92 extrae un vector de disparidad respectivo y, opcionalmente, datos de fiabilidad. Igualmente, el extractor de señal de datos 92 extrae de la señal de 35 datos multivista un vector de movimiento y, opcionalmente, datos de fiabilidad para cada subregión que tiene el modo de predicción de movimiento compensado asociado con la misma. Igualmente, para las subregiones que tienen un modo de intrapredicción asociado con las mismas, el extractor de señal de datos 92 puede recuperar información de intrapredicción a partir de la señal de datos multivista tal como, por ejemplo, una dirección de 40 extensión de contenido de borde principal. El extractor de señal de datos 92 pasa esta información al reconstructor predictivo 106 y el constructor de vista intermedia 108.

[0045] Los componentes 94 a 106 anteriormente mencionados están interconectados entre sí de la manera descrita anteriormente con respecto a los elementos 46, 50 a 54, y 58 a 66 de la fig. 4. La funcionalidad de estos 45 elementos es exactamente la misma. Es decir, el reconstructor predictivo 106 está configurado para generar una predicción para las subregiones de las tramas de ambos vídeos a partir de versiones decodificadas previamente y reconstruidas de porciones de tramas del vídeo usando la información de predicción asociada con las subregiones respectivas. Por ejemplo, las subregiones del modo de predicción entrevista son procesadas correlacionando las posiciones de muestra de las mismas tal como prescriben los vectores de disparidad respectivos y el muestreo, es decir, deduciendo los valores de muestra en las posiciones de muestra correlacionadas – a partir de la trama del otro 50 vídeo en el mismo instante de tiempo en las posiciones de muestra así obtenidas mediante correlación. El muestreo puede implicar una interpolación de posiciones de submuestra dependiendo de la resolución del vector de disparidad. Tal como se indicó anteriormente, la correlación puede implicar o ser un desplazamiento de traslación en una dirección, y en una cantidad, prescritas por el vector de disparidad. Se aplica lo mismo a subregiones de modo 55 de predicción de movimiento compensado, excepto para la trama de referencia, donde tiene lugar el muestreo o la interpolación, que es una trama reconstruida decodificada previamente del mismo vídeo de vista.

[0046] La predicción así obtenida se suma al residuo de predicción reconstruido producido como salida por los bloques 94 y 96, con el resultado de que la suma representa una reconstrucción de la subregión respectiva que,

opcionalmente, tal como se indicó anteriormente, puede mejorarse desagrupando el filtro 102 y 104.

[0047] Hasta ahora, la descripción se ha centrado en el modo de funcionamiento del decodificador de vídeo híbrido de la fig. 5, que concierne a la construcción de las vistas primarias con las cuales están relacionadas realmente los vídeos que han sido codificados en la señal de datos multivista. Sin embargo, tal como se indicó anteriormente, el decodificador de vídeo híbrido de la fig. 5 es particularmente adecuado para sintetizar un vídeo de vista intermedia a partir del vídeo de primera y de segunda vista transmitido explícitamente por medio de la señal de datos multivista. Esta sintetización es realizada por el constructor de vista intermedia 108 que recibe en su entrada las tramas reconstruidas de ambos vídeos transmitidos explícitamente así como la información de predicción, de al menos un subconjunto apropiado de las subregiones, concretamente al menos la información de predicción de las subregiones que tienen el modo de predicción entrevista asociado con las mismas. Tal como se describirá con más detalle más adelante, el constructor de vista intermedia 108 puede, sin embargo, aprovechar también otra información de predicción de otros modos de predicción y opcionalmente datos de fiabilidad.

[0048] En términos generales, el constructor de vista intermedia 108 está configurado para reconstruir porciones del vídeo de vista intermedia usando la versión reconstruida de porciones de las tramas de cualquiera de los vídeos de vista primaria transmitidos explícitamente dentro de la señal de datos multivista, y los vectores de disparidad extraídos de la señal de datos multivista. Se recuerda al lector acerca de la declaración presentada anteriormente según la cual, lo más probable, simplemente un subconjunto apropiado de las subregiones de las tramas del vídeo de primera y de segunda vista son del modo de predicción entrevista. Las otras subregiones no tienen vectores de disparidad asociados con las mismas. No obstante, los inventores de la presente invención averiguaron que debido a la similitud entre tramas de vistas vecinas, el porcentaje de subregiones del modo de predicción entrevista es suficiente con el fin de sintetizar un vídeo de vista intermedia a partir de los vídeos de primera y de segunda vista, evitando así la necesidad de proporcionar a la señal de datos multivista datos de profundidad por píxel/disparidad patentados.

[0049] Haciendo referencia a la figura 6, se describen diferentes posibilidades sobre cómo el constructor de vista intermedia sintetiza las tramas del vídeo de vista intermedia. La fig. 6 muestra cuatro tramas 120_1 a 120_4 del vídeo de primera vista que se suceden inmediatamente en el tiempo una a otra en el lado izquierdo, cuatro tramas consecutivas 122_1 a 122_4 del vídeo de segunda vista de las mismas instancias de tiempo que las tramas 120_1 a 120_4 , en el lado derecho, y cuatro tramas que han de ser sintetizadas 124_1 a 124_4 en el medio entre las tramas del primer vídeo y las tramas del segundo vídeo también de las mismas instancias de tiempo.

[0050] En aras de una mejor comprensión, se muestra que cada uno de las tramas 120_1 a 120_4 y 122_1 a 122_4 está subdividida en una matriz regular de 3×3 subregiones. Con el fin de mantener el número de signos de referencia a un nivel razonable, sólo se indica explícitamente una subregión por vídeo, concretamente la subregión 130_{111} , la subregión 132_{232} , y la subregión 134_{433} . Entre los índices, el primer dígito indica el número de trama, el segundo dígito indica el número de columna de la subregión respectiva (de izquierda a derecha), y el tercer dígito indica el número de fila de la subregión respectiva (de arriba a abajo), en donde el último dígito del propio número de referencia corresponde al último dígito respectivo de los signos de referencia que indican las tramas de los vídeos respectivos, es decir, "0" para el vídeo de primera vista, "2" para el vídeo de segunda vista, y "4" para el vídeo de vista intermedia.

