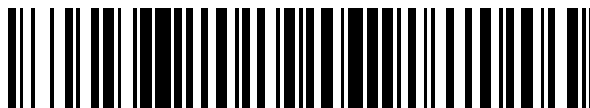


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 550 128**

21 Número de solicitud: 201430919

51 Int. Cl.:

H04N 13/02

(2006.01)

12

SOLICITUD DE PATENTE

A2

22 Fecha de presentación:

17.06.2014

30 Prioridad:

18.06.2013 IT TO2013A000503

43 Fecha de publicación de la solicitud:

04.11.2015

71 Solicitantes:

SISVEL TECHNOLOGY S.R.L. (100.0%)

Via Castagnole 59

I-10060 Turín, None IT

72 Inventor/es:

GRANGETTO, Marco;

LUCENTEFORTE, Maurizio y

D'AMATO, Paolo

74 Agente/Representante:

CURELL AGUILÁ, Mireia

54 Título: **Procedimiento para generar una imagen de color compuesta por una pluralidad de componentes partiendo de por lo menos un mapa de profundidad o disparidad y aparato correspondiente**

57 Resumen:

Procedimiento para generar una imagen de color compuesta por una pluralidad de componentes partiendo de por lo menos un mapa de profundidad o disparidad y aparato correspondiente. En dicho procedimiento, se introduce un primer conjunto de píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) en la componente de luminancia (Y) de dicha imagen de color, y en el que se introducen un segundo y un tercer conjunto de píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) en las dos componentes de crominancia (U, V) de dicha imagen de color.

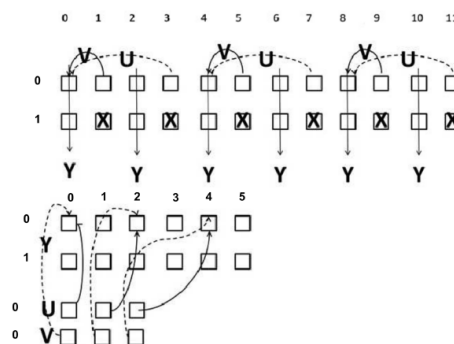


FIG. 2a

DESCRIPCIÓN

Procedimiento para generar una imagen de color compuesta por una pluralidad de componentes partiendo de por lo menos un mapa de profundidad o disparidad y aparato correspondiente.

5 Campo de la invención

La presente invención se refiere a un procedimiento y a un dispositivo para generar, almacenar, transmitir, recibir y reproducir mapas de profundidad utilizando las componentes de color de una imagen que pertenece a un flujo de vídeo tridimensional.

Estado de la técnica

10 El desarrollo de aplicaciones de vídeo estereoscópicas depende en gran parte de la disponibilidad de formatos eficaces para representar y comprimir la señal de vídeo tridimensional. Además, en aplicaciones de difusión televisiva (3D-TV), es necesario mantener el grado más alto posible de retrocompatibilidad con los sistemas en 2D existentes.

15 Para una distribución (o transmisión), las soluciones técnicas más extendidas en la actualidad se basan en la denominada “disposición de fotograma compatible”, en la que las dos vistas estereoscópicas, en relación con el mismo instante en el tiempo, se reajustan a escala y se componen para formar una única imagen compatible con los formatos existentes. Entre estas soluciones, se conocen los formatos encima y debajo, de lado a lado y de baldosa. Estas soluciones permiten utilizar toda la infraestructura de distribución de señal de vídeo existente
20 (terrestre, por satélite o de difusión por cable o la transmisión en flujo continuo a través de la red IP), y no requieren normas nuevas para la compresión del flujo de vídeo. Además, la norma de codificación AVC/H.264 actual (codificación de vídeo avanzada) y la norma HEVC futura (codificación de vídeo de alta eficacia) ya incluyen la posibilidad de señalar este tipo de organización para permitir una reconstrucción y visualización apropiadas por el receptor.

25 Para la visualización, las dos soluciones técnicas más extendidas en la actualidad se basan o bien en el principio de “fotograma alterno” (es decir las dos vistas se presentan en sucesión de tiempo en la pantalla) o bien en el principio de “línea alterna”, es decir las dos vistas se disponen en la pantalla con filas alternas (es decir se “entrelazan”). En ambos casos, para que cada ojo reciba la vista correspondiente, es necesario que el espectador utilice un par de gafas,
30 que pueden ser o bien “activas”, es decir gafas de obturación, en el caso de fotograma alterno, o bien “pasivas”, es decir con lentes polarizadas de manera diferente, en el caso de línea alterna.

35 El futuro de una visualización tridimensional se determinará mediante la difusión de pantallas autoestereoscópicas nuevas que no requerirán que el usuario utilice gafas, ya sean pasivas o activas. Estos dispositivos de visualización en 3D, que actualmente están todavía en la fase de prototipo, se basan en la utilización de lentes o barreras de paralaje que pueden hacer que el espectador perciba dos vistas estereoscópicas diferentes para cada punto de vista en el que puede encontrarse el usuario cuando se mueve de manera angular alrededor de la pantalla. Por tanto, estos dispositivos pueden mejorar la experiencia de visión en 3D, pero pueden
40 requerir la generación de un gran número de vistas (algunas decenas de las mismas).

En cuanto a la representación de vídeo en 3D, la gestión de la producción y distribución de un gran número de vistas es una tarea muy exigente. En los últimos años, la comunidad científica ha evaluado la posibilidad de crear un número arbitrariamente grande de vistas intermedias utilizando técnicas de representación basada en imagen de profundidad (DIBR) conocidas, que
45 aprovechan el denominado mapa de profundidad de escena. Estos formatos también se denominan “Vídeo + Profundidad” (V+D), en el que cada vista va acompañada de un mapa de profundidad denso. Un mapa de profundidad denso es una imagen en la que cada píxel en coordenadas planas (x, y), es decir columna, fila, representa un valor de profundidad (z) que

corresponde al píxel de la vista respectiva que presenta las mismas coordenadas. Los valores de los mapas de profundidad pueden calcularse partiendo de las dos vistas obtenidas por una cámara de vídeo estereoscópica, de lo contrario pueden medirse mediante sensores adecuados. Tales valores se representan generalmente utilizando imágenes con 256 niveles de escala de grises, que se comprimen utilizando técnicas convencionales. Las técnicas de representación basada en imagen de profundidad aprovechan el hecho de que, dadas las coordenadas (x, y, z), es decir la posición en el plano de profundidad más la profundidad asociada con cada píxel, es posible proyectar de nuevo el píxel sobre otro plano de imagen en relación con un punto de vista nuevo. El contexto de aplicación más extendido es el de un sistema de cámaras de vídeo estereoscópicas, en el que las dos cámaras de vídeo están situadas horizontalmente a una distancia b entre sus dos centros ópticos, con ejes ópticos paralelos y planos de imagen coplanarios. En una configuración de este tipo, hay una relación simple entre la profundidad z, asociada con un píxel, y la denominada disparidad d, es decir la traslación horizontal que debe aplicarse a un píxel de la imagen de la cámara de vídeo derecha (o izquierda) con el fin de obtener la posición correspondiente en el plano de imagen de la cámara de vídeo izquierda (o derecha). La disparidad puede ser o bien positiva o bien negativa (la traslación a la izquierda o a la derecha), dependiendo de la cámara de vídeo que se tiene en cuenta. Indicando f la longitud focal de las dos cámaras de vídeo, existe la siguiente relación entre profundidad z y disparidad d:

$$d = f b / z.$$

Para más detalle, véase el artículo: Paradiso, V.; Lucenteforte, M.; Grangetto, M., "A novel interpolation method for 3D view synthesis", 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012, vol., n.º, pp. 1, 4, 15-17 de octubre de 2012.

