(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 564 937**

(51) Int. Cl.:
H04N 13/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **28.07.2011 E 11749928 (5)**

(97) Fecha y número de publicación de la concesión europea: **16.12.2015 EP 2599319**

(54) Título: **Método para combinar imágenes referentes a un contenido tridimensional**

(30) Prioridad:

28.07.2010 IT TO20100652

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
30.03.2016

(73) Titular/es:

S.I.SV.EL. SOCIETA' ITALIANA PER LO SVILUPPO DELL'ELETTRONICA S.P.A. (100.0%)
Via Sestriere 100
10060 None (TO), IT

(72) Inventor/es:

PENNISI, DARIO;
BALLOCCA, GIOVANNI y
D'AMATO, PAOLO

(74) Agente/Representante:

CURELL AGUILÁ, Mireia

ES 2 564 937 T3

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

DESCRIPCIÓN

Método para combinar imágenes referentes a un contenido tridimensional.

Campo técnico

La presente invención se refiere a métodos y a dispositivos para combinar, dentro de un dispositivo de visualización estereoscópica, imágenes generadas localmente superpuestas sobre un contenido tridimensional recibido por el propio dispositivo.

Técnica anterior

Es conocidos que los aparatos de televisión (televisores y descodificadores) pueden generar imágenes localmente que contienen texto y gráficos, y pueden superponerlas sobre las imágenes que se estén recibiendo; por lo tanto es posible suministrar información útil de diversos tipos al usuario a medida que el vídeo se está reproduciendo en segundo plano.

Estas imágenes se pueden generar a partir de información recibida junto con la señal de vídeo (tal como es el caso, por ejemplo, de subtítulos y algunas guías electrónicas de programación, conocidas también como EPG), o pueden aportar información sobre la configuración y los ajustes del descodificador o televisor (por ejemplo, menús o la barra que indica el nivel de volumen y otros parámetros).

En la actualidad, la cantidad de contenido 3D disponible en el mercado ha crecido considerablemente; la proyección de estos contenidos ya no se limita a los cines, y los usuarios pueden ver vídeos 3D en casa en su propio televisor.

Por lo tanto, también para los vídeos 3D existe una necesidad de superponer imágenes generadas localmente sobre las imágenes de televisión que se estén recibiendo.

En comparación con un flujo continuo de vídeo 2D, la superposición de imágenes sobre un flujo continuo de vídeo 3D es más compleja, puesto que es necesario tener en cuenta la diferente disposición de profundidades de los objetos incluidos en las imágenes individuales que constituyen el flujo continuo de vídeo.

La solicitud de patente EP 2 157 803 A1 divulga cómo posicionar un texto en una posición tal que permanezca siempre delante de la imagen de televisión. En particular, si el contenido 3D se radiodifunde como una imagen bidimensional más una matriz de profundidades, esta última también se puede utilizar para definir la posición del texto.

Esta solución tiene el inconveniente de que utiliza un mapa de profundidad muy grande, ya que el mismo se usa para crear el par de imágenes derecha e izquierda (las cuales, cuando se combinan entre sí, producen el efecto 3D) a partir de una imagen bidimensional básica. Además de ser necesario un esfuerzo computacional considerable cuando el mapa debe analizarse para definir la posición del texto, el tamaño de este mapa implica también un uso grande del ancho de banda cuando el mapa se transmite a un receptor.

La solicitud de patente WO 2008/038205 describe un método para componer imágenes 3D (que se visualizarán en un módulo de visualización 3D) compuestas por imágenes 3D principales e imágenes 2D O 3D adicionales, como texto o gráficos, no superpuestas en las principales. Cada tipo de imagen está compuesto por una imagen 2D y un mapa de profundidad relativo. El mapa de profundidad de cada tipo de imagen se combina en la recepción con una imagen 2D correspondiente para reconstruir la imagen a visualizar.

La solicitud de patente EP 1 705 929 describe un método para organizar la transmisión de imágenes 3D combinando en un cuadro una de las dos imágenes 2D que componen la imagen 3D, con información relativa a la profundidad, útil para reconstruir la imagen 3D en la recepción.

Objetivos y breve descripción de la invención

El objetivo de la presente invención es proporcionar un método y un sistema para combinar imágenes con un contenido tridimensional transportado por un flujo continuo de vídeo estereoscópico, que permite superar los inconvenientes de la técnica anterior.

En particular, es un objetivo de la presente invención proporcionar un método para superponer imágenes sobre contenido 3D, que requiere un coste computacional menor en el nivel del reproductor de contenido 3D.

Otro objetivo de la presente invención es proporcionar un método para transmitir información necesaria para superponer imágenes sobre aquellas transportadas por un flujo continuo de vídeo estereoscópico, que no requiere un uso elevado del ancho de banda y que es robusto ante las operaciones de codificación y de descodificación del flujo continuo de vídeo estereoscópico.

Estos y otros objetivos de la presente invención se alcanzan a través de un método y un sistema para superponer imágenes sobre aquellas transportadas por un flujo continuo de vídeo estereoscópico, y que incorporan las características que se exponen en las reivindicaciones adjuntas, las cuales están destinadas a constituir una parte integral de la presente descripción.

La idea general que fundamenta la presente invención es visualizar un elemento superpuesto sobre un flujo continuo estereoscópico utilizando en la fase de reproducción un mapa de profundidades de superposición codificado en una imagen contenida en un cuadro de un flujo continuo estereoscópico. El mapa de profundidades usado en la presente invención no está destinado a codificar el flujo continuo de vídeo estereoscópico, ya que se transmite con la finalidad única de aportar al descodificador o al televisor información útil para superponer de una manera apropiada sobre la imagen estereoscópica imágenes generadas localmente. Con este fin, el mapa de profundidades tiene una menor resolución, y por lo tanto un número más reducido de píxeles, en comparación con el par estereoscópico, con lo cual se limita la ocupación del ancho de banda. Esto es posible debido a que el mapa no se usa para generar la imagen tridimensional, sino únicamente para posicionar de manera correcta las capas superpuestas.

En una forma de realización preferida, el cuadro es portador de una imagen compuesta que comprende una imagen derecha, una imagen izquierda y el mapa de profundidades, multiplexados apropiadamente.

En una forma de realización, las imágenes derecha e izquierda están dispuestas de acuerdo con un formato tradicional, por ejemplo, un formato de lado-con-lado, de arriba-abajo o de tablero de ajedrez, mientras que el mapa de profundidades se inserta en un área libre del cuadro compuesto y no está destinado a visualización.

En una forma de realización alternativa, las imágenes derecha e izquierda están dispuestas de una manera innovadora en el cuadro compuesto. En esta forma de realización, el cuadro comprende un número de píxeles que es mayor que la suma de los píxeles del formato original (es decir, antes de la codificación) de las imágenes tanto derecha como izquierda, las cuales se introducen por tanto sin someterse al diezmado.

En esta forma de realización, los píxeles de la primera imagen (por ejemplo, la imagen izquierda) se introducen en la imagen compuesta sin experimentar ningún cambio, mientras que la segunda imagen se subdivide en regiones cuyos píxeles están dispuestos en áreas libres de la imagen compuesta. Esta solución ofrece la ventaja de que una de las dos imágenes se deja sin variaciones, lo cual da como resultado una mejor calidad de la imagen reconstruida.

De forma ventajosa, a continuación la segunda imagen se descompone en el menor número posible de regiones, para maximizar la correlación espacial entre los píxeles y reducir la generación de artefactos durante la etapa de compresión.

En una forma de realización ventajosa, las regiones de la segunda imagen se introducen en la imagen compuesta por medio solamente de operaciones de traslación o rototraslación, dejando así sin variaciones la relación entre la resolución horizontal y vertical.

En otra forma de realización, por lo menos una de las regiones en las cuales se ha descompuesto la segunda imagen experimenta una etapa de inversión especular, es decir, se le da la vuelta con respecto a un eje (en particular un lateral) y se dispone en la imagen compuesta de tal modo que uno de sus laterales linda con un lateral de la otra imagen que tiene píxeles idénticos o similares en el lateral lindante debido a la fuerte correlación existente entre píxeles homólogos de las dos imágenes derecha e izquierda, es decir, píxeles de las dos imágenes que están posicionados en la misma fila y columna.

Esta solución ofrece la ventaja de reducir la generación de artefactos en el área delimitadora. Más ventajosamente, las regiones en las que se subdivide la segunda imagen tienen una forma rectangular; en comparación con la solución que hace uso de regiones triangulares dispuestas con áreas delimitadoras que cruzan la imagen compuesta en direcciones diagonales, esta opción proporciona una reducción de los artefactos producidos por una etapa de compresión subsiguiente, especialmente si esta última actúa sobre bloques cuadrados de píxeles (por ejemplo, 16x16 para la norma H.264).

De acuerdo con una forma de realización particularmente ventajosa, la formación de artefactos se reduce adicionalmente o incluso se elimina por completo al introducir redundancia en la imagen compuesta, es decir copiando algunos grupos de píxeles varias veces. En particular, esto se logra descomponiendo la imagen básica que se va a introducir en la imagen compuesta, en regiones que presentan unas dimensiones tales que el número total de píxeles de estas regiones supera el número de píxeles de la imagen a descomponer. En otras palabras, la imagen se descompone en regiones, de las cuales al menos dos comprenden una parte de imagen en común. La parte de imagen en común es un área delimitadora entre regiones adyacentes entre sí en la imagen desensamblada. El tamaño de esta parte común depende preferentemente del tipo de compresión que se aplicará posteriormente a la imagen compuesta, y puede actuar como área amortiguadora la cual se eliminará parcial o completamente cuando se reconstruya la imagen desensamblada. Puesto que la compresión puede introducir artefactos en las áreas

delimitadoras de dichas regiones, al eliminar las áreas amortiguadoras, o al menos la parte más externa de las mismas, es posible eliminar cualquier artefacto y reconstruir una imagen que resulte fiel a la original.

Otros objetivos y ventajas de la presente invención se pondrán más claramente de manifiesto a partir de la siguiente descripción de algunas de sus formas de realización, las cuales se aportan a título de ejemplo no limitativo.

Breve descripción de los dibujos

Dichas formas de realización se describirán haciendo referencia a los dibujos adjuntos, en los que:

la figura 1 muestra un diagrama de bloques de un dispositivo para multiplexar la imagen derecha y la imagen izquierda en una imagen compuesta;

la figura 2 es un diagrama de flujo de un método ejecutado por el dispositivo de la figura 1;

la figura 3 muestra una primera forma de desensamblaje de una imagen que se va a introducir en una imagen compuesta;

la figura 4 muestra una primera etapa de construcción de una imagen compuesta de acuerdo con una forma de realización de la presente invención;

la figura 5 muestra la imagen compuesta completa de la figura 4;

la figura 6 muestra una segunda forma de desensamblaje de una imagen que se va a introducir en una imagen compuesta;

la figura 7 muestra una imagen compuesta que incluye la imagen de la figura 6;

la figura 8 muestra una tercera forma de desensamblaje de una imagen a introducir en una imagen compuesta;

la figura 9 muestra una imagen compuesta que incluye la imagen de la figura 8;

la figura 10 muestra un diagrama de bloques de un receptor para recibir una imagen compuesta generada de acuerdo con el método de la presente invención;

la figura 11 muestra algunas etapas de reconstrucción de la imagen desensamblada de acuerdo con el método de la figura 8 e introducida en la imagen compuesta recibida por el receptor de la figura 10;

la figura 12 es un diagrama de flujo de un método para reconstruir las imágenes derecha e izquierda multiplexadas en una imagen compuesta del tipo que se muestra en la figura 9;

la figura 13 muestra una imagen compuesta de acuerdo con una cuarta forma de realización de la presente invención;

las figuras 14a a 14f muestran una imagen derecha y una imagen izquierda en diferentes etapas de procesado llevadas a cabo para su introducción en la imagen compuesta de la figura 13.