[0051] A continuación, la fig. 6 muestra vectores de movimiento y de disparidad. Los vectores de movimiento se muestran apuntando desde el medio de la subregión asociada hasta una trama del mismo vídeo. Es decir, los vectores de movimiento generalmente se extienden verticalmente. Los vectores de disparidad apuntan desde el medio de una cierta subregión hasta una trama del otro de los vídeos de primera y de segunda vista. Es decir, los vectores de disparidad apuntan a la trama precedente inmediata del mismo vídeo. Sin embargo, como resulta evidente para un experto en la materia, no es necesario que la trama de referencia, a la cual apuntan los vectores de movimiento, esté restringida a la trama inmediatamente precedente. En cambio, la trama de referencia a la cual se refiere el vector de movimiento puede estar separada de la trama actual por otras tramas en el tiempo. Además, la trama de referencia puede ser incluso una trama subsiguiente a la trama actual en el tiempo de reproducción, pero precedente en el tiempo de decodificación, tal como en las tramas B. Además, puede usarse predicción multihipótesis. Es decir, más de un vector de movimiento puede estar asociado con una subregión, con la predicción resultante de las predicciones parciales de movimiento compensado siendo, por ejemplo, una suma ponderada de las predicciones parciales. Se aplica lo mismo a los vectores de disparidad. Por ejemplo, se muestra que los vectores de disparidad de la fig. 6 apuntan desde subregiones del vídeo de segunda vista hasta las tramas 120_1 a 120_4 del vídeo de primera vista solamente. Aunque sería posible restringir los modos de predicción entrevista a una cierta o ciertas vistas transmitidas explícitamente dentro de la señal de datos multivista, también es posible que las

subregiones del modo de predicción entrevista estén presentes en ambos vídeos, el vídeo de primera vista y el vídeo de segunda vista. Esto se ilustra mediante la línea de puntos que apunta desde la trama 120₂ hasta la 122₂. Además, como también se indicó anteriormente, dentro de la señal de datos multivista puede estar presente más de una vista principal transmitida explícitamente, y por consiguiente, también puede usarse predicción multihipótesis con respecto a las subregiones del modo de predicción entrevista porque más de un vector de disparidad está asociado con una cierta subregión, estando las tramas de referencia de estos vectores de disparidad colocadas en la misma instancia de tiempo pero dentro de vídeos multivista diferentes. Por último, se observa que por medio de la multihipótesis, los modos de predicción de movimiento y los modos de predicción entrevista podrían mezclarse de manera que al menos un vector de disparidad y al menos un vector de movimiento está asociado a la misma subregión. En este caso, el modo respectivo asociado a esa subregión estará subsumido bajo el término “modo de predicción entrevista” ya que la subregión tiene un vector de disparidad asociado a la misma.

[0052] Después de haber descrito los elementos de la fig. 6, la funcionalidad del constructor de vista intermedia de la fig. 5 con respecto a diversas realizaciones se describe más adelante con respecto a la fig. 6. Como ya se expuso anteriormente, el constructor de vista intermedia 108 usa los vectores de disparidad de subregiones del modo de predicción entrevista con el fin de reconstruir las primeras porciones del vídeo de vista intermedia. Se dispone de diferentes posibilidades para el constructor de vista intermedia con el fin de reconstruir estas primeras porciones que están indicadas como áreas blancas dentro de las tramas 124₁ a 124₄ en la fig. 6. Por ejemplo, el constructor de vista intermedia 108 puede reconstruir las primeras porciones del vídeo de vista intermedia de una manera muy similar al procesamiento de subregiones de modo de predicción entrevista mediante el reconstructor predictivo 106. Es decir, el constructor de vista intermedia 108 puede correlacionar las posiciones de muestra de la subregión de un modo de predicción entrevista tal como es prescrito por el vector de disparidad asociado y determinar los valores de muestra en estas posiciones correlacionadas dentro de la trama de la misma instancia de tiempo de la vista vecina. En lugar de usar el vector de disparidad tal como es, sin embargo, el constructor de vista intermedia 108 cambia la escala del vector de acuerdo con la ubicación espacial del parámetro de posición de vista intermedia que corresponde al vídeo de vista intermedia en relación con la posición de vista de los vídeos de primera y de segunda vista. Por ejemplo, según la fig. 6, la subregión 132₁₁₁, es decir, la subregión superior izquierda de la trama 122₁, es una subregión del modo de predicción entrevista. Por consiguiente, el constructor de vista intermedia 108 correlaciona las posiciones de muestra de la subregión según un vector de disparidad a escala, concretamente el vector de disparidad asociado con esa subregión a escala con un valor de escala determinado del modo descrito, por ejemplo, con respecto a la fig. 3. En las posiciones de muestra así obtenidas, el constructor de vista intermedia 108 muestrea la versión reconstruida de la trama 120₁ con el fin de lograr una reconstrucción 136 de una primera porción de la trama 124₁ del vídeo de vista intermedia.

[0053] Alternativamente y/o adicionalmente, el constructor de vista intermedia puede intentar aplicar el vector de disparidad en el sentido inverso. Es decir, el constructor de vista intermedia 108 puede correlacionar las posiciones muestreadas de la subregión actual 132₁₁₁ en una dirección prescrita por el vector de disparidad pero apuntando en la dirección opuesta y cambiando la escala por 1-. Entonces, el constructor de vista intermedia 108 puede muestrear (o interpolar) los valores de muestra reconstruidos de la trama actual, es decir, la trama 122₁ con el fin de obtener una reconstrucción de la primera porción 136 del vídeo de vista intermedia. En el último caso, el constructor de vista intermedia 108 puede restringir la reconstrucción a un área donde las posiciones de muestra correlacionadas y la subregión actual 132₁₁₁ se superponen. Como ya se indicó anteriormente, el constructor de vista intermedia 108 puede combinar ambas posibilidades con el fin de aumentar el área de la porción 136 obtenida aprovechando el vector de disparidad de la subregión actual 132₁₁₁.

[0054] Tratando todas las subregiones del modo de predicción entrevista tal como se explicó representativamente con respecto a la subregión 132₁₁₁, el constructor de vídeos intermedios 108 obtiene las áreas no sombreadas dentro de la trama 124₁ a 124₄ del vídeo de vista intermedia. Obviamente, todavía hay porciones restantes 138 que han de ser rellenadas.

[0055] Con el fin de rellenar estas porciones restantes 138, el constructor de vista intermedia 108 puede estar configurado para finalizar una reconstrucción del vídeo de vista intermedia por grupo de imágenes (GOP), en lugar de por trama, en donde el grupo puede ser cualquier grupo de imágenes consecutivas temporalmente sin tener en cuenta si los grupos se pueden decodificar independientemente unos de otros, es decir, ninguna imagen de un grupo es imagen de referencia para una imagen/trama de otro grupo, aunque el último caso puede ser ventajoso. Es decir, el constructor de vista intermedia 108 puede almacenar en memoria intermedia las primeras porciones 136 de un grupo de un cierto número de tramas consecutivas del vídeo de vista intermedia, tras haber reconstruido el número máximo de porciones reconstruibles mediante el uso de los vectores de disparidad dentro de la señal de datos multivista, usando estas primeras porciones 136 como puntos de apoyo y extrapolando y/o

interpolando las porciones restantes 138 del vídeo de vista intermedia, siendo posible la extrapolación y/o interpolación en una dirección temporal y/o espacial. Sin embargo, el constructor de vídeos intermedios 108 además puede, antes o en lugar de la extrapolación y/o interpolación, aprovechar la información adicional transmitida dentro de la señal de datos multivista con el fin de rellenar porciones adicionales de la porción restante 138.

5

[0056] Por ejemplo, el constructor de vista intermedia puede estar configurado para identificar porciones espaciales de la escena donde los vectores de disparidad disponibles pueden ser interpolados temporalmente y/o espacialmente – concretamente entre vistas. Por ejemplo, las dos subregiones 132₂₂₂ y 132₄₂₂ tienen vectores de disparidad asociados con las mismas. El constructor de vista intermedia 108 puede decidir interpolar temporalmente entre los vectores de disparidad asociados con estas subregiones para obtener un vector de disparidad interpolado temporalmente 140 para reconstruir la porción restante 142 entre las porciones restantes 138 del vídeo de vista intermedia. El constructor de vista intermedia 108 puede aplicar este vector de disparidad interpolado temporalmente 140 de la misma manera a las porciones reconstruidas de las tramas 120₃ y 122₃ de los vídeos de primera y de segunda vista tal como se describió anteriormente con respecto a los vectores de disparidad originales transmitidos explícitamente dentro de la señal de datos multivista.