Puesto que, según las hipótesis descritas anteriormente, la disparidad es una función simple de profundidad, el mapa de profundidad y el mapa de disparidad llevan la misma información y, por tanto, son intercambiables. Además, debe señalarse que las imágenes a las que se hace referencia como mapas de profundidad dentro del contexto de MPEG representan los valores de $1/z$, a diferencia de z, mapeados en el intervalo 0-255. A continuación, el término "mapa de profundidad" sólo se utilizará para indicar cualquier representación de profundidad o disparidad.

Debe observarse que la señal de vídeo compuesta por un par de imágenes (izquierda y derecha) y los mapas de profundidad respectivos también se ha elegido como caso de uso por el comité de normalización de MPEG para evaluar las técnicas que se introducirán en las normas de codificación de 3D futuras.

Esto lleva a la necesidad de gestionar eficazmente el almacenamiento, transmisión, recepción y reproducción de señales de televisión que comprenden mapas de profundidad.

Sumario de la invención

Por tanto, el propósito de la presente invención es proporcionar un procedimiento y un dispositivo para generar, almacenar, transmitir, recibir y reproducir mapas de profundidad utilizando las componentes de color de una imagen, que puedan superar las limitaciones inherentes en las soluciones conocidas en la materia.

Tal como se mencionó anteriormente, un mapa de profundidad es adecuado para su representación como imagen en escala de grises, es decir constituida por un único valor por posición (x, y). El término "píxel" se utilizará a continuación para indicar un único elemento (o punto) de una imagen; cada píxel está caracterizado por su posición (x, y) y por valores tales como color o intensidad, que varían en función del sistema de representación que se utiliza. En el campo de la televisión, generalmente se adopta el sistema de representación de píxeles conocido como luminancia (Y) y crominancia (U, V). En general, los mapas de profundidad se

representan como vídeo digital utilizando sólo la componente Y.

La idea básica de la presente invención es aprovechar también las componentes de crominancia U y V para representar los valores del mapa de profundidad, creando por tanto una denominada imagen de “color ficticio”, que permite una mejor compactación de imagen.

5 Una imagen de color ficticio (a continuación en la presente memoria denominada, para abreviar, “imagen de color”) es, en este contexto, una imagen en la que también las componentes de crominancia llevan información útil que, sin embargo, no es información de color sino, en este caso, información de luminancia en relación con los píxeles que se han eliminado de la componente Y.

10 La elección de los píxeles de la matriz de profundidad que van a introducirse en las componentes U y V de la imagen de color de tamaño WxH se realiza de tal manera que se obtiene una correlación espacial entre las componentes Y, U, V, es decir entre la luminancia real y estas componentes ficticias adicionales, que es útil para garantizar el cumplimiento de la operación de los algoritmos de codificación convencionales, por ejemplo MPEG, que suponen
15 la existencia de una correlación entre las tres componentes Y, U, V de los diversos píxeles. Esto también permite utilizar codificadores y decodificadores convencionales para las técnicas de compresión empleadas para las imágenes de color real.

Correlación espacial significa que los valores dispuestos en las componentes Y, U, V pertenecen a píxeles que están espacialmente cerca en el mapa de profundidad.

20 Una solución preferida utiliza el rectángulo de WxH para introducir dos mapas de profundidad, haciendo referencia cada uno a una imagen de vídeo respectiva del par estereoscópico, obteniendo por tanto una única imagen de color.

Otras soluciones preferidas emplean, para los mapas de profundidad, los formatos que se utilizan para imágenes de televisión, conocidos como 4:2:0 o 4:2:2 Y, U, V. En el primer
25 formato, que se utiliza normalmente para distribución, en las filas pares (0, 2, 4, etc.) sólo los píxeles en las columnas pares (0, 2, 4, etc.) contienen información de crominancia, mientras que todos los píxeles en las filas impares contienen sólo información de luminancia: se deduce que, en sentido general, sólo un píxel de cuatro contiene información de crominancia. En el
30 último formato, que se emplea normalmente en la cadena de producción, en todas las filas sólo los píxeles en las columnas pares contienen información de crominancia; se deduce que, en sentido general, un píxel de dos contiene tal información.

Normalmente, se hace referencia a mapas de profundidad con valores representados en 256 niveles, es decir 8 bits por píxel. En una primera serie de ejemplos de soluciones preferidas, que se describirán a continuación, se hará referencia al formato 4:2:0 YUV, mientras que una
35 segunda serie de ejemplos se referirá al formato 4:2:2 YUV.

Entre los diversos sistemas de empaquetado de fotogramas, el formato de baldosa (descrito en el documento WO2011/077343-A1) permite introducir un mapa de profundidad sin sustraer espacio de las imágenes en relación con las dos vistas.

40 Con referencia a la figura 1, en el formato de baldosa se introducen un par de imágenes estereoscópicas L y R en un fotograma de vídeo más grande (C), copiándose de nuevo sin cambios una de las dos imágenes (por ejemplo L) y dividiéndose la otra imagen (por ejemplo R) en tres zonas (R1, R2 y R3). Estas zonas se disponen, en el fotograma compuesto C, en el área que quedó disponible por la primera imagen; todavía quedará un área no utilizada, en la que puede introducirse un mapa de profundidad con una resolución horizontal y vertical
45 reducida a la mitad. Utilizando el procedimiento propuesto por la invención pueden introducirse dos mapas de profundidad, en lugar de uno, en la misma área.

Sin embargo, la presente idea inventiva no se limita a una disposición de empaquetado de fotogramas particular, sino que permite introducir un par de mapas de profundidad en una

imagen rectangular genérica con W columnas y H filas. Naturalmente, el área de WxH puede representar una parte de una imagen más grande utilizada para transportar el vídeo en 3D a través de mecanismos de empaquetado de fotogramas.

5 Como alternativa, el flujo de vídeo estereoscópico puede consistir en diversos flujos de datos, por ejemplo, insertados en un múltiplex de tipo paquete, portando cada uno la información de una o más imágenes o de su(s) mapa(s) de profundidad.

10 En un sistema de producción, distribución y disfrute de televisión que utiliza los diversos aspectos de la presente invención, las etapas llevadas a cabo con la recepción serán inversas a las llevadas a cabo con la transmisión, para el propósito de reconstruir los mapas de profundidad. En el receptor, los valores de los mapas de profundidad, que se asignaron a las componentes Y, U y V de la imagen de color de entrada, se posicionan de nuevo de tal manera que forman los mapas de profundidad iniciales. El procedimiento de reconstrucción puede utilizar posiblemente operaciones de filtrado y/o interpolación conocidas con el fin de estimar cualquier valor de profundidad original que podría haberse descartado debido a la falta de espacio disponible mientras se forma la imagen de color.

20 Un objetivo de la presente invención es proporcionar un procedimiento, y un aparato para el mismo, para generar una imagen de color compuesta por una pluralidad de componentes partiendo de por lo menos un mapa de profundidad, en el que se introduce un primer conjunto de píxeles de dicho por lo menos un mapa de profundidad en la componente de luminancia de dicha imagen de color, y en el que se introducen un segundo y un tercer conjunto de píxeles de dicho por lo menos un mapa de profundidad en las dos componentes de crominancia de dicha imagen de color.