Cuando resulta apropiado, las estructuras, componentes, materiales y/o elementos similares se designan por medio de referencias similares en las diferentes figuras.

Descripción detallada de la invención

La figura 1 muestra un diagrama de bloques de un dispositivo 100 para generar un flujo continuo de vídeo estereoscópico 101 con un mapa de profundidad para superponer imágenes sobre un contenido de vídeo transportado por el flujo continuo de vídeo.

A efectos de la presente invención, un contenido tridimensional (o 3D) es una imagen o un vídeo que es percibido por el observador como si tuviera una profundidad variable, en donde pueden sobresalir elementos desde el plano de la pantalla, en el cual se está visualizando o proyectando dicha imagen o vídeo.

La expresión “superponer dos imágenes” se refiere en la presente a cualquier forma de combinación de dos imágenes, por ejemplo en transparencia, semi-transparencia u opacidad completa.

La presente invención se aplica asimismo a cualquier tipo de superposición, ya sea estática o dinámica, es decir que presente características gráficas fijas o variables en el tiempo, las cuales a su vez pueden ser o bien bidimensionales o bien tridimensionales.

La profundidad de un contenido tridimensional se refiere a la dimensión del contenido tridimensional que entra en la pantalla según un eje ortogonal a la pantalla en la cual se está visualizando el contenido 3D. A efectos de la presente descripción, la pantalla se corresponde con un punto de profundidad cero, mientras que el punto de "profundidad mínima" es aquel punto del contenido 3D que es percibido por el usuario como más próximo a sí mismo, es decir el más alejado de la pantalla. Por consiguiente, el punto de "profundidad máxima" es aquel punto que es percibido por el observador como más dentro de la pantalla, es decir, más alejado de sí mismo, incluso más allá del plano de la pantalla.

En la figura 1, en el dispositivo 100 se introducen dos secuencias de imágenes 102 y 103, por ejemplo dos flujos continuos de vídeo, destinados respectivamente al ojo izquierdo (L) y al ojo derecho (R), y una secuencia de mapas de profundidades 106. Cada mapa de profundidades de la secuencia 106 está asociado a un par de imágenes derecha e izquierda pertenecientes a las secuencias 102 y 103, respectivamente. En esta forma de realización, el mapa de profundidades se genera por medio de algoritmos conocidos de por sí los cuales comparan una imagen derecha con una imagen izquierda y devuelven una matriz (el mapa de profundidades) que tiene un tamaño igual a los píxeles de una de las dos imágenes comparadas, y cuyos elementos tienen un valor que es proporcional a la profundidad de cada píxel que se está visualizando. Otra técnica de generación de mapas de profundidades se basa en la medición de la distancia del objeto en la escena desde el par de cámaras de vídeo que están filmando la escena: esta distancia se puede medir fácilmente por medio de un láser. En el caso de flujos continuos de vídeo artificiales generados con ayuda de ordenadores electrónicos, las cámaras de vídeo son virtuales, por cuanto constan de dos perspectivas de una cierta escena creadas artificialmente por un ordenador. En otra forma de realización, un mapa de profundidades está asociado a múltiples pares de imágenes derecha e izquierda; en este caso, el valor seleccionado para cada elemento del mapa de profundidades es el valor de profundidad mínima del píxel en los diferentes cuadros. Preferentemente, en esta forma de realización el mapa de profundidades se introduce una vez para cada grupo de cuadros al que está asociado, con el fin de reducir la carga del dispositivo 100, en el cual se introduce también una información que permite asociar un mapa de profundidades a múltiples pares de imágenes derecha e izquierda.

Como alternativa al ejemplo de la figura 1, los mapas de profundidades de la secuencia 106 se pueden generar dentro del dispositivo 100. En este caso, el dispositivo 100 comprende un módulo adecuado en el cual se introducen las imágenes L y R de las secuencias 102 y 103, y a continuación generan mapas de profundidades correspondientes.

El dispositivo 100 permite implementar un método para multiplexar dos imágenes de las dos secuencias 102 y 103 y el mapa de profundidades de la secuencia 106.

Con el fin de implementar el método para multiplexar las imágenes derecha e izquierda y el mapa de profundidades, el dispositivo 100 comprende un módulo desensamblador 104 para descomponer una imagen de entrada (la imagen derecha del ejemplo de la figura 1) en una pluralidad de sub-imágenes, que se corresponden cada una de ellas con una región de la imagen recibida, un módulo de submuestreo y filtrado 107 para procesar el mapa de profundidades, y un módulo ensamblador 105 con capacidad de introducir los píxeles de imágenes recibidas, incluyendo el mapa de profundidades, en una única imagen compuesta que se proporcionará en su salida. Si no es necesario ningún procesado de la secuencia 106, puede omitirse el módulo 107. Este puede ser el caso en el que, por ejemplo, el mapa de profundidades se genera por láser y tiene, ya desde el principio, una resolución inferior a la de las imágenes L y R.

A continuación se describirá en referencia a la figura 2 un ejemplo de un método de multiplexado implementado por el dispositivo 100.

El método se inicia en la etapa 200. Subsiguientemente (etapa 201), una de las dos imágenes de entrada (derecha o izquierda) se descompone en una pluralidad de regiones, tal como se muestra en la figura 3. En el ejemplo de la figura 3, la imagen desensamblada es un cuadro R de un flujo continuo de vídeo 720p, es decir, un formato progresivo con una resolución de 1280x720 píxeles, 25/30 fps (cuadros por segundo).

El cuadro R de la figura 3 proviene del flujo continuo de vídeo 103 el cual es portador de las imágenes destinadas al ojo derecho, y se desensambla en tres regiones R1, R2 y R3.

El desensamblaje de la imagen R se obtiene dividiéndola en dos partes del mismo tamaño y posteriormente subdividiendo una de estas partes en dos partes del mismo tamaño.

La región R1 tiene un tamaño de 640x720 píxeles y se obtiene tomando la totalidad de los primeros 640 píxeles de cada fila. La región R2 tiene un tamaño de 640x360 píxeles y se obtiene tomando los píxeles desde el 641 al 1280 de las primeras 360 filas. La región R3 tiene un tamaño de 640x360 píxeles y se obtiene tomando los píxeles restantes de la imagen R, es decir los píxeles desde el 641 al 1280 de las últimas 360 filas.

En el ejemplo de la figura 1, la etapa de desensamblaje de la imagen R la lleva a cabo el módulo 104, el cual recibe una imagen de entrada R (en este caso el cuadro R) y da salida a tres sub-imágenes (es decir, tres grupos de píxeles) correspondientes a las tres regiones R1, R2 y R3.

Posteriormente (etapas 202, 203 y 205) se construye la imagen compuesta C, la cual comprende la información perteneciente a las imágenes tanto derecha como izquierda y al mapa de profundidades recibido; en el ejemplo descrito en la presente, dicha imagen compuesta C es un cuadro del flujo continuo de vídeo estereoscópico de salida, y por lo tanto se le hace referencia también como cuadro contenedor.

En primer lugar (etapa 202), la imagen de entrada recibida por el dispositivo 100 y no desensamblada por el dispositivo 105 (la imagen izquierda L en el ejemplo de la figura 1) se introduce sin variaciones en un cuadro contenedor el cual está dimensionado de tal manera que incluye todos los píxeles de las dos imágenes de entrada. Por ejemplo, si las imágenes de entrada tienen un tamaño de 1280x720 píxeles, entonces un cuadro contenedor adecuado para contener las dos será un cuadro de 1920x1080 píxeles, por ejemplo, un cuadro de un flujo continuo de vídeo del tipo 1080p (formato progresivo con 1920x1080 píxeles, 25/30 cuadros por segundo).

En el ejemplo de la figura 4, la imagen izquierda L se introduce en el cuadro contenedor C y se posiciona en la esquina izquierda superior. Esto se obtiene copiando los 1280x720 píxeles de la imagen L en un área C1 compuesta por los primeros 1280 píxeles de las primeras 720 filas del cuadro contenedor C.

Cuando en la siguiente descripción se hace referencia a la introducción de una imagen en un cuadro, o a la transferencia o copia de píxeles de un cuadro a otro, se entiende que esto significa la ejecución de un procedimiento que genera (usando medios de hardware y/o software) un cuadro nuevo que comprende los mismos píxeles que la imagen de origen.

Las técnicas (de software/hardware) para reproducir una imagen de origen (o un grupo de píxeles de una imagen de origen) en una imagen de destino se considera que no tienen importancia a efectos de la presente invención y no se describirán de forma más detallada en la presente, por cuanto las mismas son conocidas de por sí para aquellos versados en la materia.

En la siguiente etapa 203, la imagen desensamblada en la etapa 201 por el módulo 104 se introduce en el cuadro contenedor. Esto se logra por medio de la copia, por parte del módulo 105, de los píxeles de la imagen desensamblada en el cuadro contenedor C en las áreas del mismo que no fueron ocupadas por la imagen L, es decir áreas que son externas al área C1.

Para obtener la mejor compresión posible y reducir la generación de artefactos cuando se descomprime el flujo continuo de vídeo, los píxeles de las subimágenes a los que da salida el módulo 104 se copian conservando las relaciones espaciales respectivas. En otras palabras, las regiones R1, R2 y R3 se copian en áreas respectivas del cuadro C sin experimentar ninguna deformación, exclusivamente por medio de operaciones de traslación y/o rotación.

En la figura 5 se muestra un ejemplo del cuadro contenedor C al que da salida el módulo 105.

La región R1 se copia en los últimos 640 píxeles de las primeras 720 filas (área C2), es decir, junto a la imagen copiada previamente L.

Las regiones R2 y R3 se copian debajo del área C1, es decir respectivamente en las áreas C3 y C4, las cuales comprenden respectivamente los primeros 640 píxeles y los siguientes 640 píxeles de las últimas 360 filas.

Como alternativa a la solución mostrada en la figura 5, las regiones R2 y R3 se pueden copiar en el cuadro contenedor C en áreas desconectadas (es decir, ni solapadas ni vecinas) separadas por un grupo de píxeles, para reducir las regiones delimitadoras.

Las operaciones para introducir las imágenes L y R en el cuadro contenedor no implican ninguna alteración en el equilibrio entre la resolución horizontal y vertical.

En los píxeles libres del cuadro C, es decir en el área C5, el módulo 105 introduce, en forma de una imagen, el mapa de profundidades (DM) perteneciente al par estereoscópico L y R (etapa 205). Antes de la etapa 205, el mapa de profundidades DM se puede submuestrear, filtrar o procesar adicionalmente por parte del módulo 107.