[0057] Igualmente, el constructor de vista intermedia 108 puede interpolar espacialmente entre los vectores de disparidad asociados con las subregiones 132₁₁₁ y 132₂₃₁ con el fin de obtener un vector de disparidad interpolado espacialmente para la porción 144 entre las porciones restantes 138 del vídeo de vista intermedia y aplicar este vector de disparidad interpolado espacialmente de la misma manera que se describió anteriormente con respecto a los otros vectores de disparidad.

[0058] Como ya se ha indicado anteriormente, la señal de datos multivista puede contener datos de fiabilidad para cada vector de disparidad y/o vector de movimiento con el fin de señalar al decodificador de vídeo híbrido la probabilidad de que el vector de disparidad transmitido dentro de la señal de datos multivista coincida realmente con la disparidad real entre las vistas, en lugar de representar sólo una similitud accidental entre porciones de vistas vecinas transmitidas dentro de la señal de datos multivista. En este sentido, el constructor de vista intermedia 108 puede estar configurado para excluir los vectores de disparidad extraídos de la señal de datos multivista para subregiones procedentes de la reconstrucción de porciones del vídeo de vista intermedia para las cuales los datos de fiabilidad extraídos de la señal multidatos no cumplen un requisito mínimo predeterminado, es decir, su fiabilidad es demasiado baja.

[0059] Los datos de fiabilidad adicionales pueden estar ocultos en la señal de datos multivista dentro de porciones de la misma que son ignoradas por los decodificadores de vídeo híbrido que no pueden realizar la síntesis de vistas intermedias recién descrita. Por último, se observa que, por ejemplo, ocultas en la porción recién mencionada de la señal de datos multivista transparente para los decodificadores de vídeo híbrido convencionales, podrían estar contenidas señales residuales para un conjunto predeterminado de vistas intermedias predeterminadas. En este caso, el constructor de vista intermedia 108 podría estar configurado para, cuando se sintetiza un vídeo de vista intermedia en tal vista intermedia predeterminada colocada entre los vídeos de primera vista y de segunda vista, en primer lugar, realizar la reconstrucción de las tramas del vídeo de vista intermedia, tal como se acaba de describir anteriormente, aplicando entonces la señal residual tal como se obtiene de la señal de datos multivista por medio del extractor de señal de datos 92, a las tramas reconstruidas del vídeo de vista intermedia. Aunque aumentarían los datos que han de ser transferidos por medio de la señal de datos multivista, también aumentaría la calidad de la síntesis de vistas intermedias. Por otra parte, se evitaría una transmisión de datos de disparidad adicionales para estas vistas intermedias predeterminadas. Sólo la señal residual para las tramas de estos vídeos de vista intermedia en las vistas intermedias predeterminadas tendría que transmitirse dentro de la señal de datos multivista, finalmente junto con los datos de fiabilidad anteriormente mencionados.

[0060] Por último, se observa que el constructor de vista intermedia además podría estar configurado para crear tramas en instancias de tiempo intermedias de una manera similar a la descrita anteriormente, con el uso de los vectores de movimiento en lugar de los vectores de disparidad, sin embargo. Por ejemplo, las posiciones de una subregión del modo de predicción de movimiento compensado son correlacionadas según un vector de movimiento a escala de acuerdo con la ubicación de la trama intermedia en relación con las tramas inmediatamente vecinas en el tiempo, con el fin de obtener una porción de estas tramas de tiempo intermedias de los vídeos de primera y de segunda vista. Las tramas del vídeo de vista intermedia en las posiciones de tiempo intermedias podrían obtenerse por medio, por ejemplo, de extrapolación/interpolación en el tiempo de la dirección de vista, o determinando vectores de movimiento del vídeo de vista intermedia mediante interpolación intervista de los vectores de movimiento del vídeo de primera y de segunda vista sólo por mencionar unas pocas posibilidades.

[0061] En otras palabras, las realizaciones descritas anteriormente con respecto a las figs. 4 a 6 permiten la extracción de nuevas vistas usando vectores de disparidad a escala en el procedimiento de decodificación. Todas las vistas requeridas, por ejemplo para una visualización de N vistas, pueden extraerse en paralelo usando vectores de disparidad a escala diferente y posiblemente información de fiabilidad. La estructura de un posible decodificador para el cual se mostró una realización en la fig. 5 puede describirse, en otras palabras, de la siguiente manera. En particular, el mismo puede estar basado en estructuras de decodificación de vídeo "clásicas", con la incorporación, sin embargo, de elementos de tecnología adicional. El decodificador puede recibir un tren de bits comprimido que contiene una representación codificada de dos o más secuencias de vídeo. Puede aplicarse una decodificación de entropía al tren de bits para extraer los coeficientes de transformación cuantificados, datos de movimiento/disparidad, datos de fiabilidad y otros datos de control. En la realización de la fig. 5, esta decodificación de entropía puede realizarse dentro del extractor de señal de datos 92.

[0062] Los coeficientes de transformación cuantificados pueden entonces cambiarse de escala y puede aplicarse una transformación inversa para obtener la diferencia de señal residual de predicción. Esto puede realizarse dentro de los bloques 94 y 96 anteriormente mencionados. A este residuo de predicción se añaden datos decodificados previamente procedentes de vistas temporales o vecinas. El tipo de información que ha de añadirse puede controlarse mediante datos de control especiales: en el caso de datos de vídeo intracodificados, no se dispone de información previa o vecina, de manera que se aplica la reconstrucción intratrama. Para datos de vídeo intercodificados, se dispone de datos decodificados previamente procedentes de todas las vistas vecinas precedentes temporalmente. Los datos decodificados previamente son desplazados por los vectores de movimiento asociados y el procedimiento de compensación de movimiento y se añaden a las diferentes señales. Los datos decodificados previamente pertenecen a una vista vecina; los datos de movimiento representan los datos de disparidad.

[0063] Para la síntesis de vistas intermedias, pueden usarse los mismos datos de disparidad con el fin de sintetizar una, o en paralelo, varias vistas intermedias. Por ejemplo, los vectores de disparidad pueden cambiarse de escala por uno o más factores de escala n , $n = 1...N$. Mediante esta medida pueden obtenerse n versiones a escala diferente de los vectores de disparidad y usarse para desplazar los datos decodificados previamente en consecuencia para reconstruir las porciones respectivas 136 de la una o más vistas intermedias.