25 Otro objetivo de la presente invención es proporcionar un procedimiento, y un aparato para el mismo, para generar un flujo de vídeo que comprende una secuencia de imágenes de color, en el que por lo menos una parte del fotograma comprende una imagen de color obtenida utilizando un procedimiento para generar una imagen de color tal como se describió anteriormente.

30 Otro objetivo de la presente invención es proporcionar un procedimiento, y un aparato para el mismo, para reconstruir por lo menos un mapa de profundidad partiendo de una imagen de color compuesta por una pluralidad de componentes, en el que se introducen un primer conjunto de píxeles tomados de la componente de luminancia, un segundo conjunto de píxeles tomados de una de las componentes de crominancia y un tercer conjunto de píxeles tomados de la otra componente de crominancia de dicha imagen de color en dicho por lo menos un mapa de profundidad.

35 Otro objetivo de la presente invención es proporcionar un procedimiento, y un aparato para el mismo, para reconstruir dos secuencias de mapas de profundidad partiendo de un flujo de vídeo estereoscópico, en el que durante la reconstrucción de cada mapa de las dos secuencias se aplica un procedimiento para reconstruir por lo menos un mapa de profundidad tal como se describió anteriormente a cada fotograma del flujo.

40 Un objetivo particular de la presente invención es proporcionar un procedimiento y un dispositivo para generar, almacenar, transmitir, recibir y reproducir mapas de profundidad aprovechando las componentes de color de una imagen de color tal como se expone en las reivindicaciones, que forman una parte integral de la presente descripción.

45 La invención se refiere a un procedimiento para generar una imagen de color compuesta por una pluralidad de componentes (Y, U, V), partiendo de al menos un mapa de profundidad o disparidad (DM1, DM2), caracterizado porque se introduce un primer conjunto de píxeles de dicho al menos un mapa de profundidad o disparidad (DM1, DM2) en la componente de luminancia (Y) de dicha imagen de color, y porque se introducen un segundo y tercer conjunto

de píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) en las dos componentes de crominancia (U, V) de dicha imagen de color.

Preferentemente, la elección de las posiciones en las que los píxeles de dichos conjuntos se introducirán en dichas componentes de luminancia (Y) y crominancia (U, V) se realiza, de manera que se garantice la correlación espacial entre la componente de luminancia (Y) y las componentes de crominancia (U, V) de dicha imagen de color.

Preferentemente, la unión de dicho primer, segundo y tercer conjuntos comprende todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).

Preferentemente, dicho primer, segundo y tercer conjuntos comprenden sólo una parte de todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).

Preferentemente, se introducen dos mapas de profundidad o disparidad (DM1, DM2) de tamaño WxH en una imagen de color del mismo tamaño utilizando una técnica de lado a lado o encima y debajo.

Preferentemente, los píxeles de dicho primer y tercer conjuntos que pertenecen a un grupo de píxeles se sustituyen por valores de diferencia en relación con un valor predicho o interpolado, obteniéndose este último a partir de valores de píxeles que pertenecen a dicho primer conjunto.

Preferentemente, comprende la etapa de sustituir los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) que pertenece a un grupo de píxeles contiguos (a, b, c, d) por sumas o diferencias de los mismos (w0, w1, w2, w3), colocando las sumas en la componente de luminancia (Y) y eligiendo sólo una de las diferencias (wd) que se colocará de manera alterna en las componentes de crominancia (U, V).

Preferentemente, comprende la etapa de sustituir los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) que pertenece a un grupo de píxeles contiguos (a, b, c, d) por sumas o diferencias de los mismos (w0, w1, w2, w3), colocando las sumas en la componente de luminancia (Y), la primera de las dos diferencias en una componente de crominancia (U), y la segunda en la otra componente de crominancia (V).

Preferentemente, comprende la etapa de someter a dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) a una transformación de dominio antes de introducir dichos conjuntos de píxeles en dichas componentes de la imagen de color (Y, U, V).

La invención también se refiere a un procedimiento para generar un flujo de vídeo, caracterizado porque comprende una secuencia de imágenes de color obtenidas utilizando el procedimiento descrito anteriormente.

Preferentemente, se utiliza el formato de empaquetado de fotogramas de formato de baldosa, y se introduce dicha imagen de color en la parte de fotograma que está libre de píxeles de imágenes relacionados con las dos vistas estereoscópicas.

La invención también se refiere a un aparato, caracterizado porque comprende unos medios de procesamiento de imagen adaptados para generar un flujo de vídeo que comprende una secuencia de imágenes de color obtenidas utilizando el procedimiento descrito anteriormente.

La invención también se refiere a un procedimiento para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por una pluralidad de componentes (Y, U, V), caracterizado porque se introducen un primer conjunto de píxeles obtenidos partiendo de la componente de luminancia (Y), un segundo conjunto de píxeles obtenidos partiendo de una de las componentes de crominancia (U) y un tercer conjunto de píxeles obtenidos partiendo de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V) en dicho por lo menos un mapa de profundidad o disparidad.

Preferentemente, se toma dicho primer conjunto de píxeles de la componente de luminancia, se toma dicho segundo conjunto de píxeles de una de las componentes de crominancia (U) y se toma dicho tercer conjunto de píxeles de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V).

- 5 Preferentemente, se toma dicho primer conjunto de píxeles de la componente de luminancia, mientras se obtienen dicho segundo o tercer conjuntos realizando una suma de los valores de los píxeles tomados de una o, respectivamente, de la otra de las componentes de crominancia, representando la diferencia en relación con un valor de referencia obtenido mediante interpolación a partir de valores de píxeles que pertenecen a dicho primer conjunto, y dicho
10 valor de referencia.

Preferentemente, se obtienen todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) obteniendo dicho primer, segundo y tercer conjuntos de píxeles.

- Preferentemente, se obtienen algunos píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) obteniendo dicho primer, segundo y tercer conjuntos de píxeles, y se
15 obtienen los píxeles que faltan de los píxeles copiados de nuevo por medio de operaciones de interpolación.

Preferentemente, se obtienen dos mapas de profundidad o disparidad (DM1, DM2), que presentan el mismo tamaño (WxH) que dicha imagen de color (Y, U, V), obteniendo dicho primer, segundo y tercer conjuntos de píxeles.

- 20 La invención también se refiere a un procedimiento para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por una pluralidad de componentes (Y, U, V), caracterizado porque comprende la etapa de combinar linealmente unos valores (w0, w1, wd; w0, w1, w2, w3) tomados de dicha pluralidad de componentes (Y, U, V), en posiciones contiguas, para reconstruir unos valores (a, b, c, d) que
25 van a ser copiados en dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).

Preferentemente, después de que se han obtenido dichos conjuntos de píxeles de dichas componentes (Y, U, V) de la imagen de color, se realiza una transformación de dominio inversa a la llevada a cabo en la fase de generación con el fin de obtener dicho por lo menos un mapa de profundidad (DM1, DM2).

- 30 La invención también se refiere a un procedimiento para reconstruir un flujo de vídeo utilizando el formato de empaquetado de fotogramas denominado formato de baldosa, caracterizado porque comprende unas etapas destinadas a reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color presente en la parte de fotograma que está libre de píxeles de imágenes que se relacionan con las dos vistas estereoscópicas,
35 según el procedimiento descrito anteriormente.

La invención también se refiere a un aparato para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por las componentes Y, U, V, caracterizado porque comprende unos medios para poner en práctica el procedimiento descrito anteriormente.