El mapa de profundidades se codifica preferentemente como una imagen en escala de grises, cuyo contenido de información puede ser transportado por lo tanto por la señal de luminancia sola, puesto que las crominancias son nulas, esto permite obtener una compresión eficaz del cuadro contenedor C.

Tal como se muestra en el ejemplo de la figura 5, el mapa de profundidades introducido en el cuadro C es preferentemente un mapa de profundidades de superposición de imágenes, y por lo tanto su resolución es inferior a

la del par L y R, puesto que se transmite con la única finalidad de posicionar las capas superpuestas en profundidad, no para generar el flujo continuo de vídeo estereoscópico. La resolución seleccionada del mapa de profundidades es un resultado de un compromiso entre la velocidad de bits requerida para la transferencia, que debe ser la más baja posible, y la calidad de la información necesaria para posicionar correctamente las capas superpuestas.

En una forma de realización preferida, el mapa de profundidades de superposición DM tiene una resolución de 640 x 360 píxeles, correspondiente a un submuestreo (o diezmado) de 4-a-1 del mapa de profundidades original que tiene una resolución de 1280 x 720 píxeles, igualando la de las imágenes L y R. Cada píxel del mapa submuestreado DM se corresponde con una región de 2 x 2 píxeles del mapa original. En particular, la etapa de submuestreo de 4-a-1 se puede ejecutar seleccionando una fila de cada dos y una columna de cada dos del mapa original.

En otra forma de realización, después del diezmado el mapa de profundidades DM de superposición experimenta una etapa de procesado en la que el mismo se divide en macrobloques de 16x16 píxeles, y a los píxeles pertenecientes al mismo macrobloque se les da un valor de profundidad único. Preferentemente, este valor equivale a la profundidad mínima dentro del macrobloque, puesto que este es el valor más significativo para posicionar correctamente las capas superpuestas.

Alternativamente, este valor es igual al valor de profundidad media dentro del macrobloque.

La elección de macrobloques de 16x16 píxeles es particularmente ventajosa cuando la normativa de compresión en uso es la H.264, puesto que dichos macrobloques coinciden con aquellos utilizados en la normativa H.264. De hecho, con esta solución, la compresión genera menos artefactos y requiere una velocidad de bits menor.

La subdivisión en bloques de 8 x 8 ó 4 x 4 también se puede considerar como ventajosa por cuanto, debido a las características particulares del algoritmo de compresión H.264, se obtienen ventajas de la compresión si los píxeles dentro de estos bloques son todos iguales.

Alternativamente, aunque dejando de lado la subdivisión en bloques o macrobloques dentro de los cuales los píxeles son todos iguales, el mapa de profundidades de 640x360 se puede filtrar con un filtro pasobajo bidimensional. También se obtienen en este caso ventajas en la compresión, puesto que se eliminan o reducen las frecuencias espaciales más altas.

Alternativamente, el mapa de profundidades puede tener una solución de 160 x 90 píxeles, resultante de un submuestreo de 64-a-1 en el que cada píxel del mapa de profundidades DM se corresponde con una región de 8x8 del mapa original.

En otra forma de realización, el mapa de profundidades DM de superposición introducido en el cuadro contenedor C puede tener una resolución irregular; en particular, la mitad o tercio inferior del mapa de profundidades de superposición tiene una resolución mayor que la parte superior. Esta solución resulta ser particularmente ventajosa por lo que respecta al posicionamiento de subtítulos u otra información, tal como el volumen del audio, que generalmente se sitúan en la parte inferior de la imagen. Así, el receptor puede usar información más precisa sobre la profundidad de los píxeles en una región de interés, por ejemplo, el tercio inferior de la imagen 3D, y por lo tanto puede posicionar las imágenes (texto o gráficos) correctamente en esa región. Como mínimo, el mapa de profundidades de superposición puede incluso contener solamente información sobre la profundidad de los píxeles (la totalidad o solamente una parte de ellos) situados en una región de interés, en particular en la mitad inferior o en el tercio inferior del contenido tridimensional.

En otra forma de realización, una región del cuadro contenedor que no está ocupada por las imágenes derecha o izquierda, por partes de las mismas o por el mapa de profundidades de superposición, está destinada a recibir un indicador el cual es necesario para reconstruir las imágenes derecha e izquierda en el nivel del desmultiplexor. Por ejemplo, dicho indicador puede hacer referencia a cómo se ha creado la imagen compuesta. Preferentemente, el indicador puede contener información útil para usar apropiadamente en un mapa de profundidades.

Los píxeles de esta región de indicador están coloreados, por ejemplo, con dos colores (por ejemplo, negro y blanco) con el fin de crear un código de barras de cualquier tipo, por ejemplo lineal o bidimensional, el cual es portador de la información del indicador.

Una vez que se ha completado la transferencia de las dos imágenes y del mapa de profundidades de superposición recibido (y posiblemente también del indicador) hacia el cuadro contenedor, el método implementado por el dispositivo 100 finaliza, y el cuadro contenedor se puede comprimir y transmitir sobre un canal de comunicaciones y/o se puede grabar en un soporte adecuado (por ejemplo, CD, DVD, Blu-ray, memoria masiva, etcétera).

Puesto que las operaciones de multiplexado explicadas anteriormente no alteran las relaciones espaciales entre los píxeles de una región o imagen, el flujo continuo de vídeo al que da salida el dispositivo 100 se puede comprimir a un nivel considerable al mismo tiempo que conservando buenas posibilidades de que la imagen se reconstruya de manera muy fiel a la transmitida sin crear artefactos significativos.

Antes de describir cualquier otra forma de realización, debe indicarse que, en la forma de realización preferida, la división del cuadro R en tres regiones R1, R2 y R3 se corresponde con la división del cuadro en el menor número posible de regiones, teniendo en cuenta el espacio disponible en la imagen compuesta y el espacio ocupado por la imagen izquierda que se introduce sin variaciones en el cuadro contenedor.

Dicho menor número es, en otras palabras, el número mínimo de regiones necesario para ocupar el espacio que queda disponible en el cuadro contenedor C por la imagen izquierda.

Por lo tanto, en general el número mínimo de regiones en el cual se debe desensamblar la imagen se define como una función del formato de las imágenes de origen (imágenes derecha e izquierda) y de la imagen compuesta de destino (cuadro contenedor C).

Preferentemente, la imagen a introducir en el cuadro se desensambla teniendo en cuenta la necesidad de descomponer la imagen (por ejemplo, R en el ejemplo anterior) en el número más pequeño de regiones rectangulares.

En otra forma de realización, la imagen derecha R se desensambla tal como se muestra en la figura 6.

La región R1' se corresponde con la región R1 de la figura 3, y comprende por lo tanto los primeros 640 píxeles de la totalidad de las 720 filas de la imagen.

La región R2' comprende las 320 columnas de píxeles adyacentes a la región R1', mientras que la región R3' comprende las últimas 320 columnas de píxeles.

Por tanto, el cuadro contenedor C se puede construir tal como se muestra en la figura 7, con las regiones R2' y R3' giradas 90º y dispuestas en las áreas C3' y C4' bajo la imagen L y la región R1'.

Las regiones R2' y R3' así giradas ocupan 720 píxeles de 320 filas; por lo tanto, las áreas C3' y C4' están separadas de las áreas C1 y C2 que contienen los píxeles copiados desde la imagen L y desde la región R1'.

Preferentemente, las áreas C3' y C4' están separadas de las otras áreas C1 y C2 por al menos una línea de salvaguarda. En particular, resulta ventajoso y preferible copiar los píxeles de las regiones R2' y R3' a las últimas filas del cuadro contenedor C.

Puesto que en este caso el cuadro contenedor está constituido por 1080 filas, en la forma de realización de la figura 7 las regiones giradas R2' y R3' están separadas de la anterior imagen L y de la región R1' por una tira de salvaguarda de 40 píxeles de altura.

En el ejemplo de la figura 7, las regiones R2' y R3' están separadas entre sí, de manera que están rodeadas por píxeles de un color predefinido (por ejemplo, blanco o negro) no provenientes de las imágenes derecha e izquierda. De esta manera, se reducen las áreas delimitadoras entre regiones que contienen píxeles provenientes de las imágenes derecha e izquierda, al mismo tiempo que se reducen también cualesquiera artefactos provocados por la compresión de imágenes y se maximiza la velocidad de compresión.

Como alternativa al posicionamiento de R2' y R3' en las últimas filas del cuadro contenedor C (tal como se ha descrito en referencia a la figura 7), en una forma de realización preferida R2' y R3' se posicionan de tal manera que se deja una tira de salvaguarda de una altura de 32 filas de píxeles entre el borde inferior de L y el borde superior de R2' y R3'. Esto proporciona una segunda tira de salvaguarda de una altura de 8 filas de píxeles entre el borde inferior de R2' y R3' y el borde inferior de C. Aprovechando más la anchura del cuadro contenedor, es posible posicionar R2' y R3' de tal manera que queden completamente rodeadas por píxeles que no provienen ni de la imagen derecha ni de la imagen izquierda.

Finalmente, en el área C5' en la esquina derecha inferior del cuadro C, el mapa de profundidades (DM') de superposición se introduce con una resolución de 160 x 90 píxeles, obtenida al submuestrear el mapa de profundidades original según se ha descrito previamente. En general, el mapa de profundidades de superposición puede presentar cualquier resolución, siempre que esté contenido dentro de un espacio libre del cuadro C. Para aprovechar mejor el espacio disponible, el mapa de profundidades de superposición puede experimentar una etapa de rotación y/o desensamblaje antes de introducirse en el cuadro C.

En otra forma de realización, la cual se describe en la presente en referencia a las figuras 8 y 9, el módulo 104 extrae tres subimágenes R1'', R2'' y R3'', cuya suma total de píxeles supera la de la imagen desensamblada.

La región R1'' se corresponde con la región R1' de la figura 6, mientras que R2'' y R3'' incluyen el área de las regiones R2' y R3' más un área adicional (Ra2 y Ra3) que permite minimizar la creación de artefactos durante la etapa de compresión de imágenes.

Por tanto, el segmento R1'' es una región que tiene un tamaño de 640 x 720 píxeles y que ocupa las primeras columnas del cuadro R a desensamblar.

El segmento R3'' ocupa las últimas columnas del cuadro R a desensamblar, y linda con la región central R2''. R3'' incluye, en el lateral izquierdo (el que linda con R2''), una tira amortiguadora Ra3 que contiene píxeles en común con la región R2''. En otras palabras, las últimas columnas de R2'' y las primeras de R3'' (que constituyen la tira amortiguadora Ra3) coinciden.

Preferentemente, el tamaño de la tira amortiguadora Ra3 se selecciona en función del tipo de compresión a aplicar posteriormente al cuadro contenedor C, y en general al flujo continuo de vídeo que lo contiene. En particular, dicha tira tiene un tamaño que es dos veces el de la unidad de procesado elemental usado en el proceso de compresión. Por ejemplo, la norma H.264 prevé el desensamblaje de la imagen en macrobloques de 16x16 píxeles, cada uno de los cuales representa la unidad de procesado elemental de esta norma. Sobre la base de esta consideración, la tira Ra3 tiene una anchura de 32 píxeles. Por tanto, el segmento R3'' tiene un tamaño de 352 (320+32)x720 píxeles, y comprende los píxeles de las últimas 352 columnas de la imagen R.