[0064] En caso de que se transmitan datos residuales para las vistas intermedias, las n versiones desplazadas pueden añadirse a estos datos diferentes transmitidos adicionalmente para generar un número de tramas reconstruidas adicionales de vistas en paralelo. Como sólo unos pocos bloques de una imagen pueden haber sido predichos a partir de vistas vecinas y de este modo tienen vectores de disparidad, mientras que la mayoría pueden haber sido predichos a partir de tramas previas, la reconstrucción inicial puede ser bastante escasa. Además, los vectores de disparidad pueden ser poco fiables y diferentes de los valores reales, ya que los vectores de disparidad fueron estimados inicialmente en el codificador basándose en la minimización de tasa global. Por lo tanto, los datos de fiabilidad pueden estar incluidos dentro del flujo de bits o la señal de datos multivista, respectivamente. Estos datos pueden usarse en la fase de mejora de vista o de datos de síntesis para juzgar la fiabilidad de los vectores de disparidad disponibles y posiblemente omitir o refinar los datos más poco fiables. Estas tramas o vistas iniciales puede ser procesadas además mediante filtros de desagrupamiento o posiblemente procedimientos de mejora, tales como relleno de huecos, suavizado de bordes, etc., para mejorar la calidad visual. Esto puede realizarse en los filtros de desagrupamiento 102 y 104 anteriormente mencionado, por ejemplo. También, los datos escasos pueden rellenarse como se describió anteriormente, tal como mediante datos precedentes temporalmente o procedimientos avanzados tales como estimación de disparidad o profundidad para obtener los vectores de disparidad que faltan y rellenar las vistas intermedias inicialmente escasas. Puesto que se dispone de N tramas o vistas a la vez, estos procedimientos de mejora pueden usar información disponible de todas ellas. Un ejemplo es la síntesis de vistas: si ha de sintetizarse una nueva vista entre dos vistas originales 1 y 2, tal como se muestra en la fig. 3, los datos de la vista 1 pueden desplazarse primero a la nueva posición. La desviación de disparidad, sin embargo, es diferente para los objetos de primer plano y de fondo, ya que la desviación es inversamente proporcional a la profundidad de escena original (distancia frontal desde la cámara). Por lo tanto, nuevas áreas de fondo se hacen visibles en la vista sintetizada que no eran visibles en la vista 1. Aquí, la vista 2 puede usarse para rellenar esta información. Además, puede usarse datos de vecindad espacial, por ejemplo, información de fondo adyacente. Después de esta fase de mejora, los datos reconstruidos pueden ser transferidos a una memoria intermedia de salida. Tal memoria intermedia de salida puede estar colocada en la salida 90. Esta memoria intermedia puede ordenar los datos codificados y producir como salida las imágenes decodificadas en el orden temporal correcto con un número correcto de vistas sintetizadas para cada instancia de tiempo. Para lo último, el número requerido de vistas puede señalizarse por medio de un control de extracción de vistas intermedias. Además, los datos de partida también pueden usarse para el siguiente ciclo de procesamiento para que sirvan como entrada a la compensación de disparidad escalable.

[0065] El decodificador de vídeo híbrido de la fig. 5 puede usarse así como decodificador de N vistas tal como se muestra en la fig. 7. Aquí, el tren de bits que contiene dos vistas v_{BS1} y v_{BS2} , han de ser decodificadas por el decodificador de N vistas para una visualización de 9 vistas con las vistas v_{D1} , v_{D2} , ..., v_{D9} , con el decodificador de 9 vistas estando implementado, por ejemplo, como se describió anteriormente con respecto a la fig. 5. La visualización señala el número de posiciones espaciales o vistas por medio del control de extracción de vistas intermedias. Aquí, se requieren nueve vistas con una distancia espacial de 0,25 cuando se mide en unidades de la distancia entrevista entre las vistas v_{BS1} y v_{BS2} , de manera que las vistas de visualización vecinas tales como v_{D2} y v_{D3} están 4 veces más cerca entre sí en cuanto a la posición espacial y la percepción estereoscópica que las vistas en el tren de bits. Por lo tanto, el constructor de vista intermedia puede usar un conjunto de factores de extracción de vistas $\{k_1, k_2, \dots, k_9\}$ que está establecido en $\{-0,5, -0,25, 0, 0,25, 0,5, 0,75, 1, 1,25, 1,5\}$ con el fin de cambiar la escala de los vectores de disparidad y reconstruir las tramas de vídeos de vista intermedia a partir de ello. Tal como se muestra, las vistas de tren de bits v_{BS1} y v_{BS2} coinciden en su posición espacial con las vistas de visualización v_{D3} y v_{D7} ya que $k_3 = 0$ y $k_7 = 1$. Además, los factores v_{D3} , v_{D4} y v_{D5} son interpolados entre v_{BS1} y v_{BS2} . Finalmente, v_{D1} y v_{D2} así como v_{D8} y v_{D9} son extrapolados a cada lado del par del tren de bits v_{BS1} , v_{BS2} . Con el conjunto de factores de extracción de vistas, los vectores de disparidad entre v_{BS1} y v_{BS2} son cambiados de escala en consecuencia para obtener 9 versiones desplazadas de manera diferente de los datos decodificados previamente. Junto con los nuevos datos de diferencia procedentes del bloque de escalado/transformación inversa en el decodificador, se proporcionan 9 vistas nuevas para el bloque de desagrupamiento /postprocesamiento. Las 9 vistas finales se proporcionan entonces a la visualización por medio de la memoria intermedia de imágenes de salida.

[0066] Por lo tanto, las realizaciones anteriores describen un procedimiento para extracción de vistas a partir de datos de vídeo de dominio comprimido, donde las vistas extraídas pueden ser diferentes de las vistas contenidas en la representación comprimida o el tren de datos en cuanto al número de vistas y la posición espacial. A diferencia de los procedimientos del estado de la técnica actual, no se requería de antemano ninguna extracción de geometría, tal como la estimación de profundidad. En cambio, la información contenida en la representación comprimida se usó para la extracción de vistas, tal como los vectores de movimiento y disparidad. Además, el tren de bits también puede contener información de fiabilidad acerca de los vectores de movimiento/disparidad con el fin de juzgar lo adecuado de estos vectores para una correcta síntesis de vistas adicionales. Estos vectores están incluidos en el tren de bits comprimido en los procedimientos de codificación de vídeo clásicos, tales como la codificación de vídeo multivista.

[0067] Por lo tanto, las realizaciones anteriores pueden usarse para generar directamente diferentes vistas en la posición intermedia requerida de una manera eficiente. Por lo tanto, diferentes visualizaciones multivista pueden usar la misma representación comprimida y no se requiere ningún procesamiento externo adicional.

[0068] Aunque algunos aspectos se han descrito en el contexto de un aparato, resulta evidente que estos aspectos también representan una descripción del procedimiento correspondiente, donde un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunas o todas las etapas de procedimiento pueden ejecutarse mediante (o usando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o algunas de las etapas de procedimiento más importantes pueden ser ejecutadas mediante tal aparato.

[0069] La señal de vídeo codificada inventiva puede ser almacenada en un medio de almacenamiento digital o puede ser transmitida en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión por cable tal como Internet.

[0070] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blue-Ray, un CD, una ROM, una PROM, una EEPROM o una memoria FLASH, que tenga señales de control legibles electrónicamente en el mismo, las cuales cooperan (o son capaces de cooperar) con un sistema informático programable de manera que se realiza el procedimiento respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

[0071] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control legibles electrónicamente, las cuales son capaces de cooperar con un sistema informático programable, de manera que se realiza uno de los procedimientos descritos en este documento.