40

Breve descripción de los dibujos

Objetivos y ventajas adicionales de la presente invención resultarán evidentes a partir de la siguiente descripción detallada de algunos ejemplos de realización de la misma y a partir de los dibujos adjuntos, que se proporcionan sólo a modo de ejemplo no limitativo, en los que:

- 45 la figura 1 muestra el formato de imagen de vídeo estereoscópico denominado "formato de baldosa";

las figuras 2a y 2b muestran dos representaciones gráficas de una primera variante de la invención con respecto a la manera de obtener una imagen de color que contiene dos mapas de profundidad según la solución de lado a lado 4:2:0;

5 la figura 3 muestra un ejemplo de implementación de una técnica de interpolación que puede aplicarse en el lado de recepción para cubrir determinados valores que faltan en el mapa de profundidad reconstruido;

la figura 4 muestra una segunda variante de la invención con respecto a la manera de obtener una imagen de color que contiene dos mapas de profundidad según la solución de lado a lado 4:2:0;

10 la figura 5 muestra imágenes correspondientes a las componentes Y, U, V de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad en el modo de lado a lado 4:2:0;

15 la figura 6 muestra una variante de la invención con respecto a la manera de obtener una imagen de color que contiene dos mapas de profundidad según la solución encima y debajo 4:2:0;

la figura 7 muestra imágenes correspondientes a las componentes Y, U, V de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad en modo encima y debajo 4:2:0;

20 la figura 8 muestra una variante de la invención con respecto a la manera de obtener una imagen de color que contiene dos mapas de profundidad según la solución encima y debajo 4:2:0 con una reducción de las componentes cromáticas;

la figura 9 muestra imágenes correspondientes a las componentes Y, U, V de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad en modo encima y debajo 4:2:0 con una reducción de las componentes cromáticas;

25 la figura 10 muestra una variante de la invención con respecto a la manera de transformar muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad según la solución encima y debajo 4:2:0 con sumas y diferencias medias;

30 la figura 11 muestra imágenes correspondientes a las componentes Y, U, V de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad en modo encima y debajo 4:2:0 con sumas y diferencias medias;

la figura 12 muestra imágenes correspondientes a las componentes Y, U, V de un primer ejemplo de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad tratados previamente con operaciones de transformada de subbanda;

35 la figura 13 muestra una variante de la invención con respecto a la manera de transformar muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad tratados previamente con operaciones de transformada de subbanda;

40 la figura 14 muestra imágenes correspondientes a las componentes Y, U, V de un segundo ejemplo de una imagen de color obtenida disponiendo uno al lado del otro dos mapas de profundidad tratados previamente con operaciones de transformada de subbanda;

la figura 15 muestra una variante de la invención con respecto a la manera de transformar muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad según la solución de lado a lado 4:2:2;

la figura 16 muestra una variante de la invención con respecto a la manera de transformar

muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad según la solución encima y debajo 4:2:2;

la figura 17 muestra una variante de la invención con respecto a la manera de transformar muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad según la solución encima y debajo o de lado a lado 4:2:2 con reducción de las componentes cromáticas;

la figura 18 muestra una variante de la invención con respecto a la manera de transformar muestras de mapas de profundidad para obtener una imagen de color que contiene dos mapas de profundidad tratados previamente con operaciones de transformada de subbanda para el formato 4:2:2;

las figuras 19a y 19b muestran diagramas de bloques de un ejemplo de realización de un sistema completo para regenerar y, respectivamente, reconstruir y visualizar imágenes tridimensionales, que utiliza el formato de empaquetado de fotogramas conocido como formato de baldosa para organizar un par de secuencias de vídeo estereoscópico con sus mapas de profundidad respectivos según la invención.

En los dibujos, los mismos números y letras de referencia identifican los mismos elementos o componentes.

Descripción detallada de unas formas de realización de la invención

A continuación se describirán diversas variantes específicas de implementación de las ideas básicas de la presente invención con respecto a las diversas posibles disposiciones de dos mapas de profundidad en una única imagen de color que presenta un tamaño de $W \times H$ píxeles (que, tal como ya se explicó, pueden introducirse a su vez en un fotograma compuesto más grande que también contiene otras imágenes en relación con el mismo flujo de vídeo estereoscópico).

Considérense dos mapas de profundidad con una resolución $W \times H$, correspondiente a $2 \times W \times H$ muestras en 8 bits, es decir $2 \times W \times H$ bytes. Dicho par de mapas va a introducirse en una única imagen de color “compuesta” 4:2:0 o 4:2:2 YUV que presenta una resolución $W \times H$.

Las diversas formas de realización de la invención descritas a continuación se basan en la utilización de un formato de empaquetado de fotogramas para introducir dos mapas de profundidad en el rectángulo de tamaño $W \times H$. Dos mapas de profundidad pueden disponerse, por ejemplo, en modo encima y debajo, reduciendo su resolución a $W \times (H/2)$ y colocándolos en la parte superior y la parte inferior, respectivamente, de sólo la componente Y, o en modo de lado a lado, reduciendo su resolución a $(W/2) \times H$ y colocándolos en la parte izquierda y la parte derecha, respectivamente, de sólo la componente Y. Después, según la invención, los píxeles descartados se recuperan e introducen en las señales U y V.

Solución de lado a lado 4:2:0.

A continuación se describirán dos técnicas que difieren entre sí en el posicionamiento de las muestras en las componentes de crominancia.

La primera técnica (a continuación en la presente memoria denominada modo A) para obtener una imagen de color partiendo de por lo menos un mapa de profundidad se muestra en la figura 2a. La figura se refiere a un par de filas del mapa de profundidad (por ejemplo el mapa de profundidad DM1 en relación con la vista izquierda de un vídeo estereoscópico) y muestra cómo asociar los valores de profundidad con las componentes Y, U, V de una imagen de color compuesta con un número de columnas reducido a la mitad.

Se utiliza la notación $\mathbf{D}(j,i)$ para indicar el píxel de la fila j y la columna i del mapa de profundidad DM1 y DM2, con $j=0,1,\dots,H-1$ e $i=0,1,\dots,W-1$. Utilizando la técnica descrita en la

figura 2a, el mapa de profundidad debe reorganizarse en una imagen de color constituida por 3 componentes:

- $Y(j,i)$, con $j=0,\dots,H-1$, e $i=0,\dots,W/2-1$
- $U(j,i)$ y $V(j,i)$, con $j=0,\dots,H-1$, e $i=0,\dots,(W/2)-1$, donde los índices j e i toman sólo valores pares debido al submuestreo 4:2:0 YUV.

Con estas notaciones, se obtiene la reorganización de píxeles mostrada en la figura 2a aplicando las siguientes reglas:

- Para cada píxel $D(j,i)$
 1. Si i es par: $Y(j,i/2)=D(j,i)$
 2. Si j es par y el módulo relativo a 4 de i es igual a 1: $V(j,\frac{i-1}{2})=D(j,i)$.
 3. Si j es par y el módulo relativo a 4 de i es igual a 3: $U(j,\frac{i-3}{2})=D(j,i)$.
 4. De lo contrario, se descarta $D(i,j)$.