El segmento R2'' ocupa la parte central de la imagen R a desensamblar e incluye, en su lateral izquierdo, una tira amortiguadora Ra2 que tiene el mismo tamaño que la tira Ra3. En el ejemplo que tiene en cuenta la norma de compresión H.264, la tira Ra2 tiene una anchura por tanto de 32 píxeles y comprende píxeles en común con la región R1''. Por lo tanto, el segmento R2'' tiene un tamaño de 352x720 píxeles y comprende los píxeles de las columnas desde el 608 (640 de R1''–32) al 978 del cuadro R.

Las tres subimágenes pertenecientes a las regiones R1'', R2'' y R3'' a las que da salida el módulo 104 (visible en la figura 8) se introducen a continuación en el cuadro contenedor C tal como se muestra en la figura 9. Las regiones R2'' y R3'' se giran 90 grados y los píxeles se copian en las últimas filas del cuadro C (áreas designadas C3'' y C4'') proporcionando un cierto número de píxeles de salvaguarda que separan las áreas C3'' y C4'' con respecto a las áreas C1 y C2 que incluyen los píxeles de las imágenes L y R1''. En el caso mostrado en la figura 9, esta tira de salvaguarda tiene una anchura de 8 píxeles.

También en esta forma de realización el mapa de profundidades (DM') de superposición se introduce en el área C5' en la esquina derecha inferior del cuadro C.

El cuadro C así obtenido se comprime posteriormente y se transmite o guarda en un soporte de almacenamiento (por ejemplo, un DVD). Con este fin, se proporcionan medios de compresión que están adaptados para comprimir una imagen o una señal de vídeo, junto con medios para grabar y/o transmitir la imagen o señal de vídeo comprimida.

La figura 10 muestra un diagrama de bloques de un receptor 1100 el cual descomprime el cuadro contenedor (en caso de que estuviese comprimido) recibido o leído de un soporte, reconstruye las dos imágenes derecha e izquierda, y las pone a disposición, junto con el mapa de profundidades de superposición relacionado, para un dispositivo de visualización (por ejemplo, un televisor) permitiendo el desfrute de contenido 3D con imágenes superpuestas sobre un contenido de vídeo. El receptor 1100 puede ser una caja particular del televisor o un receptor incorporado en un televisor. Debe indicarse de que, cuando el receptor 1100 es una caja particular del televisor no integrada en este último, debe usar el mapa de profundidades internamente para posicionar correctamente los gráficos que genera (por ejemplo, subtítulos, EPG y menús relacionados). Adicionalmente, el receptor 1100 tendrá que enviar el mapa de profundidades (por ejemplo, por medio de la interfaz de HDMI) al televisor, por cuanto este último lo necesitará para posicionar correctamente sus propios gráficos (por ejemplo, sus menús).

Las mismas observaciones realizadas para receptor 1100 son también aplicables para un lector (por ejemplo, un lector de DVD) que lea un cuadro contenedor (posiblemente comprimido) y lo procese con el fin de obtener un par de cuadros correspondientes a las imágenes derecha e izquierda introducidas en el cuadro contenedor (posiblemente comprimido) leído por el lector.

En referencia de nuevo a la figura 10, el receptor recibe (por cable o antena) un flujo continuo de vídeo estereoscópico comprimido 1101 y lo descomprime por medio de un módulo de descompresión 1102, obteniendo así un flujo continuo de vídeo que comprende una secuencia de cuadros C' correspondientes a los cuadros C. Si existe un canal ideal o si se están leyendo cuadros contenedores de una memoria masiva o un soporte de datos (Blu-ray, CD, DVD), los cuadros C' se corresponden con los cuadros contenedores C portadores de la información sobre las imágenes derecha e izquierda y el mapa de profundidades de superposición, excepto por cualquier artefacto introducido por el proceso de compresión.

A continuación, estos cuadros C' se suministran a un módulo de reconstrucción 1103, el cual ejecuta un método de reconstrucción de imágenes y de extracción de mapas de profundidades según se describe posteriormente en referencia a las figuras 11 y 12.

Resulta evidente que, si el flujo continuo de vídeo no está comprimido, el módulo de descompresión 1102 se puede omitir y la señal de vídeo se puede suministrar directamente al módulo de reconstrucción 1103.

El proceso de reconstrucción comienza en la etapa 1300, cuando se recibe el cuadro contenedor descomprimido C'. El módulo de reconstrucción 1103 extrae (etapa 1301) la imagen izquierda L copiando los primeros 720x1280 píxeles del cuadro descomprimido en un cuadro nuevo el cual es menor que el cuadro contenedor, por ejemplo un cuadro de un flujo continuo de 720p. A la imagen L así reconstruida se le da salida hacia el receptor 1100 (etapa 1302).

Posteriormente, el método prevé la extracción de la imagen derecha R del cuadro contenedor C'.

La etapa de extracción de la imagen derecha comienza copiando (etapa 1303) una parte del área R1'' incluida en el cuadro C'. De forma más detallada, los píxeles de las primeras 624(640-16) columnas de R1'' se copian en las primeras 624 columnas correspondientes del cuadro nuevo que representa la imagen reconstruida Rsalida, tal como se muestra en la figura 11. De hecho, esto elimina de la etapa de reconstrucción las 16 columnas de R1'' que están más sujetas a la creación de artefactos, por ejemplo a través del ejemplo de procedimiento de estimación de movimiento llevado a cabo por la norma de compresión H.264.

A continuación, se extrae una parte central de R2'' (etapa 1304). Del cuadro descomprimido C' (el cual, tal como se ha mencionado anteriormente, se corresponde con el cuadro C de la figura 9), se seleccionan los píxeles del área C3'' (correspondientes a la región de origen R2'') y se realiza una rotación de 90º inversa a la ejecutada en el multiplexor 100, lo cual los devuelve a la condición original de filas/columnas, es decir la que se muestra en la figura 8. Llegado este momento, se eliminan las primeras y últimas dieciséis (16) columnas de R2'' y las restantes 352-32=320 columnas de píxeles se copian en las columnas libres adyacentes a las recién copiadas de R1''.

Al cortar las 16 columnas más externas de la región R2'', se eliminan aquellas columnas en la que es más probable que se produzca una formación de artefactos. La anchura del área cortada (en este caso 16 columnas) depende del tipo de compresión utilizado. Dicha área es preferentemente igual a la unidad de procesado elemental utilizada por el proceso de compresión; en el caso que se describe en la presente, la norma H.264 actúa sobre bloques de 16x16 píxeles, y por lo tanto se deben cortar 16 columnas.

En cuanto a R3'' (etapa 1305), los píxeles de la región C4'' se extraen del cuadro C' y la subimagen R3'' se devuelve al formato original de filas/columnas (véase la figura 8). Posteriormente, se eliminan las primeras 16 columnas de píxeles (correspondientes a la mitad del área Ra3) y las restantes 352-16=336 columnas de píxeles se copian en las últimas columnas libres a la izquierda del cuadro reconstruido. Igual que en R2'', también en R3'' el área cortada es igual a la unidad de procesado elemental utilizada por el proceso de compresión.

Evidentemente, para ambas regiones R2'' y R3'' la etapa de rotación se puede llevar a cabo de una manera virtual, es decir se puede obtener el mismo resultado en términos de extracción de los píxeles de interés copiando en el cuadro reconstruido los píxeles de una fila del área C3'' (en el caso de R2'', C4'' en el caso de R3'') en una columna del cuadro nuevo Rsalida, excepto por las últimas 16 filas del área C3'' (en caso de R2'', C4'' en caso de R3'') correspondiente a las 16 columnas a cortar, lo que se muestra en la figura 8.

Llegado este punto, la imagen derecha Rsalida se ha reconstruido por completo y se puede dar salida a la misma (etapa 1306).

Finalmente, el módulo de reconstrucción 1103 extrae (etapa 1308) el mapa de profundidades DM' de superposición copiando en un registro los valores de luminancia de los últimos 160 x 90 píxeles del cuadro contenedor descomprimido C', correspondiente al área C5'. Al contenido de dicho registro se le da salida hacia el receptor 1100 (etapa 1309) y el mismo se usará para definir la posición en profundidad de imágenes (texto o gráficos) a combinar con el contenido tridimensional transportado por el flujo continuo de vídeo estereoscópico; en particular, se usará para combinar imágenes a superponer sobre el contenido tridimensional.

Como alternativa a dar salida al contenido del mapa de profundidades y a las imágenes L y R extraídas de los cuadros de entrada o de forma adicional a esto último, el receptor 1100 comprende un generador de caracteres y/o un generador de gráficos y combina otras imágenes con las imágenes L y R, es decir con el contenido tridimensional. Las imágenes a combinar se seleccionan de un área de memoria del receptor y se pueden almacenar cuando se fabrica el receptor (por ejemplo, los gráficos de algunos menús o de los números de canal) o se pueden extraer del flujo continuo de vídeo (por ejemplo, información de guías de programación y subtítulos).

Estas imágenes se combinan con el contenido tridimensional en posiciones en cuanto a profundidad que dependen de los mapas de profundidades de superposición extraídos del flujo continuo de vídeo. En particular, para cada imagen estereoscópica (producida por el par de imágenes L y R), la imagen combinada se coloca en el punto de profundidad mínima de la imagen estereoscópica.

Después de que las imágenes se hayan combinado con el contenido 3D, en esta forma de realización el receptor 1100 da salida a un par de imágenes L* y R* las cuales, cuando se reproducen, serán percibidas por el usuario como contenido tridimensional correspondiente a la original (producida por las imágenes L y R) con imágenes superpuestas sobre ella, por ejemplo, subtítulos, menús, gráficos, etcétera.

El proceso para reconstruir las imágenes derecha e izquierda y el mapa de profundidades contenido en el cuadro contenedor C' se completa de esta manera (etapa 1307). Dicho proceso se repite para cada cuadro del flujo continuo de vídeo recibido por el receptor 1100, de manera que la salida estará compuesta por dos flujos continuos de vídeo 1104 y 1105 para la imagen derecha y para la imagen izquierda, respectivamente, y una señal de datos que se deduce a partir del mapa de profundidades de superposición.

El proceso para reconstruir las imágenes derecha e izquierda y el mapa de profundidades de superposición que se ha descrito anteriormente en referencia a las figuras 10, 11 y 12 se basa en la suposición de que el desmultiplexor 1100 sabe cómo se construyó el cuadro contenedor C y por lo tanto puede extraer las imágenes derecha e izquierda y el mapa de profundidades de superposición.

Evidentemente, esto es posible si el método de multiplexado está normalizado.

Para tener en cuenta el hecho de que el cuadro contenedor se puede generar de acuerdo con uno cualquiera de los métodos antes descritos, o en cualquier caso de acuerdo con uno cualquiera de los métodos que utilizan la solución que es materia de las reivindicaciones adjuntas, el desmultiplexor usa la información de indicador contenido en una región predefinida de la imagen compuesta (por ejemplo, un código de barras, según se ha descrito previamente) con el fin de saber cómo debe desempaquetarse el contenido de la imagen compuesta y cómo reconstruir las imágenes derecha e izquierda y el mapa de profundidades de superposición.