[0072] Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa
5 puede estar almacenado, por ejemplo, en un soporte legible por una máquina.

[0073] Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en este documento, almacenado en un soporte legible por una máquina.

10 **[0074]** En otras palabras, una realización del procedimiento inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en este documento, cuando el programa informático se ejecuta en un ordenador.

[0075] Una realización adicional del procedimiento inventivo es, por lo tanto, un soporte de datos (o un medio de
15 almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los procedimientos descritos en este documento. El soporte de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transitorios.

[0076] Una realización adicional del procedimiento inventivo es, por lo tanto, un tren de datos o una secuencia de
20 señales que representa el programa informático para realizar uno de los procedimientos descritos en este documento. El tren de datos o la secuencia de señales puede estar configurado, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

[0077] Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un
25 dispositivo lógico programable, configurado para o adaptado para realizar uno de los procedimientos descritos en este documento.

[0078] Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los procedimientos descritos en este documento.

30

[0079] En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo una matriz de
puertas programables in situ) para realizar algunas o todas las funcionalidades de los procedimientos descritos en este documento. En algunas realizaciones, una matriz de puertas programables in situ puede cooperar con un
microprocesador con el fin de realizar uno de los procedimientos descritos en este documento. Generalmente, los
35 procedimientos se realizan, con preferencia, mediante algún aparato de hardware.

[0080] Las realizaciones descritas anteriormente son simplemente ilustrativas de los principios de la presente
invención. Se comprende que a otros expertos en la materia les resultarán evidentes modificaciones y variaciones
de las disposiciones y los detalles descritos en este documento. Por lo tanto, la intención es que esté limitada sólo
40 por el ámbito de las reivindicaciones de patente inminente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de este documento.

REIVINDICACIONES

1. Decodificador de vídeo híbrido que soporta síntesis de vistas intermedias de un vídeo de vista intermedia a partir de un vídeo de una primera y de una segunda vista que son codificadas predictivamente en una
5 señal de datos multivista, estando las tramas (122₁ – 122₄) del vídeo de segunda vista subdivididas espacialmente en subregiones (132₂₃₂) y teniendo la señal de datos multivista un modo de predicción de un conjunto de posibles modos de predicción, asociado con cada una de las subregiones, en el que el conjunto de posibles modos de predicción comprende al menos un modo de predicción intervista y un modo de predicción intravista, comprendiendo el decodificador de vídeo híbrido:
- 10 un extractor (82) configurado para extraer respectivamente, de la señal de datos multivista, para las subregiones (132₂₃₂) de las tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista, un vector de disparidad y un residuo de predicción;
- 15 un reconstructor predictivo (84) configurado para reconstruir las subregiones (132₂₃₂) de las tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista, generando una predicción a partir de una versión reconstruida de una porción de tramas (120₁ – 120₄) del vídeo de primera vista usando los vectores de disparidad extraídos de las señales de datos multivista para las subregiones respectivas, y el residuo de predicción para las subregiones respectivas; y
- 20 un sintetizador de vista intermedia (86) configurado para reconstruir las primeras porciones (136) del vídeo de vista intermedia usando la versión reconstruida de las porciones de las tramas (120₁ – 120₄) del vídeo de primera vista,
- caracterizado porque**
- 25 el sintetizador de vista intermedia (86) está configurado para reconstruir las primeras porciones (136) del vídeo de vista intermedia usando los vectores de disparidad extraídos de la señal multivista, y configurado para reconstruir las cuartas porciones (142) del vídeo de vista intermedia aparte de las primeras porciones interpolando temporalmente y/o espacialmente los vectores de disparidad extraídos de la señal de datos multivista para las subregiones de las
30 tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista, para obtener los vectores de disparidad para las subregiones con las cuales está asociado el modo de predicción intravista.
2. Decodificador de vídeo híbrido según la reivindicación 1, en el que el sintetizador de vista intermedia (86) está configurado para usar la reconstrucción de las primeras porciones del vídeo de vista intermedia como
35 puntos de apoyo en la extrapolación y/o interpolación de terceras porciones del vídeo de vista intermedia.
3. Decodificador de vídeo híbrido según cualquiera de las reivindicaciones anteriores, en el que el sintetizador de vista intermedia (86) está configurado para finalizar una reconstrucción del vídeo de vista intermedia por grupo de imágenes.
- 40 4. Decodificador de vídeo híbrido según cualquiera de las reivindicaciones anteriores, en el que el reconstructor predictivo (84) y el sintetizador de vista intermedia (86) están configurados para correlacionar linealmente, dependiendo de los vectores de disparidad, las posiciones de muestra de subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista en el vídeo
45 de primera vista, y muestrear la versión reconstruida de las porciones de las tramas (120₁ -120₄) del vídeo de primera vista en las posiciones de muestra así obtenidas para obtener respectivamente la predicción para las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista, y la reconstrucción de las primeras porciones (136) del vídeo de vista intermedia, con una reducción de la cantidad de correlación lineal del sintetizador de vista intermedia (86) en relación con el
50 reconstructor predictivo (84), que depende de una ubicación espacial de una vista intermedia que corresponde al vídeo de vista intermedia, en relación con una primera vista que corresponde al vídeo de primera vista, y una segunda vista que corresponde al vídeo de segunda vista.
5. Decodificador de vídeo híbrido según cualquiera de las reivindicaciones anteriores, en el que el
55 reconstructor predictivo (84) está configurado para correlacionar linealmente, dependiendo de los vectores de disparidad, las posiciones de muestra de subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista con las cuales está asociado el modo de predicción intervista en el vídeo de primera vista, y muestrear la versión reconstruida de las porciones de las tramas (120₁ -120₄) del vídeo de primera vista en las posiciones de muestra así obtenidas, para obtener la predicción para las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista

con las cuales está asociado el modo de predicción entrevista, y el sintetizador de vista intermedia (86) está configurado para correlacionar linealmente, dependiendo del vector de disparidad, las posiciones de muestra de las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista con las cuales está asociado el modo de predicción entrevista, en una dirección opuesta a los vectores de disparidad, y muestrear la reconstrucción de las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista en las posiciones de muestra así obtenidas, con una reducción de una cantidad de correlación lineal del sintetizador de vista intermedia (86) en relación con el reconstructor predictivo (84), que depende de una ubicación espacial de una vista intermedia que corresponde al vídeo de vista intermedia, en relación con una primera vista que corresponde al vídeo de primera vista, y una segunda vista que corresponde al vídeo de segunda vista, respectivamente.

10

6. Decodificador de vídeo híbrido según cualquiera de las reivindicaciones anteriores, en el que el extractor (82) está configurado para extraer, de la señal de datos multivista, para las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista, también datos de fiabilidad, en el que el sintetizador de vista intermedia (86) está configurado para excluir los vectores de disparidad extraídos de la señal de datos multivista para las subregiones de las tramas (122₁ – 122₄) del vídeo de segunda vista para las cuales los datos de fiabilidad extraídos de la señal de datos multivista no cumplen un requisito mínimo predeterminado, de su utilización en la reconstrucción de las primeras porciones del vídeo de vista intermedia.