Aplicando estas reglas, se obtiene una componente Y de resolución $H \times W/2$ que contiene sólo las columnas pares del mapa de profundidad. La componente V recopila los valores de profundidad con un índice de fila par correspondiente a las columnas $4k+1$, siendo k un número entero positivo mayor que o igual a cero, es decir las columnas 1, 5, 9,... del mapa de profundidad se posicionan en la componente V en las columnas 0, 2, 4... del mapa de color. Finalmente, la componente U recopila los valores de profundidad con un índice de fila par correspondiente a las columnas $4k+3$, siendo k un número entero mayor que o igual a cero, es decir las columnas 3, 7, 11,... del mapa de profundidad se posicionan en la componente U en las columnas 0, 2, 4... del mapa de color.

Por tanto, se pierden los valores de profundidad de fila impar y columna impar, es decir los indicados en la figura mediante una cruz o letra "X". Por tanto, para cada mapa de profundidad las muestras están organizadas espacialmente en una imagen de color 4:2:0 YUV nueva con resolución $W/2 \times H$, mientras se mantiene una correlación espacial alta entre las componentes Y , U , V , es decir las imágenes representadas por las componentes Y , U y V representan versiones de la misma imagen con submuestreos alternos. Esto es muy importante porque los algoritmos de compresión de MPEG suponen una correlación espacial entre las componentes U y V ; por tanto, no funcionarían bien si no existiera tal correlación.

Disponiendo una al lado de la otra (horizontalmente) dos imágenes así obtenidas partiendo de un par de mapas de profundidad, se genera finalmente una imagen con resolución $W \times H$, tal como se muestra en la figura 5. En dicha figura, puesto que no pueden utilizarse colores, las tres componentes Y , U y V están representadas por separado.

Para una mayor claridad, la figura 2b muestra una representación esquemática de la manera en la que los dos mapas de profundidad de tamaño $W \times H$, designados $DM1$ y $DM2$, están distribuidos en la imagen de color 4:2:0 (Y , U , V), todavía en el modo A. Los píxeles están representados por cuadrados pequeños con figuras geométricas inscritas; los bloques de 2×2 píxeles de los mapas de profundidad $DM1$ y $DM2$, determinados implícitamente utilizando la técnica mostrada en la figura 2a y las fórmulas anteriores, están enumerados de manera secuencial, en el orden de exploración de filas, o bien sin apóstrofes ($1, 2, \dots, N \times M$, con $N=W/2$ y $M=H/2$) o bien con apóstrofes ($1', 2', \dots, N \times M'$, con $N=W/2$ y $M'=H/2$), dependiendo de si pertenecen a $DM1$ o $DM2$, y cuadrados pequeños iguales corresponden al mismo píxel del bloque inicial. Los píxeles descartados se marcan con una cruz.

Cuando se reorganizan los píxeles de los mapas de profundidad DM1 y DM2, se dividen en bloques de 4 píxeles de tamaño 2x2. Los dos píxeles de la columna izquierda de cada bloque de 2x2 de DM1 (los bloques de 2 píxeles de cuadrados con un círculo de tamaño 1x2) están dispuestos uno al lado del otro en la componente de luminancia Y de la imagen de color ficticio, siguiendo la exploración de filas del mapa de profundidad. De esta manera, se ocupa la mitad izquierda de Y; se hace lo mismo para que DM2 ocupe la mitad derecha.

Todavía utilizando una exploración de filas, el píxel izquierdo superior de dichos bloques de 2x2 de DM1 (bloque de 1 cuadrado con un cuadro inscrito) se coloca de manera alterna en la primera posición disponible en la esquina izquierda superior de las componentes cromáticas V (cuadrado relleno) y U (cuadrado hueco). De esta manera, se ocupan las mitades izquierdas de U y V. Lo mismo se realiza para que los bloques de 2x2 de DM2 ocupen las mitades derechas. Se descartan los píxeles de los bloques de 2x2 de DM1 y DM2 que están ubicados en la esquina derecha inferior (marcado con una cruz en la figura 2b).

Obsérvese la configuración de lado a lado tomada por Y, U y V, y la correspondencia entre estas componentes y las obtenidas de manera experimental utilizando esta realización de la invención (figura 5).

Una vez que se ha recibido la imagen de color, posiblemente después de la codificación, transmisión, recepción y decodificación, los valores de profundidad pueden posicionarse de nuevo invirtiendo las etapas descritas en la figura 2.

En particular, con referencia a la figura 2b, DM1 y DM2 se dividen en bloques de $W/2 \times H/2$ de 2x2 píxeles. Todavía siguiendo una exploración de filas, cada bloque de 1x2 de 2 píxeles de la componente Y de las primeras $W/2$ columnas de la imagen de color ficticio se copia de nuevo en la columna izquierda de los bloques de 2x2 homólogos de DM1, mientras que los bloques de 1x2 de las $W/2$ columnas restantes de Y se copian de nuevo en la columna izquierda de los bloques de 2x2 homólogos de DM2 (cuadrados con círculos en la figura 2).

Cuando se somete a las componentes cromáticas y los mapas a exploración de filas, los píxeles de las primeras $W/4$ columnas de V (cuadrado relleno) y U (cuadrado hueco) se copian de nuevo de manera alterna en la posición derecha superior del bloque homólogo de DM1, tomándolos de la misma posición de fila y columna de U y V. Se hace lo mismo para los píxeles de las $W/4$ columnas derechas restantes de V y U con los bloques de 2x2 de DM2 para reconstruir también el segundo mapa de profundidad. Ambos mostrarán, en esta realización, una cuadrícula que incluye $W \times H/4$ posiciones con valores de píxeles que faltan, es decir faltará un valor de dos en las filas pares del mapa de profundidad. Tales valores pueden interpolarse fácilmente a través de técnicas en sí conocidas, aprovechando la correlación espacial alta que caracteriza a los mapas de profundidad.

La figura 3 muestra un ejemplo de una técnica de interpolación que utiliza un filtro espacial simple (por ejemplo un filtro de media o filtro de mediana) con una máscara de 3x3. Utilizando una máscara de 3x3 centrada alrededor del valor que falta, pueden aprovecharse los 8 valores de profundidad recibidos alrededor del que falta con el fin de estimar el de la posición central. Los experimentos llevados a cabo con un filtro de mediana han demostrado que, cuando se utilizan los mapas de profundidad proporcionados por el comité de MPEG, el mapa de profundidad con resolución $W \times H$ puede reconstruirse con una fidelidad mayor que 50 dB de una relación señal a ruido pico (PSNR).

Esta técnica de reconstrucción mediante interpolación puede utilizarse, naturalmente, en todas las variantes descritas en la presente memoria.

Lo siguiente introducirá una variante de la solución de lado a lado anterior, denominada modo B. La disposición de las componentes U, V de la solución anterior puede mejorarse con respecto a la correlación espacial entre las componentes, tal como se muestra en la figura 4.

Este resultado se alcanza adoptando las siguientes reglas de colocación:

- Para cada píxel $D(j,i)$
 1. Si i es par: $Y(j, i/2) = D(j,i)$
 2. Si j es par y el módulo relativo a 4 de i es igual a 1: $V(j, \frac{i-1}{2}) = D(j,i)$.
 - 5 3. Si j es par y el módulo relativo a 4 de i es igual a 3: $U(j, \frac{i+1}{2}) = D(j,i)$.
 4. De lo contrario, se descarta $D(i,j)$.

En comparación con la solución anterior, ésta garantiza una distancia espacial más corta entre los píxeles asignados de las componentes U , V e Y . En particular, las asignaciones realizadas en las etapas 2 y 3 del algoritmo garantizan una mejor alineación entre las componentes.