Después de descodificar el indicador, el desmultiplexor conocerá la posición de la imagen sin variaciones (por ejemplo, la imagen izquierda en los ejemplos antes descritos), así como las posiciones y cualesquiera transformaciones (rotación, traslación o similares) de las regiones en las cuales se desensambló la otra imagen (por ejemplo, la imagen derecha en los ejemplos antes descritos) y la posición del mapa de profundidades de superposición.

Con esta información, el desmultiplexor puede extraer así la imagen sin variaciones (por ejemplo, la imagen izquierda) y el mapa de profundidades, y reconstruir la imagen desensamblada (por ejemplo, la imagen derecha).

Aunque la presente invención se ha ilustrado hasta el momento en referencia a algunas formas de realización preferidas y ventajosas, es evidente que la misma no se limita a dichas formas de realización y que en ellas se pueden realizar muchos cambios por parte de una persona versada en la materia que desee combinar en una imagen compuesta dos imágenes referentes a dos perspectivas diferentes (derecha e izquierda) de un objeto o una escena.

Por ejemplo, los módulos electrónicos que proporcionan los dispositivos antes descritos, en particular el dispositivo 100 y el receptor 1100, se pueden subdividir y distribuir de forma variada; además, se pueden proporcionar en forma de módulos de hardware o como algoritmos de software implementados por un procesador, en particular un procesador de vídeo equipado con áreas de memoria adecuadas para almacenar temporalmente los cuadros de entrada recibidos. Por lo tanto, estos módulos pueden ejecutar en paralelo o en serie una o más de las etapas de procesado de vídeo de los métodos de multiplexado y de desmultiplexado de imágenes de acuerdo con la presente invención.

Resulta evidente también que, aunque las formas de realización preferidas se refieren al multiplexado de dos flujos continuos de vídeo de 720p en un flujo continuo de vídeo de 1080p, también pueden utilizarse otros formatos, tales como, por ejemplo, dos flujos continuos de vídeo de 640x480 en un flujo continuo de vídeo de 1280x720, o dos flujos continuos de vídeo de 320x200 en un flujo continuo de vídeo de 640x480.

Tampoco la invención se limita a un tipo particular de disposición de la imagen compuesta, puesto que diferentes soluciones para generar la imagen compuesta pueden ofrecer ventajas específicas.

Por ejemplo, las formas de realización descritas anteriormente en referencia a las figuras 1 a 12 ofrecen la ventaja de que únicamente llevan a cabo operaciones de traslación o rototraslación, siendo así necesario solamente un poder computacional reducido.

Alternativamente, es concebible que las imágenes se sometan también a etapas de inversión especular, además de dichas operaciones de rotación y/o traslación, con el fin de obtener una imagen compuesta del tipo que se muestra en la figura 13.

Estas operaciones adicionales se llevan a cabo con la finalidad de maximizar los perímetros delimitadores entre regiones que contienen píxeles homólogos, aprovechando así la fuerte correlación existente entre ellas y reduciendo

al mínimo los artefactos introducidos por la etapa de compresión subsiguiente. En el ejemplo de las figuras 13 y 14 se ha supuesto, por motivos de claridad, que las dos imágenes derecha e izquierda son idénticas, aún cuando en general las mismas difieren ligeramente.

En esta figura, la imagen izquierda L (mostrada en la figura 14a) está posicionada en la esquina derecha superior del cuadro contenedor C, de manera que ocupa los últimos 1280 píxeles de las primeras 720 filas. Como en los ejemplos previamente descritos, la imagen L se copia por tanto sin variaciones en el cuadro contenedor C.

La imagen derecha R en cambio se desensambla de acuerdo con el ejemplo de la figura 3; la figura 14b muestra la imagen R descompuesta en tres regiones R1, R2 y R3.

Posteriormente, algunas regiones (las regiones R1 y R3 en el ejemplo de la figura 14) experimentan una operación de inversión especular; la inversión puede producirse con respecto a un eje vertical (es decir, paralelo a una columna de la imagen) o con respecto a un eje horizontal (es decir, paralelo a una fila de la imagen).

En el caso de una inversión con respecto a un eje vertical, los píxeles de la columna N (donde N es un entero entre 1 y 1080, siendo 1080 el número de columnas de la imagen) se copian en la columna 1080+1-N.

En el caso de una inversión con respecto a un eje horizontal, los píxeles de la fila M (donde M es un entero entre 1 y 720, siendo 720 el número de filas de la imagen) se copian en la fila 720+1-N.

Las figuras 14c y 14d muestran la región R1 extraída de la imagen R e invertida (R1rot) con respecto a un eje vertical, en particular con respecto a un lateral vertical.

La región invertida R1inv se introduce en los primeros 640 píxeles de las primeras 640 filas de píxeles. Tal como puede observarse en el ejemplo de la figura 13, cuando R1inv se introduce girada en el cuadro contenedor C, los píxeles de R1inv que lindan con L son muy similares a los píxeles de L que lindan con R1inv. La correlación espacial entre estos píxeles tiene la ventaja de reducir la formación de artefactos.

Las figuras 14e y 14f muestran la región R3 extraída de la imagen R de la figura 14b y a continuación invertida (R3inv) con respecto a un eje horizontal, en particular con respecto a un lateral horizontal.

La región R3inv se introduce en los últimos 640 píxeles de las últimas 360 filas. Esto reduce la generación de artefactos, puesto que los píxeles de las regiones delimitadoras entre R3inv y L son píxeles que tienen una alta correlación espacial. De hecho, los píxeles en esta región delimitadora reproducen partes similares o idénticas de la imagen.

A continuación, el cuadro contenedor C se completa introduciendo la región R2.

En este ejemplo R2 no se invierte y/o gira debido a que no sería posible, en ningún caso, hacer coincidir una región delimitadora de R2 con una región delimitadora constituida por píxeles homólogos de otra región de R o L.

Finalmente, resulta también evidente que la invención se refiere asimismo a cualquier método de desmultiplexado que permita extraer una imagen derecha y una imagen izquierda de una imagen compuesta invirtiendo uno de los procesos de multiplexado antes descrito, y situándose dentro del alcance de protección de la presente invención.

Por tanto, la invención se refiere también a un método para generar un par de imágenes a partir de una imagen compuesta, el cual comprende las etapas siguientes:

- generar una primera (por ejemplo, la imagen izquierda) de dichas imágenes derecha e izquierda copiando un único grupo de píxeles contiguos desde una región de dicha imagen compuesta,

- generar una segunda imagen (por ejemplo, la imagen derecha) copiando otros grupos de píxeles contiguos de regiones diferentes de dicha imagen compuesta.

Según una forma de realización, la información para generar dicha segunda imagen se extrae de un área de dicha imagen compuesta. Dicha información se codifica preferentemente de acuerdo con un código de barras.

En una forma de realización del método para generar las imágenes derecha e izquierda, la generación de la imagen que se desensambló en la imagen compuesta comprende por lo menos una etapa de inversión especular de un grupo de píxeles de una de dichas regiones diferentes.

En una forma de realización del método para generar las imágenes derecha e izquierda, la generación de la imagen que se desensambló en la imagen compuesta comprende por lo menos una etapa de eliminación de píxeles de una de las regiones de la imagen compuesta que comprenden los píxeles de esta imagen a reconstruir. En particular, los píxeles se eliminan de un área delimitadora de esta región.

En una forma de realización, la imagen que se desensambló en diferentes regiones de la imagen compuesta se reconstruye sometiendo las regiones de píxeles que incluyen los píxeles de la imagen a desensamblar, únicamente a operaciones de traslación y/o rotación.

Aunque el ejemplo de forma de realización antes descrito se refiere a la introducción de un mapa de profundidades de superposición en un cuadro contenedor en el cual una cualquiera de las dos imágenes derecha e izquierda se desensambla en varias partes, es evidente que la invención no depende de la manera en la que se da formato a las dos imágenes derecha e izquierda dentro del cuadro contenedor. Por ejemplo, las dos imágenes se pueden submuestrear y disponer lado con lado (formato de lado-con-lado) o una encima de la otra (formato de arriba-abajo) con el fin de dejar un espacio libre en el cuadro en el que se pueda colocar el mapa de profundidades de superposición. Además, una cualquiera de las imágenes derecha e izquierda se puede dejar sin variaciones, mientras que la otra se puede submuestrear con el fin de liberar espacio para el mapa de profundidades.

Finalmente, debe observarse que los ejemplos de forma de realización antes descritos en referencia a los dibujos adjuntos se refieren a un mapa de profundidades “completo”, es decir un mapa de profundidades que se calcula diezmando o filtrando un mapa de profundidades del contenido 3D aunque sin subdividirlo en varias partes, a diferencia de una de las dos imágenes L y R, por ejemplo. Sin embargo, esto no constituye una limitación de la presente invención, y el mapa de profundidades de superposición, una vez generado (o recibido), puede ser introducido en el cuadro contenedor por un codificador, el cual lo descompondrá en múltiples partes que se dispondrán en diferentes regiones del cuadro contenedor. Por ejemplo, tal como es sabido, con el fin de codificar un contenido estereoscópico, un codificador H.264 debe introducir ocho filas adicionales las cuales serán cortadas por el descodificador; en una forma de realización, el mapa de profundidades de superposición se puede introducir en estas ocho filas adicionales dividiéndolo, por ejemplo, en 240 bloques de un tamaño de 8x8, los cuales, cuando se reensamblen apropiadamente, formarán una imagen con dimensiones proporcionales al contenido estereoscópico transportado. Se puede obtener un ejemplo de disposición de bloques explorando las filas de un mapa de profundidades diezmado por 16, y por lo tanto con una resolución de 120x72, en donde tiras de 120x8 píxeles se alinean con el fin de obtener una imagen de 1080x8 píxeles. En otra forma de realización, el mismo mapa de profundidades diezmado se puede subdividir en un número mayor de tiras de una altura de 8 píxeles utilizando un desplazamiento de 6 píxeles en lugar de uno de 8 píxeles, de manera que el contenido se hace redundante y se fomenta la protección del mismo en el límite con la imagen principal. Esto parece resultar particularmente ventajoso siempre que el contenido estereoscópico incluya un par de imágenes derecha e izquierda multiplexadas en un formato de arriba-abajo, de lado-con-lado, o de tablero de ajedrez, con una resolución tal que ocupe todos los píxeles potencialmente visualizables en el cuadro, por ejemplo los píxeles de un formato de 1920x1080.

Preferentemente, en el caso de que el cuadro incluya un par de imágenes diezmadas asimétricamente (por ejemplo, un formato de lado-con-lado en el que las columnas se diezman más que las filas, o un formato de arriba-abajo en el que se diezman únicamente las filas, no las columnas), entonces el mapa de profundidades de superposición se obtiene diezmando un mapa de profundidades con una relación de diezmado de filas/columnas proporcional a la utilizada para muestrear las imágenes colocadas en el mismo cuadro. A título de ejemplo, suponiendo que se use un formato de lado-con-lado para multiplexar las imágenes derecha e izquierda en el cuadro, la relación de diezmado de filas/columnas será 1:2, puesto que se mantienen todas las filas y las columnas se diezman en dos. En este caso, el mapa de profundidades de superposición se puede obtener diezmando un mapa de profundidades con una relación de diezmado de filas/columnas de 1:2.