7. Procedimiento de decodificación de vídeo híbrido que soporta síntesis de vistas intermedias de un vídeo de vista intermedia de un vídeo de primera vista y de segunda vista que son codificados predictivamente en una señal de datos multivista, estando las tramas (122₁ – 122₄) del vídeo de segunda vista subdivididas espacialmente en subregiones (132₂₃₂) y teniendo la señal de datos multivista un modo de predicción de un conjunto de posibles modos de predicción, asociado con cada una de las subregiones, en el que el conjunto de posibles modos de predicción comprende al menos un modo de predicción entrevista y un modo de predicción intravista, comprendiendo el procedimiento de decodificación de vídeo híbrido:

extraer respectivamente, de la señal de datos multivista, para las subregiones (132₂₃₂) de las tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción entrevista, un vector de disparidad y un residuo de predicción;

30

reconstruir predictivamente las subregiones (132₂₃₂) de las tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción entrevista, generando una predicción a partir de una versión reconstruida de una porción de tramas (120₁ – 120₄) del vídeo de primera vista usando los vectores de disparidad extraídos de las señales de datos multivista para las subregiones respectivas, y el residuo de predicción para las subregiones respectivas; y

35

reconstruir las primeras porciones (136) del vídeo de vista intermedia usando la versión reconstruida de las porciones de las tramas (120₁ – 120₄) del vídeo de primera vista,

40 **caracterizado porque** la etapa de reconstruir las primeras porciones (136) del vídeo de vista intermedia usa los vectores de disparidad extraídos de la señal de datos multivista y **porque** el procedimiento comprende además reconstruir las cuartas porciones (142) del vídeo de vista intermedia aparte de las primeras porciones interpolando temporalmente y/o espacialmente los vectores de disparidad extraídos de la señal de datos multivista para las subregiones de las tramas del vídeo de segunda vista con las cuales está asociado el modo de predicción entrevista, para obtener los vectores de disparidad para las subregiones con las cuales está asociado el modo de predicción intravista.

45

8. Programa informático que tiene un código de programa para realizar, cuando es ejecutado en un ordenador, un procedimiento según la reivindicación 7.