10 Aplicando estas reglas, se obtiene una componente Y de resolución $H \times W/2$ que contiene sólo las columnas pares del mapa de profundidad. La componente V recopila los valores de profundidad con un índice de fila par correspondiente a las columnas $4k+1$, siendo k un número entero positivo mayor que o igual a cero, es decir las columnas 1, 5, 9,... del mapa de profundidad original se posicionan en la componente V en las columnas 0, 2, 4,... de la imagen de color. Finalmente, la componente U recopila los valores de profundidad con un índice de fila par correspondiente a las columnas $4k+3$, siendo k un número entero mayor que o igual a cero, es decir las columnas 3, 7,... del mapa de profundidad original se posicionan en la componente U en las columnas 0, 2,... del mapa de color.

20 Debe observarse que, en la etapa 3 del algoritmo, pueden obtenerse valores de índice de columna $\frac{i+1}{2} \geq \frac{W}{2}$ en el borde derecho del mapa de profundidad (por ejemplo el valor de profundidad con $i=11$ en la figura no puede representarse en la imagen de color). En la fase de reconstrucción, estos valores pueden interpolarse utilizando técnicas conocidas en la técnica.

25 De la misma manera, los píxeles en la primera columna de la imagen de color no presentan ningún valor asignado a la componente U (puede asignarse libremente un valor por defecto, normalmente 128).

30 Finalmente, puesto que no pueden proporcionarse imágenes de color, la figura 5 muestra las imágenes correspondientes a las únicas componentes Y , U , V de una imagen de color obtenida disponiendo uno al lado del otro, en modo de lado a lado, dos mapas de profundidad correspondientes a un par de imágenes estereoscópicas, que se obtienen según los procedimientos descritos anteriormente.

Solución encima y debajo 4:2:0.

35 Con el fin de obtener una configuración encima y debajo, es necesario transformar el mapa de profundidad con resolución $W \times H$ en una imagen de color con resolución $W \times H/2$. Un resultado de este tipo se alcanza aplicando los mismos procedimientos descritos anteriormente, intercambiando los papeles de fila y columna.

40 A modo de ejemplo, la figura 6 muestra cómo representar un par de columnas del mapa de profundidad en las componentes YUV reduciendo a la mitad el número de filas con el enfoque denominado modo B en la solución de lado a lado. El modo A descrito para el procedimiento de lado a lado 4:2:0 puede utilizarse de una manera completamente similar intercambiando los papeles de fila y columna y, por tanto, no se describirá adicionalmente por motivos de brevedad.

Superponiendo (verticalmente) dos imágenes con resolución $W \times H/2$ así obtenidas, finalmente se genera una imagen con resolución $W \times H$ que representa el par de mapas de profundidad en modo encima y debajo, tal como se muestra en la figura 7. Puesto que no pueden utilizarse colores, dicha figura muestra las tres componentes Y, U y V por separado.

5 Solución encima y debajo 4:2:0 con reducción de componentes cromáticas.

Las soluciones anteriores permiten que algunos valores de los mapas de profundidad se representen como componentes de crominancia de una imagen de color, mientras que se garantiza una buena coherencia espacial entre las componentes Y, U y V. En el caso de una imagen común (con colores reales), las componentes de crominancia se caracterizan por un contenido energético bajo y, por tanto, por una entropía baja, y pueden comprimirse fácilmente utilizando técnicas conocidas. Por el contrario, las componentes U, V de las soluciones anteriores consisten en píxeles que presentan el mismo contenido energético que la luminancia.

Con el fin de resolver este problema, es posible sustituir los valores de profundidad que van a introducirse en las componentes de crominancia por valores de diferencia en relación con un valor predicho (o interpolado) utilizando los valores de profundidad representados en la componente Y; en otras palabras, puede introducirse un error de predicción en las componentes de crominancia. Puesto que los mapas de profundidad no contienen muchos detalles, la predicción es eficaz y el error de predicción presenta un contenido energético muy bajo y, por tanto, es adecuado para representarse como crominancia.

A modo de ejemplo, en la figura 8, que muestra la configuración encima y debajo en el modo B, las flechas dobles identifican pares de valores de profundidad a través de los que es posible interpolar los valores de profundidad que van a colocarse en las componentes U y V.

En la figura 8, se identifican las muestras de profundidad de la componente Y que pueden utilizarse como predicciones para los valores de profundidad previstos para las componentes U y V según las etapas 2 y 3 de los algoritmos de reposicionamiento anteriores. Utilizando la notación de los algoritmos anteriores, de hecho pueden calcularse los siguientes errores de predicción:

$$D_e(4h+1,2k) = D(4h+1,2k) - \text{redondear}\left(\frac{1}{2}[D(4h,2k) + D(4h+2,2k)]\right)$$

$$D_e(4h+3,2k) = D(4h+3,2k) - \text{redondear}\left(\frac{1}{2}[D(4h+2,2k) + D(4h+4,2k)]\right)$$

donde h, k son índices de número entero mayores que o igual a cero, que van a sustituirse por los valores de profundidad correspondientes en las columnas pares antes de que se coloquen en las componentes V y U, respectivamente. Con el fin de evitar almacenar valores fraccionarios del error de predicción, en las fórmulas anteriores se redondea el valor predicho (el operador de redondeo identifica la operación de redondeo de número entero). Debe señalarse que en este ejemplo se utiliza un predictor simple, que consiste en la media entre las dos muestras de profundidad en las filas por encima y por debajo del píxel en consideración.

El error de predicción D_e es un valor con un signo. En la implementación práctica, se añade preferiblemente un desplazamiento de 128 a tales valores, de modo que pueden representarse en valores de 8 bits sin signo, con la saturación de los valores obtenidos dentro del intervalo [0..255].

Por tanto, los errores de predicción sin signo se representan en las componentes V y U:

$$\tilde{D}_e(4h+1,2k) = D_e(4h+1,2k) + 128$$

$$\tilde{D}_e(4h+3,2k) = D_e(4h+3,2k) + 128$$

Por motivos de simplicidad, en las fórmulas anteriores se ha omitido la etapa de recortar o saturar los valores del intervalo de 0-255.

Con la recepción, antes de la interpolación de las muestras que faltan (identificadas por la letra “X” en los dibujos), por ejemplo según la técnica descrita anteriormente, se reconstruirán los valores de profundidad predichos invirtiendo las fórmulas anteriores:

$$D(4h+1,2k) = \text{redondear}\left(\frac{1}{2}[D(4h,2k) + D(4h+2,2k)]\right) + \tilde{D}_e(4h+1,2k) - 128$$

$$D(4h+3,2k) = \text{redondear}\left(\frac{1}{2}[D(4h+2,2k) + D(4h+4,2k)]\right) + \tilde{D}_e(4h+3,2k) - 128$$

Los errores de redondeo de error de predicción provocan un error de reconstrucción con efectos insignificantes en comparación con los errores de cuantificación obtenidos utilizando un codificador convencional.