Resulta también evidente que se pueden usar métodos diferentes para señalizar al receptor el área ocupada por el mapa de profundidades, que no sean los descritos anteriormente, y los cuales prevean la introducción de un indicador en la imagen; de hecho, dicho indicador también se puede incluir en un paquete de datos de la señal portadora del flujo continuo de vídeo.

REIVINDICACIONES

1. Método para superponer imágenes sobre un contenido tridimensional, en el que se recibe un flujo continuo de vídeo, que comprende dicho contenido tridimensional y un mapa de profundidades (DM, DM') para superponer imágenes sobre dicho contenido tridimensional, conteniendo dicho mapa de profundidades (DM, DM') información sobre la profundidad de dicho contenido tridimensional, y siendo insertado como una imagen en un cuadro (C) de dicho flujo continuo de vídeo,

estando el método caracterizado por que utiliza dicho mapa de profundidades (DM, DM') en la fase de reproducción únicamente para la superposición de imágenes a dicho contenido tridimensional, siendo dicha superposición realizada en una posición en cuanto a profundidad en función de dicho mapa de profundidades (DM, DM') de superposición,

y por que dicho mapa de profundidades (DM, DM') tiene un número de píxeles menor que el de una imagen bidimensional incluida en dicho cuadro (C) y asociada a dicho contenido tridimensional.

2. Método según la reivindicación 1, en el que dicho mapa de profundidades de superposición únicamente contiene información sobre la profundidad de píxeles situados en la mitad inferior, preferentemente en el tercio inferior, de dicho contenido tridimensional.

3. Método según la reivindicación 1 o 2, en el que el mapa de profundidades de superposición tiene una resolución no uniforme, en particular la mitad o tercio inferior de dicho mapa de profundidades tiene una resolución mayor que la parte superior.

4. Método según la reivindicación 1 o 2 o 3, en el que dicho mapa de profundidades de superposición tiene una resolución menor que una imagen bidimensional asociada a dicho contenido tridimensional.

5. Método según la reivindicación 4, en el que dicho contenido tridimensional es una imagen que consiste en una pluralidad de píxeles, y en el que dicho mapa de profundidades se obtiene submuestreando un mapa de profundidades, cuyos elementos se corresponden con la profundidad de los píxeles de dicho contenido tridimensional.

6. Método según la reivindicación 5, en el que, después de submuestrear dicho mapa de profundidades, el mapa submuestreado se divide en bloques y a cada píxel del bloque se le da un mismo valor igual a la profundidad mínima de los píxeles de dicho bloque o al valor medio de la profundidad de los píxeles del bloque.

7. Método según la reivindicación 5, en el que, antes de submuestrear dicho mapa de profundidades, el mapa de profundidades se divide en bloques y a cada píxel del bloque se le da un mismo valor igual a la profundidad mínima de los píxeles de dicho bloque o al valor medio de la profundidad de los píxeles del bloque.

8. Método según la reivindicación 6 o 7, en el que dichos bloques tienen un tamaño igual a un múltiplo de un bloque elemental de 2x2 píxeles.

9. Método según cualquiera de las reivindicaciones anteriores, en el que dicho mapa de profundidades de superposición se introduce en una parte de dicho cuadro no destinada a la visualización.

10. Método según cualquiera de las reivindicaciones anteriores, en el que dicho mapa de profundidades se descompone en bloques distribuidos en áreas de dicho cuadro (C) que no están ocupadas por dicho contenido tridimensional.

11. Método según cualquiera de las reivindicaciones anteriores, en el que dicho cuadro comprende una imagen derecha, una imagen izquierda y dicho mapa de profundidades, en el que dicho mapa de profundidades se descompone en bloques distribuidos en regiones del cuadro (C) que no están ocupadas por dicho contenido tridimensional, y en el que dicho cuadro (C) se codifica según la norma de codificación H.264.

12. Método según cualquiera de las reivindicaciones 1 a 9, en el que dicho contenido tridimensional comprende una imagen bidimensional e información que permite reconstruir la otra imagen de un par estereoscópico, y en el que dicho mapa de profundidades de superposición es introducido en una parte de la imagen bidimensional.

13. Método según cualquiera de las reivindicaciones 1 a 12, en el que dicho cuadro comprende un indicador adaptado para indicar al receptor la posición de dicho mapa de profundidades de superposición dentro de dicho cuadro.

14. Método según cualquiera de las reivindicaciones 1 a 13, en el que dicho flujo continuo de vídeo comprende un indicador adaptado para indicar al receptor la posición de dicho mapa de profundidades de superposición dentro de dicho cuadro, siendo dicho indicador externo a dicho cuadro.

15. Dispositivo para reproducir contenido tridimensional, que comprende unos medios adaptados para recibir un flujo continuo de vídeo que contiene contenido tridimensional, unos medios adaptados para combinar una imagen con dicho contenido tridimensional, caracterizado por que dichos medios adaptados para combinar una imagen con dicho contenido tridimensional están adaptados para implementar un método según cualquiera de las reivindicaciones 1 a 14, y comprenden unos medios para extraer dicho mapa de profundidades de superposición y para combinar dicha imagen que se va a superponer sobre el contenido tridimensional, basándose en una profundidad en función de dicho mapa de profundidades de superposición.

16. Flujo continuo de vídeo estereoscópico (1101), que comprende una pluralidad de cuadros y caracterizado por que dichos cuadros comprenden por lo menos un contenido tridimensional y por lo menos un mapa de profundidades (DM, DM') de superposición insertado como imagen dentro de cada cuadro de entre dicha pluralidad de cuadros, siendo dicho mapa de profundidades (DM, DM') de superposición un mapa de profundidades para superponer imágenes sobre dicho contenido tridimensional, siendo dicha superposición realizada en una posición en cuanto a profundidad en función de dicho mapa de profundidades de superposición, y comprendiendo un número de píxeles menor que el de una imagen bidimensional asociada a dicho contenido tridimensional, estando dicho flujo continuo de vídeo estereoscópico adaptado para ser utilizado en un método según cualquiera de las reivindicaciones 1 a 14.

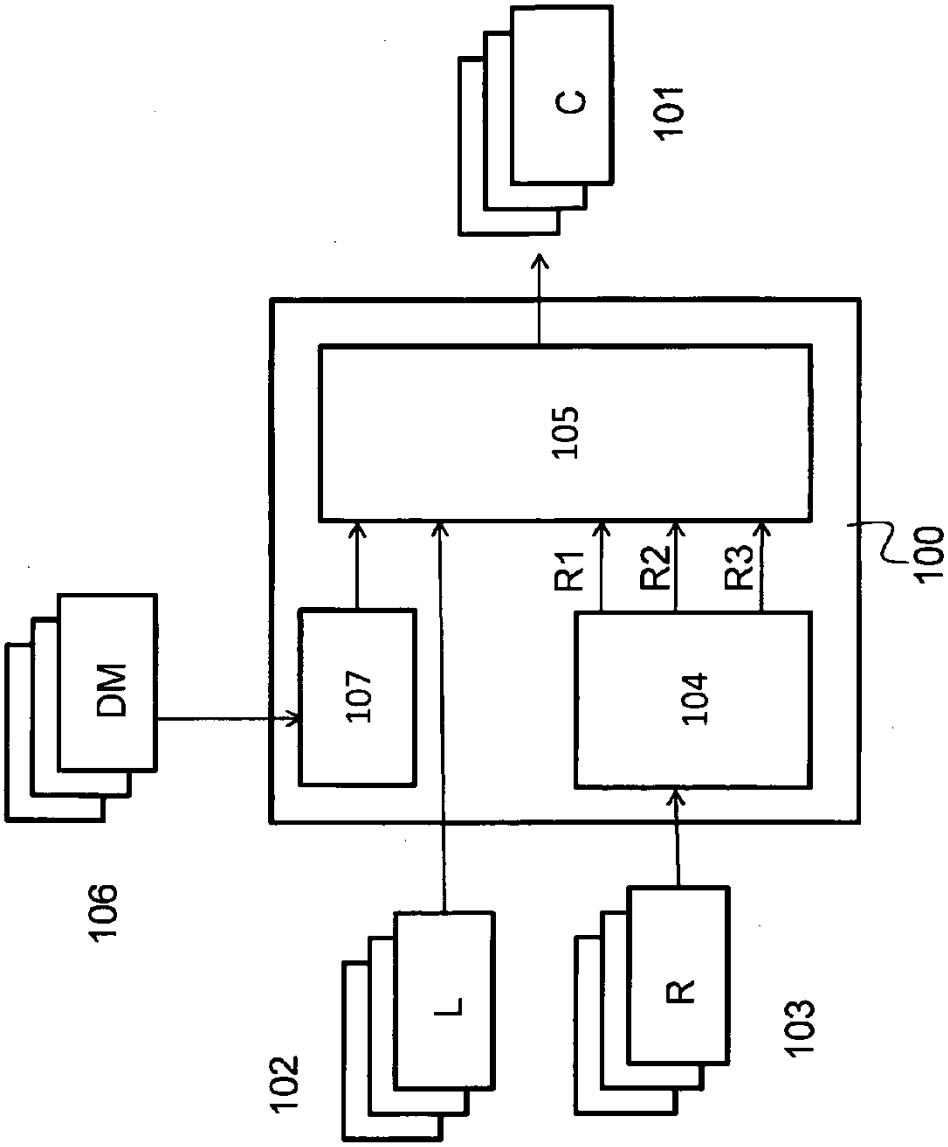

Fig. 1

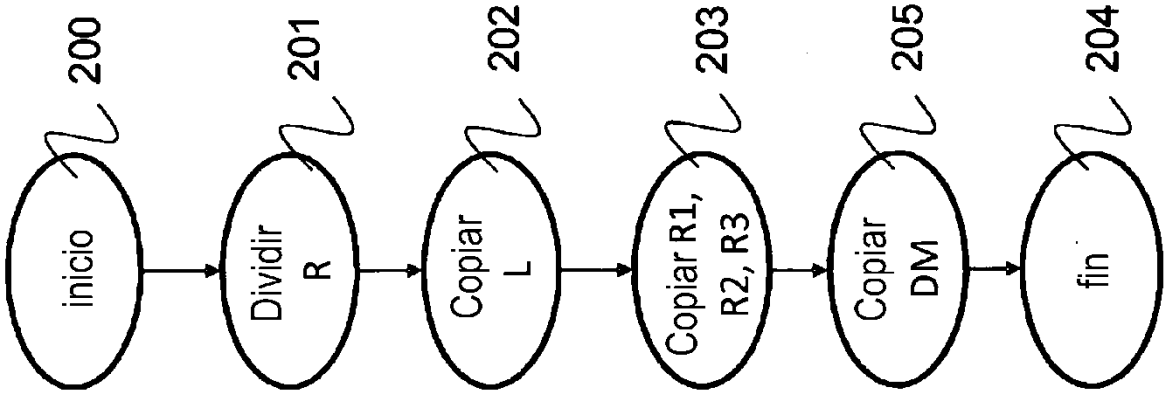
inicio
200
Dividir R
201
Copiar L
202
Copiar R1, R2, R3
203
Copiar DM
205
fin
204