50

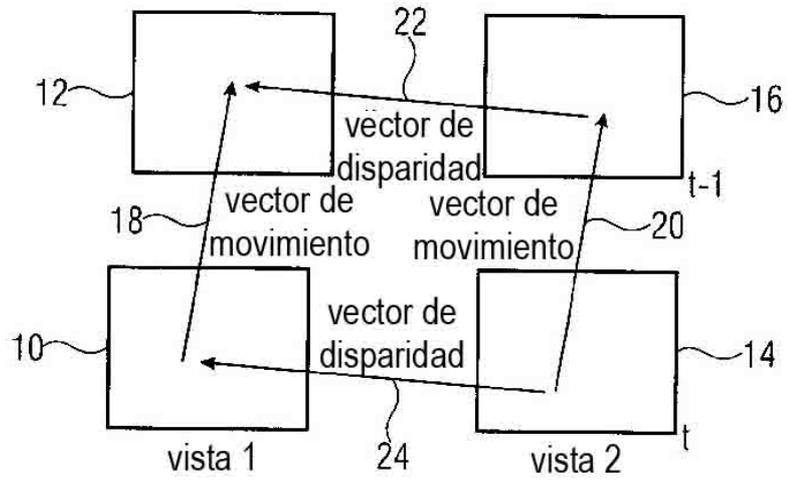


FIGURA 1

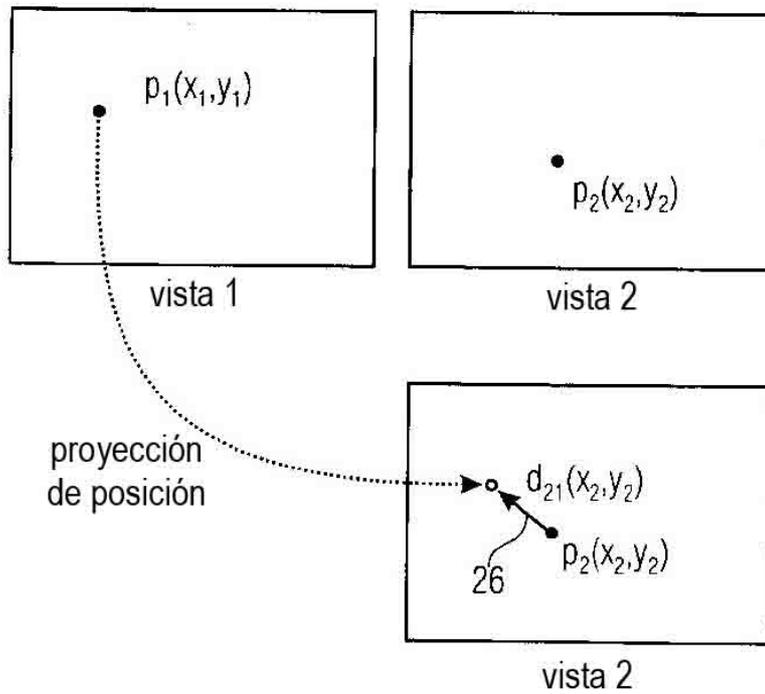


FIGURA 2

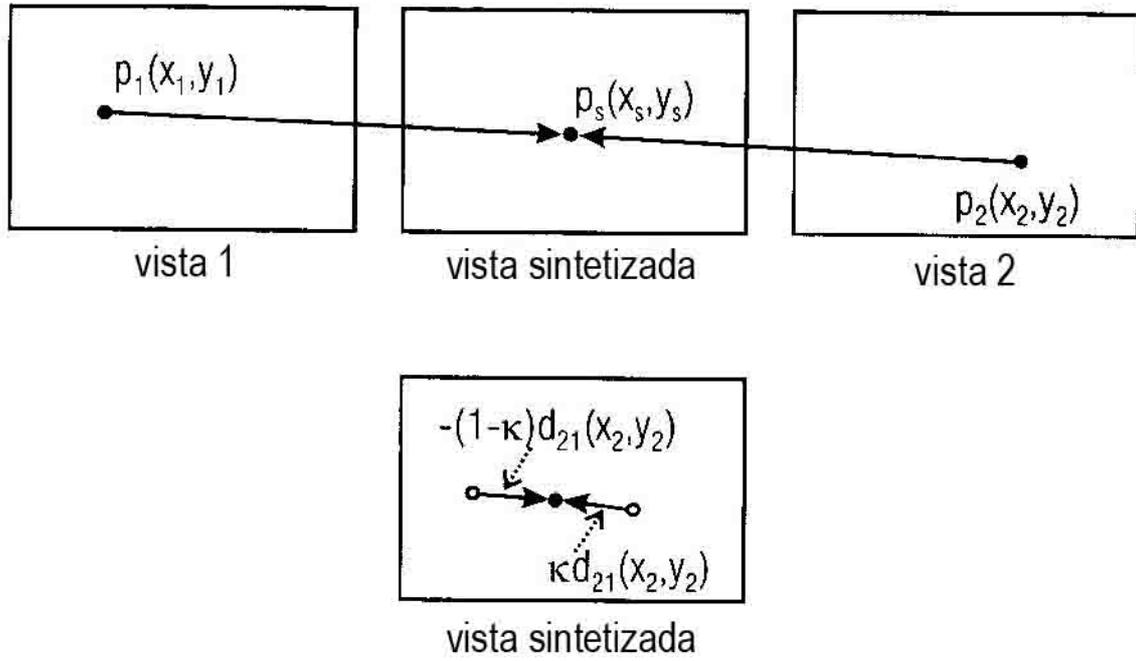


FIGURA 3

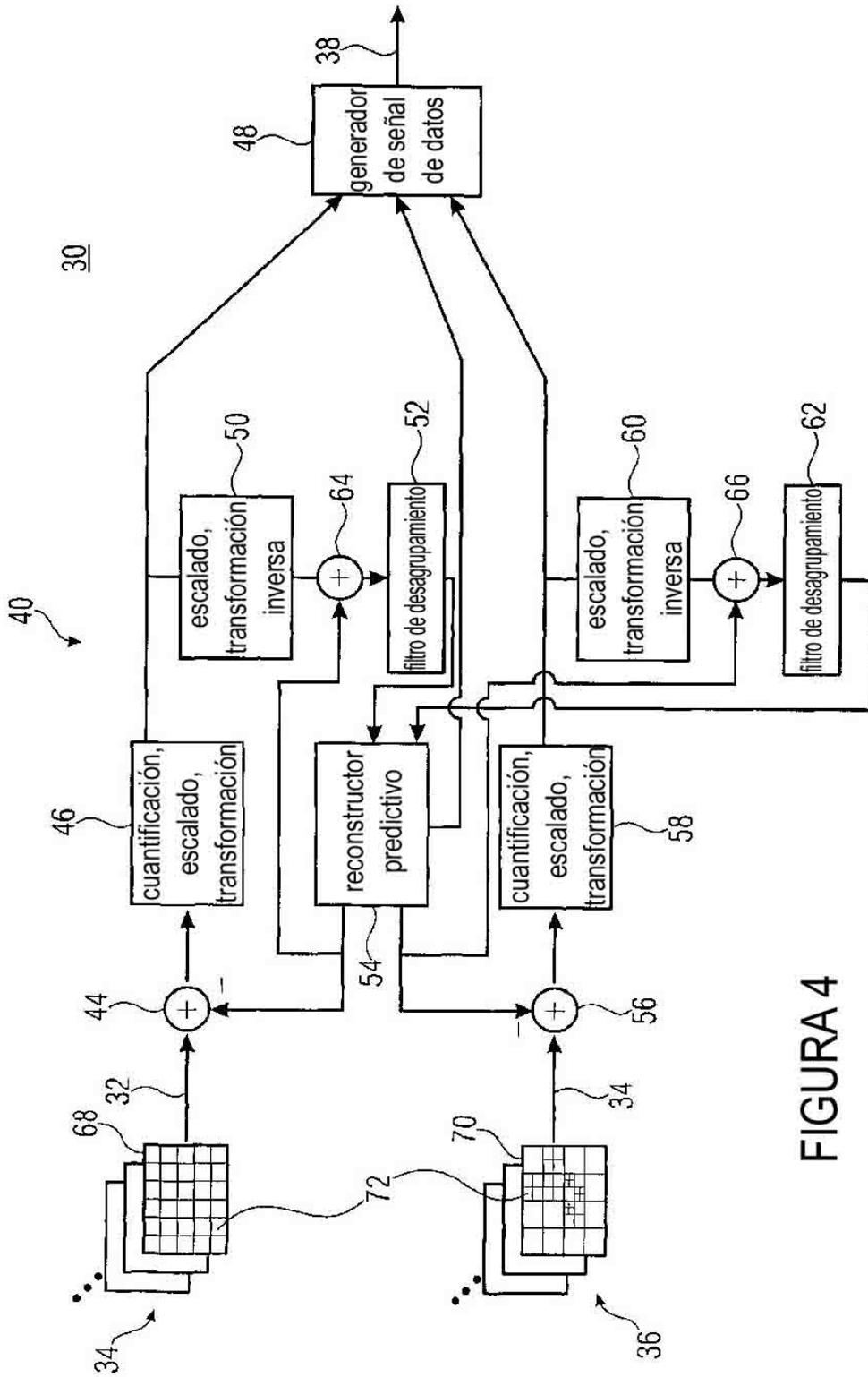