La figura 9 muestra un ejemplo de una imagen compuesta obtenida utilizando el sistema descrito anteriormente. Puede observarse el contenido energético bajo de las componentes U y V representado a través de imágenes que, también debido a la aproximación de impresión, parecen ser uniformes. La ventaja de la solución en cuestión, de hecho, es que efectúa una reducción de la amplitud media de las componentes de crominancia, reflejando así más la distribución energética de la señal YUV tradicional, en la que el contenido energético se concentra mayormente en la componente Y, en comparación con las componentes U y V. También debe observarse que las componentes U y V son señales de “diferencia” (para ser precisos, se denominan señales de diferencia de color) y, por tanto, presentan un valor medio nulo, de modo que, en una representación de 8 bits, el valor 128 corresponde a cero, mientras que los valores menores que 128 se consideran negativos y los valores mayores que 128 se consideran positivos. Los algoritmos de compresión de MPEG se basan en tales propiedades de las señales U y V; por tanto, si dichas señales no representan ninguna diferencia, como en la realización anterior, los algoritmos de compresión de MPEG pueden resultar menos eficaces.

El enfoque descrito anteriormente se ha ejemplificado, en el caso del formato encima y debajo, a través de la utilización de un procedimiento de interpolación con sólo dos muestras, pero puede extenderse obviamente a todas las soluciones descritas hasta ahora y a otros procedimientos de interpolación.

Solución encima y debajo 4:2:0 con sumas y diferencias medias.

Según una variante adicional que permite reducir la contribución energética de las componentes U y V, los valores de los mapas de profundidad se sustituyen por sumas y diferencias medias, que se representan utilizando las componentes de luminancia y crominancia, respectivamente. Las ventajas de esta variante son las mismas que las de la solución anterior, porque en este caso también las señales que representan diferencias se introducen en U y V.

Las muestras de cada bloque de 2x2 del mapa de profundidad se transforman en sumas y diferencias medias tal como se muestra en la figura 10 y a continuación, donde a, b, c, d representan los valores de profundidad en un bloque de 2x2 del mapa de profundidad. Los coeficientes w0, w1 son valores medios y deben redondearse al número entero más cercano (o truncarse al número entero inferior), de modo que entonces puedan representarse en muestras de 8 bits sin signo.

Tiene lugar el mismo proceso para los coeficientes de diferencia w2, w3; además, puesto que estos últimos presentan un signo, sus valores absolutos se saturan a 128, de modo que su valor puede representarse en muestras de 8 bits añadiendo un desplazamiento de 128.

La figura 10 también muestra cómo reconstruir las muestras de profundidad a partir de las

sumas y diferencias medias, tal como se cita a continuación.

$$\begin{array}{ll}
 w_0 = (a+c)/2 & a = (w_0 + w_2) \\
 w_1 = (b+d)/2 & b = (w_1 + w_3) \\
 w_2 = (a-c)/2 & c = (w_0 - w_2) \\
 w_3 = (b-d)/2 & d = (w_1 - w_3)
 \end{array}
 \quad \text{---} \rightarrow$$

Debido a los redondeos adoptados, la reconstrucción de los valores de profundidad a, b, c, d será exacta a menos de un error de +/-1.

- 5 Con el fin de obtener una imagen de color de dimensiones reducidas, por ejemplo $W \times H/2$ en el caso de encima y debajo, es necesario descartar un coeficiente de 4. Entonces los 3 coeficientes seleccionados pueden disponerse en las componentes Y, U, V según una de las técnicas descritas anteriormente.

- 10 Esto se consigue manteniendo las dos sumas medias w_0 y w_1 , que se introducirán en la componente Y, y moviendo o bien la diferencia $w_d = w_2$ o bien la diferencia $w_d = w_3$ a U, V, dependiendo de cuál de los dos valores w_2 o w_3 garantiza el menor error de reconstrucción según las siguientes fórmulas.

- 15 El error de reconstrucción introducido por la selección o bien de w_2 o bien de w_3 puede evaluarse tal como sigue. Se calculan el error medio $e(w_d = w_2)$, obtenido en el caso de $w_d = w_2$, y el error medio $e(w_d = w_3)$, obtenido en el caso de $w_d = w_3$; el coeficiente w_d que se seleccionará entonces para la transmisión es el que minimiza el error tal como sigue:

$$\begin{aligned}
 e(w_d = w_2) &= |w_1 + w_2 - b| + |w_1 - w_2 - d| \\
 e(w_d = w_3) &= |w_0 + w_3 - a| + |w_0 - w_3 - c|
 \end{aligned}$$

$$w_d = \begin{cases} w_2 & \text{si } e(w_d = w_2) < e(w_d = w_3) \\ w_3 & \text{de lo contrario} \end{cases}$$

Con el fin de obtener una imagen de color con resolución $W \times H/2$, puede seguirse el mismo algoritmo ya descrito.

- 20 Para cada índice de fila $j=0, 2, 4 \dots$ y cada índice de columna $i=0, 2, 4, \dots$
- Se calculan los coeficientes w_0, w_1, w_2, w_3 del bloque de 2×2 en la posición j, i del mapa de profundidad
 - Se determina la mejor diferencia media w_d
 - Se asignan las sumas medias a la componente Y según las fórmulas $Y(j/2, i) = w_0$ e $Y(j/2, i+1) = w_1$
 - Si el resto de la división del índice j entre 4 es igual a 0, se asigna la diferencia media según la fórmula $U(j/2, i) = w_d$
 - De lo contrario (si el índice j no es un múltiplo de 4), $V(j/2, i) = w_d$

- 30 La figura 11 muestra un ejemplo de una imagen de color obtenida utilizando el sistema descrito anteriormente en la configuración encima y debajo, obtenida introduciendo dos mapas de profundidad con sumas y diferencias medias. Puede observarse que también en este caso las

componentes de crominancia recopilan un contenido energético bajo debido a la dinámica muy baja asociada con el valor w_d .

Con la reconstrucción, la técnica con sumas y diferencias medias no requerirá un filtro de interpolador de 3×3 , puesto que será suficiente aplicar las fórmulas inversas que permiten reconstruir los valores iniciales a , b , c , d de los coeficientes w_0 , w_1 y w_d :

$$a \approx w_0 + w_d$$

$$b \approx w_1 + w_d$$

$$c \approx w_0 - w_d$$

$$d \approx w_1 - w_d$$

El error de reconstrucción realizado depende de la elección de la diferencia media w_d y del error de representación de número entero de todos los coeficientes. Las pruebas han mostrado que tales errores son insignificantes en mapas de profundidad de referencia con relaciones señal a ruido pico de más de 50 dB.

Solución con transformada de subbanda.

La técnica anterior basada en medias y diferencias de los valores de profundidad puede generalizarse adoptando un enfoque de subbanda. Utilizando una transformada de dominio, por ejemplo del tipo de DCT de 2×2 , es decir una transformada de Walsh-Hadamard, o sólo un nivel de cualquier transformada de ondícula conocida en la bibliografía, la imagen constituida por el mapa de profundidad DM1 (y DM2) se divide en cuatro subbandas con una resolución reducida a la mitad tanto en las filas como en las columnas, por tanto con una resolución $W/2 \times H/2$. Tales subbandas se indican generalmente de izquierda a derecha y de arriba abajo como subbandas LL, LH, HL y HH. La figura 12 muestra un ejemplo de una transformada de subbanda obtenida aplicando la transformada de Walsh-Hadamard por separado a las filas y columnas de un mapa de profundidad. Puede observarse que la transformada concentra la mayor parte de la energía de la señal de vídeo en la subbanda LL (en la parte izquierda superior). Para imágenes que contienen pocos detalles, tal como en el caso de mapas de profundidad, la subbanda HH recopila muy poca energía. También puede observarse que los coeficientes de las subbandas HL, LH y HH presentan un signo, y que debe utilizarse un desplazamiento de 128 con el fin de representar el valor 0, como en el caso de las diferencias medias descritas para la solución anterior. Cuando se desee representar los coeficientes de transformada en muestras de 8 bits sin signo, también será necesario realizar redondeos apropiados que pueden variar dependiendo del tipo de transformada utilizada.