Fig. 2

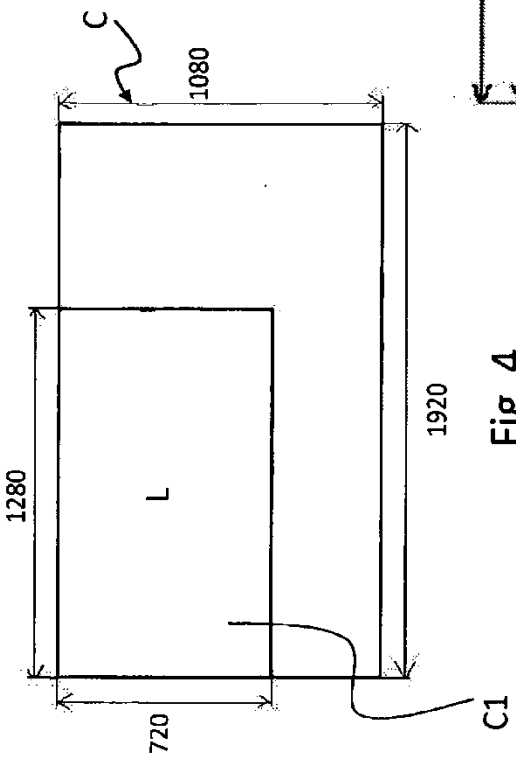
C
1080
1280
L
1920
720
C1

Fig. 4

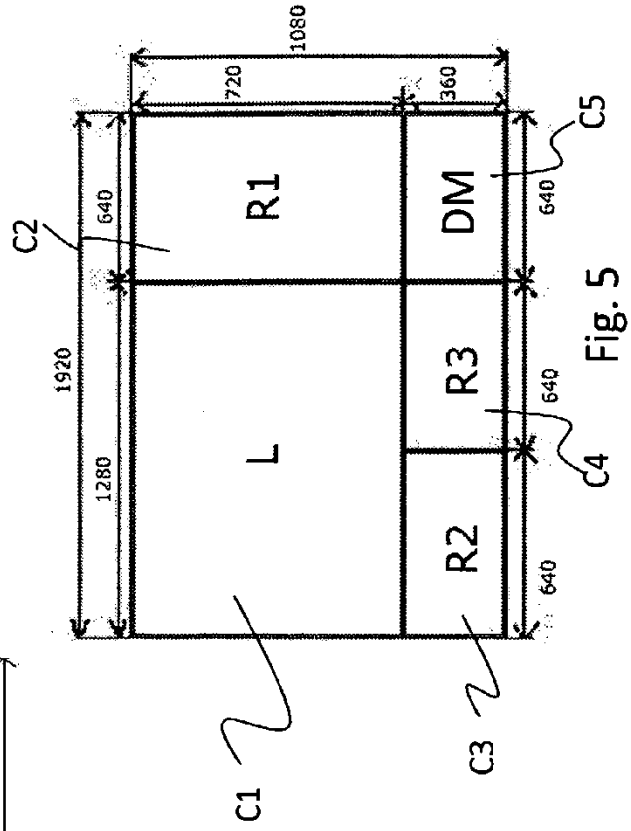
1080
720
360
C5
C2
640
R1
DM
640
1920
R3
640
C4
1280
L
R2
640
C1
C3

Fig. 5

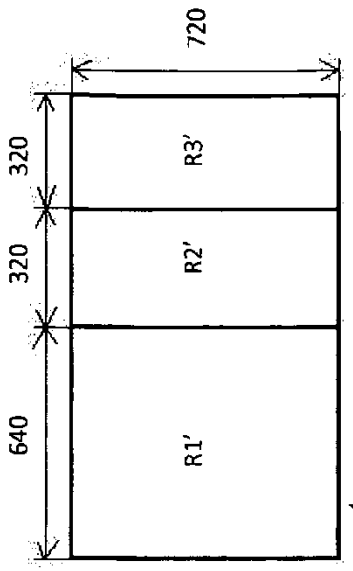

Fig. 6

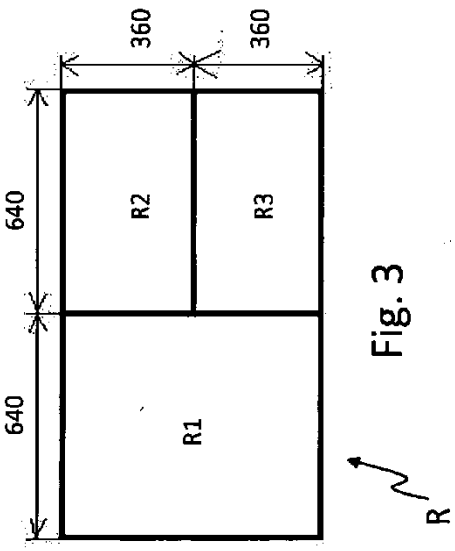

Fig. 3

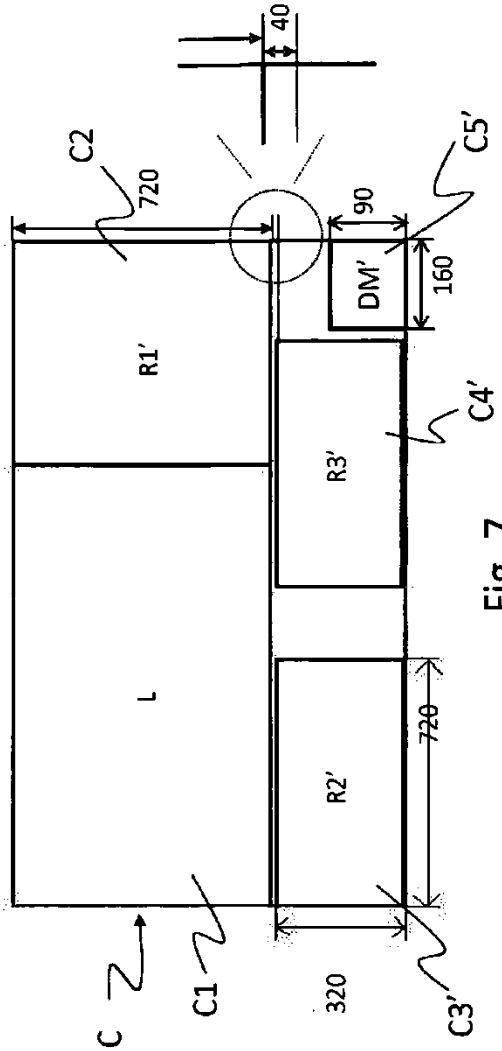

Fig. 7

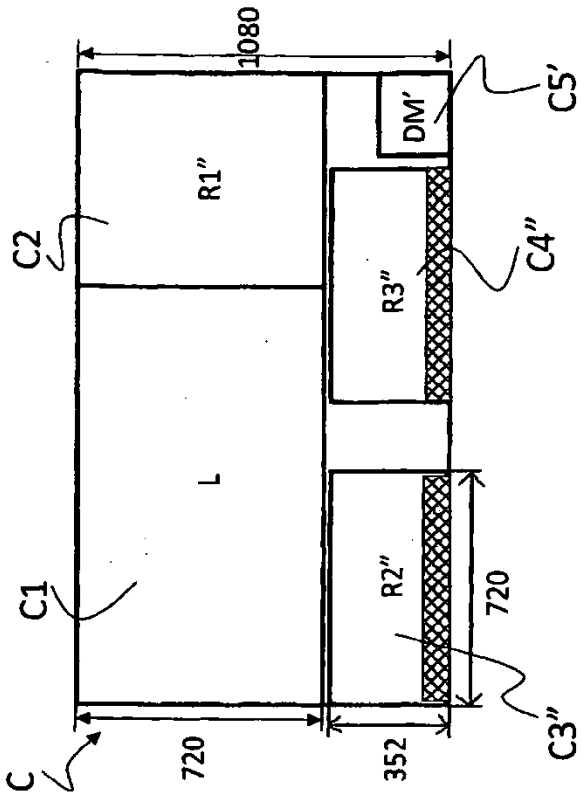
1080
C5'
DM'
R1"
C2
R3"
C4"
L
C1
R2"
720
C3"
C
720
352

Fig.9

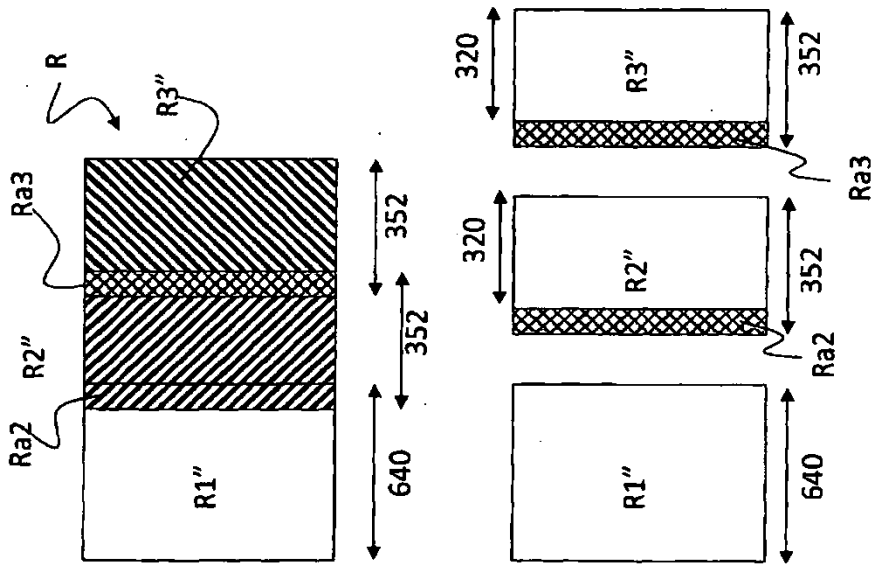
R
R3"
Ra3
R2"
Ra2
R1"
352
352
640
320
R3"
352
Ra3
320
R2"
352
Ra2
R1"
640