FIGURA 4

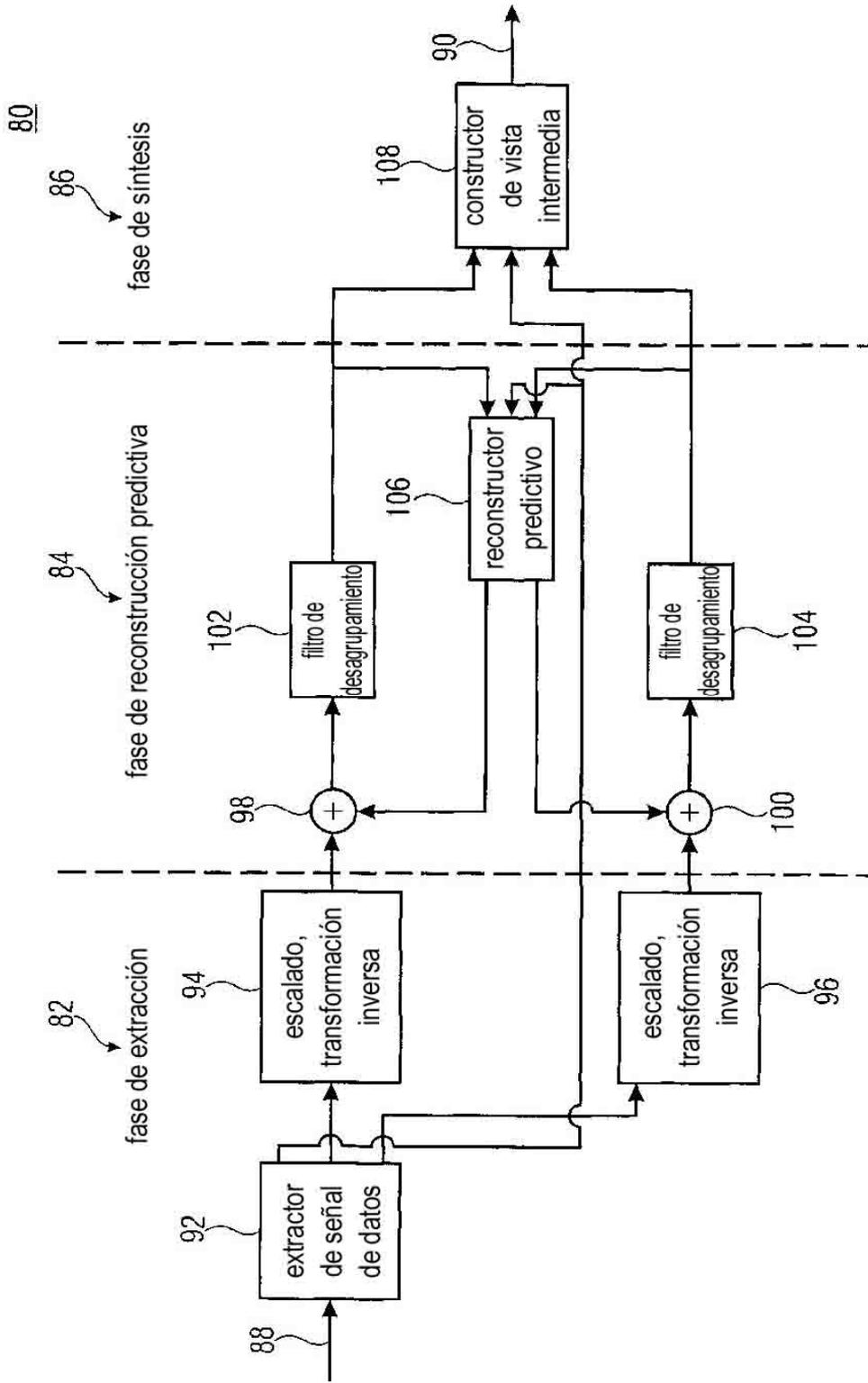


FIGURA 5

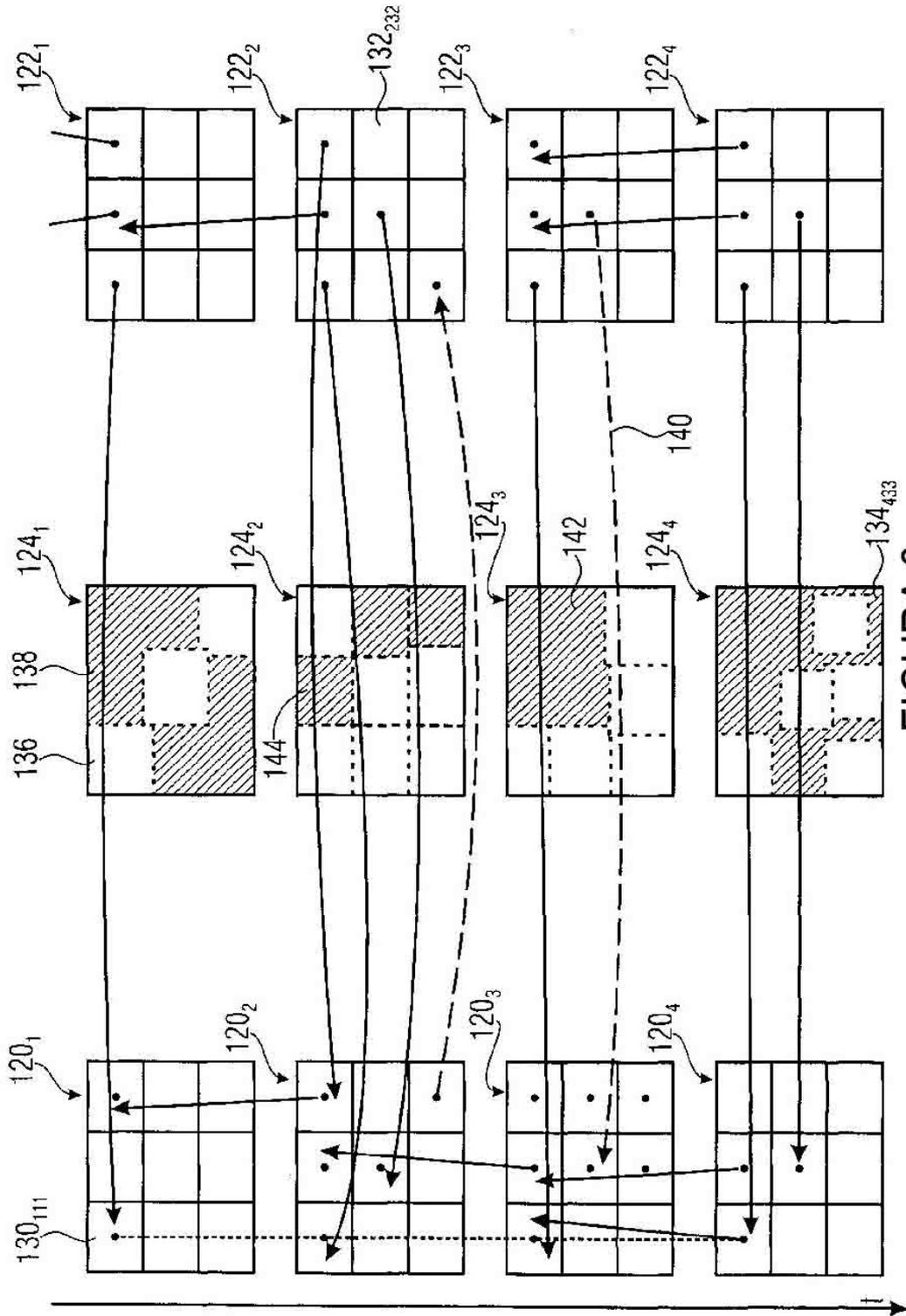


FIGURA 6

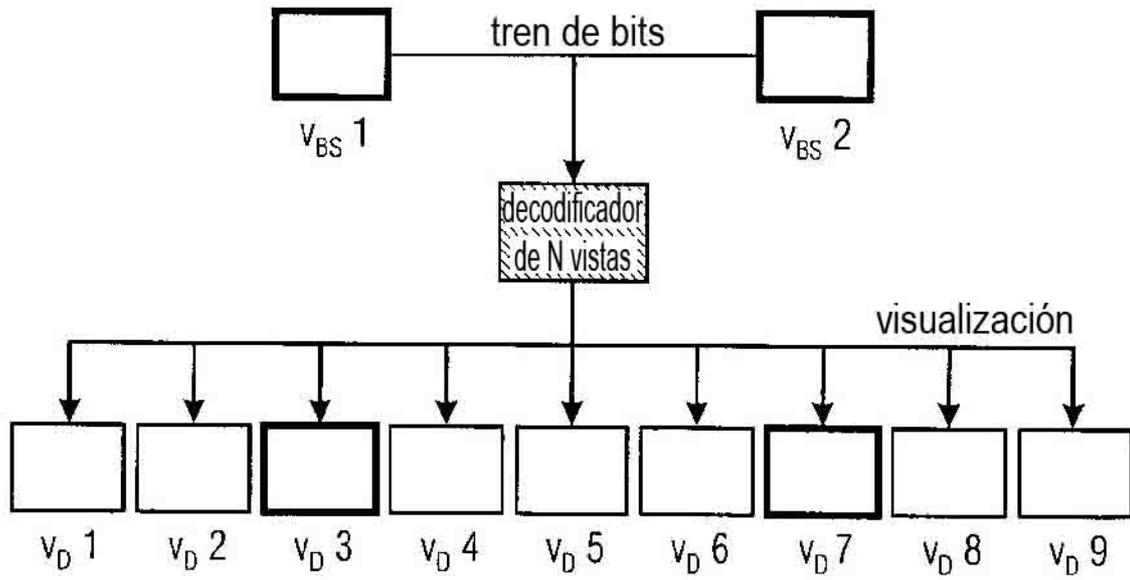


FIGURA 7