Fig. 8

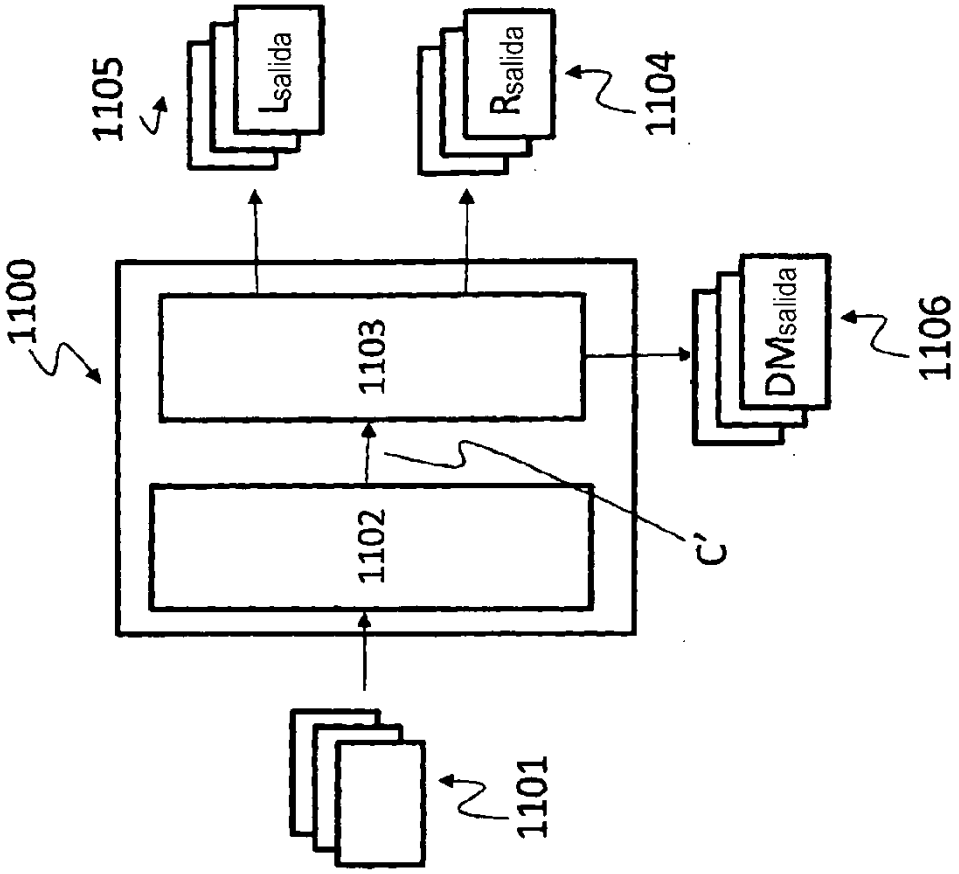
1100
1101
1102
1103
C'
1104
1105
1106
Lsalida
Rsalida
DMsalida

Fig. 10

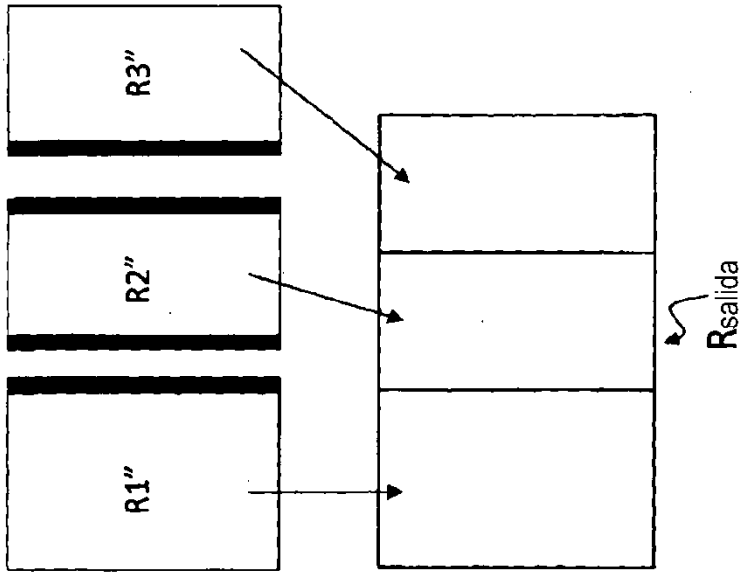
R1"
R2"
R3"
Rsalida

Fig. 11

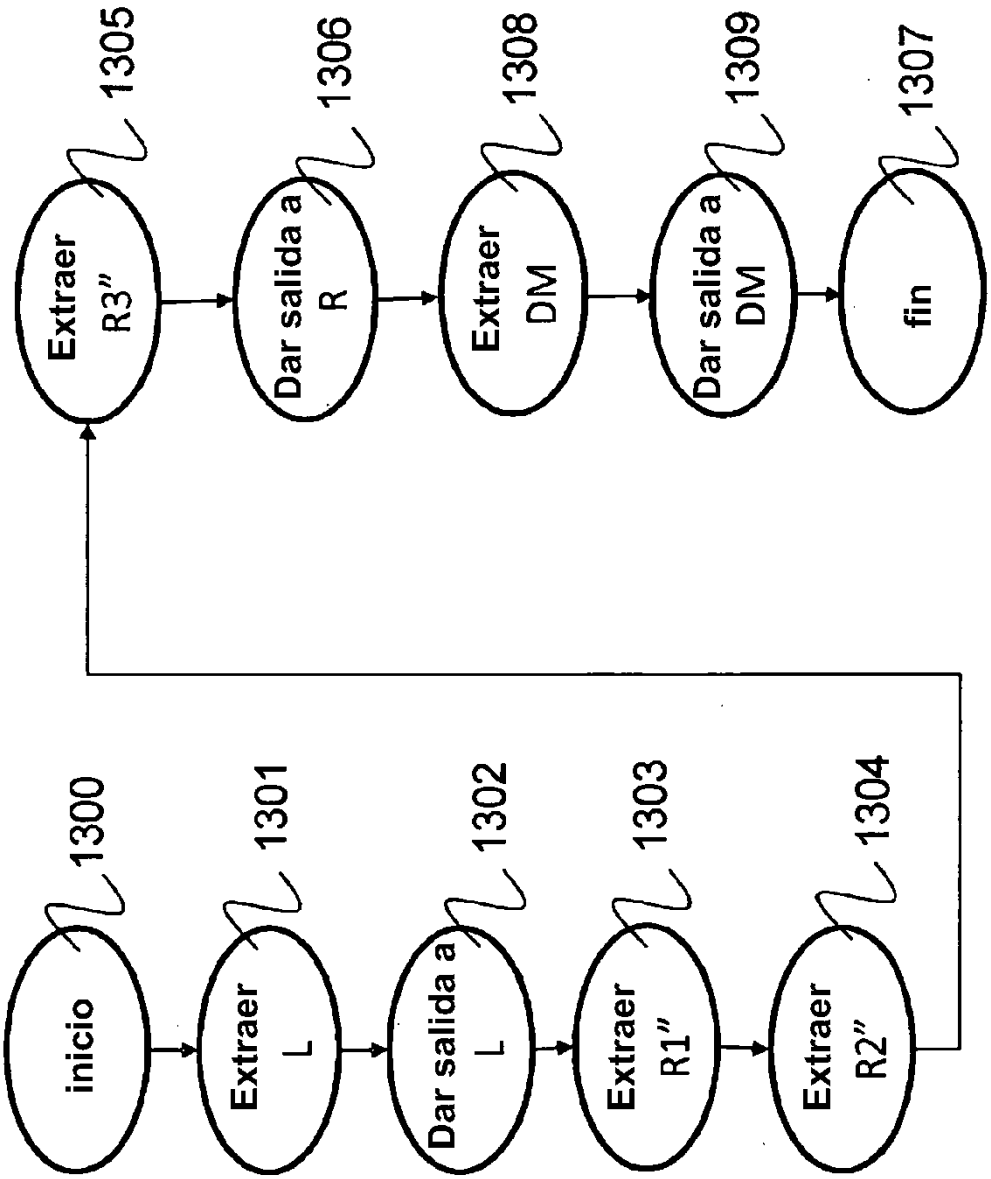
inicio
1300
Extraer L
1301
Dar salida a L
1302
Extraer R1”
1303
Extraer R2”
1304
Extraer R3”
1305
Dar salida a R
1306
Extraer DM
1308
Dar salida a DM
1309
fin
1307

Fig. 12

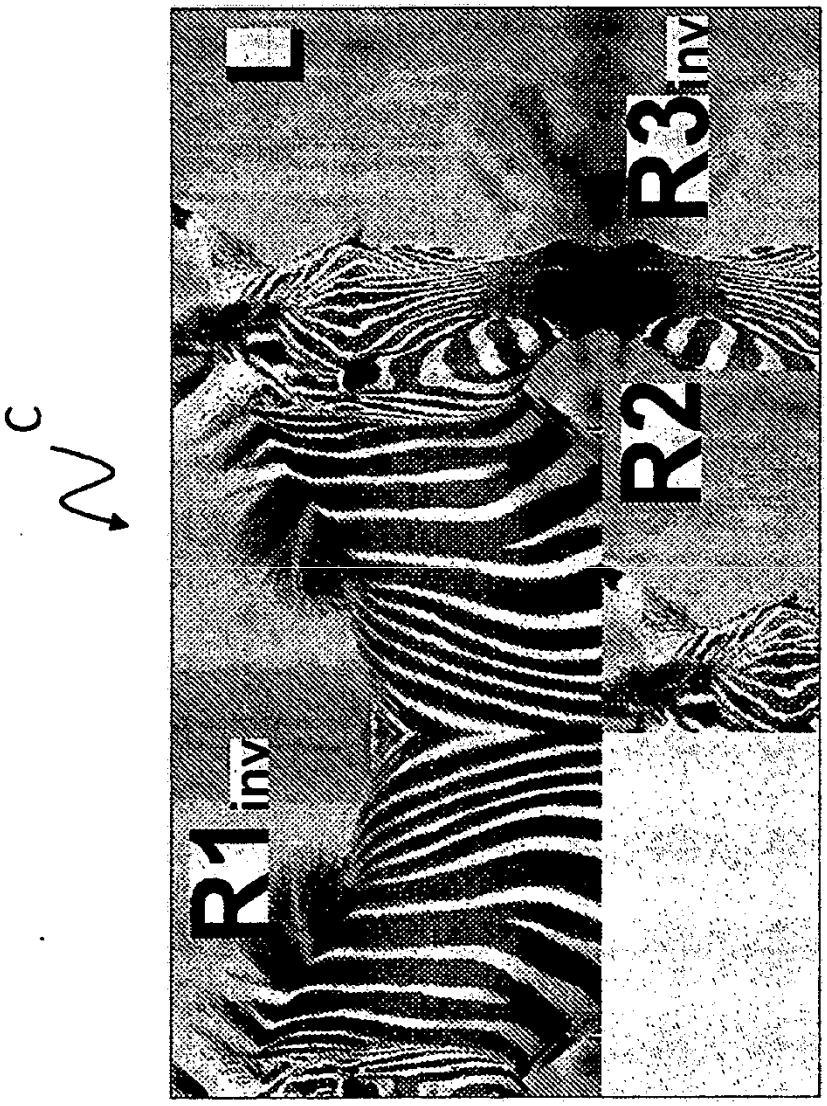

Fig. 13

Fig. 14a

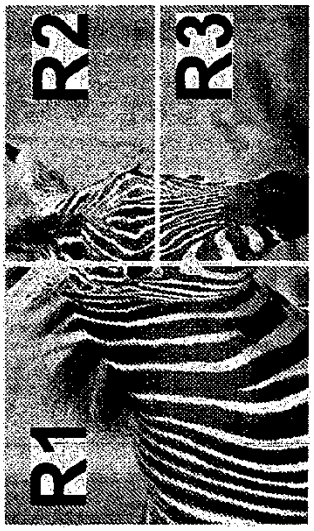

Fig. 14b

Fig. 14c

Fig. 14d

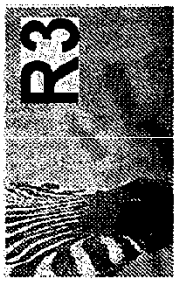

Fig. 14e

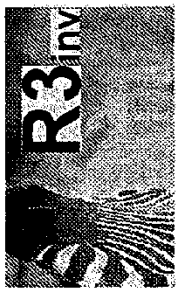

Fig. 14f