Puesto que la resolución del mapa de profundidad debe reducirse en presencia de un dominio transformado, pueden eliminarse algunos coeficientes: la mejor elección es eliminar la banda HH entera de ambos mapas de profundidad, introduciendo de ese modo una pérdida de calidad con la reconstrucción. La reconstrucción se consigue naturalmente aplicando la transformada inversa, donde se supone que los coeficientes eliminados son cero.

Con el fin de construir una única imagen YUV con una resolución $W \times H$ partiendo de las subbandas LL', LH' y HL' del mapa de profundidad izquierdo y de las subbandas LL'', LH'' y HL'' del mapa de profundidad derecho, se sigue el diagrama mostrado en la figura 13, en el que DM1' y DM2' indican las transformadas de Hadamard de los dos mapas de profundidad DM1 y DM2. La figura muestra cómo disponer los coeficientes de las subbandas LL', LH' y LL'', LH'' en los cuatro cuadrantes de la componente de luminancia Y. En particular, se ha elegido asociar la mitad superior de la componente Y con las subbandas LL' y LH' del mapa de profundidad izquierdo, y la mitad inferior con las subbandas LL'' y LH'' del mapa de profundidad derecho. Para completar la reconstrucción de la imagen YUV, será necesario disponer de nuevo los coeficientes $W/2 \times H/2$ restantes de las subbandas HL' y HL'' en las componentes U y V, sin crear bordes falsos debido a una desalineación de las tres componentes. La figura 13 muestra

cómo obtener el resultado deseado. La subbanda HL' del mapa de profundidad izquierdo se subdivide en 4 versiones submuestreadas por un factor de 2 en ambas direcciones, tal como se muestra en la figura 13; en la práctica, los coeficientes de cada bloque de 2x2 se colocan en 4 imágenes submuestreadas. Por tanto, se obtienen 4 imágenes de la subbanda HL', designadas HL_a', HL_b', HL_c' y HL_d', con una resolución de W/4xH/4. Las dos primeras imágenes obtenidas se disponen una al lado de la otra en la parte superior de la componente U, ocupando un área equivalente a muestras de H/4xW/2. Las dos imágenes restantes de la subbanda HL' se copian (una al lado de la otra) en la parte superior de la componente V. Finalmente se adopta el mismo procedimiento para la subbanda HL'' del mapa de profundidad derecho, pero las imágenes de W/4xH/4 se copian en la mitad inferior de las componentes U y V. Este enfoque permite mantener una coherencia espacial entre la componente Y y las componentes U, V con una resolución reducida a la mitad.

En la figura 14 se muestra un ejemplo de una imagen de color obtenida utilizando la transformada de Walsh-Hadamard. Además en este caso, la contribución energética asociada con las componentes U y V es muy baja, lo que es beneficioso en términos de compresibilidad de la señal de vídeo.

Síntesis de dos mapas de profundidad en una cadena de producción de televisión (formato YUV 4:2:2).

En la cadena de producción de televisión, normalmente se utiliza el formato YUV 4:2:2. En comparación con el caso del formato 4:2:0, las soluciones para empaquetar dos mapas en una única imagen de color son más simples, porque la información de crominancia (U, V) está presente en el 50% de los píxeles. Por tanto, en general, están disponibles 16 bits por píxel (ocho para Y y ocho para U, V) y, por tanto, es posible introducir los dos mapas en la imagen de color sin ninguna pérdida de información y, por tanto, sin la necesidad de interpolar los píxeles que faltan con la recepción.

Una primera solución puede consistir en transmitir un mapa en Y y el otro mapa en U y V. Sin embargo, esta solución no es óptima porque los algoritmos de compresión de vídeo se basan habitualmente en la suposición de que la luminancia y la crominancia están correlacionadas entre sí, tal como en el caso de una imagen de color real, mientras que en la solución en cuestión, Y pertenece a una imagen y U, V pertenecen a una imagen diferente. Por tanto, también en el caso de 4:2:2, será mejor disponer de nuevo las muestras de tal manera que se garantice la construcción de una imagen con componentes Y, U, V correlacionadas, utilizando mecanismos similares a los empleados para el caso de YUV 4:2:0.

Solución de lado a lado 4:2:2.

La creación de la imagen de color puede entenderse fácilmente examinando la figura 15. En la primera fila de píxeles, los píxeles en las columnas impares (1, 3, 5 ...) del mapa de profundidad original se transfieren a U y V de los píxeles en las columnas pares de la imagen de color, de una manera completamente similar a la figura 2a. Puesto que, a diferencia del caso de 4:2:0, los píxeles de las columnas pares (0, 2, 4...) de la imagen de color presentan crominancia, sucede lo mismo en la segunda fila de píxeles que en la primera fila, es decir los píxeles en las columnas impares del mapa de profundidad original se transfieren a U y V de los píxeles en las columnas pares de la imagen de color. Por tanto, ya no habrá ningún píxel que falte (los marcados con una cruz en la figura 2a) para su reconstrucción mediante interpolación con la recepción.

Lo mismo se aplica al modo B. En este caso también, en todas las filas (par e impar) se produce lo que se muestra en la figura 4 para la fila 0 de píxeles y, por tanto, ya no habrá ningún píxel que falta para su reconstrucción mediante interpolación con la recepción.

Solución encima y debajo 4:2:2.

A modo de ejemplo, en la figura 16 se muestra una posible manera de crear la imagen de color. Los píxeles en las filas impares, que se descartan, se reasignan a U y V de los píxeles en las filas pares, por ejemplo tal como se indica por las flechas. Puede efectuarse una reasignación también de otras maneras sustancialmente equivalentes, que no se describirán en la presente memoria por motivos de brevedad.

Soluciones de lado a lado y encima y debajo con reducción de las componentes cromáticas, formato 4:2:2.

Las soluciones anteriores permiten que algunos valores de los mapas de profundidad se representen como componentes de crominancia de una imagen de color, mientras que se garantiza una buena coherencia espacial entre las componentes Y, U y V. En el caso de una imagen común (con colores reales), las componentes de crominancia se caracterizan por un contenido energético bajo y, por tanto, por una entropía baja, y pueden comprimirse fácilmente utilizando técnicas conocidas. Por el contrario, las componentes U, V de las soluciones anteriores consisten en píxeles que presentan el mismo contenido energético que la luminancia. Con el fin de resolver este problema, es posible sustituir los valores de profundidad que van a introducirse en las componentes de crominancia por valores de diferencia en relación con un valor predicho (o interpolado) utilizando los valores de profundidad representados en la componente Y; en otras palabras, puede introducirse un error de predicción en las componentes de crominancia. Puesto que los mapas de profundidad no están muy detallados, la predicción es eficaz y el error de predicción presenta un contenido energético muy bajo y, por tanto, es adecuado para representarse como crominancia.

A modo de ejemplo, en la figura 17 las flechas dobles identifican dos valores de profundidad a través de los que es posible interpolar los valores de profundidad que van a colocarse en las componentes U y V en la configuración encima y debajo de la figura 16. En la representación de la figura 16, las muestras en las filas impares se interpolan partiendo de las muestras en las filas por encima y por debajo. Tales valores también están disponibles con la recepción de la imagen de color, puesto que se ubican en la componente Y. En este punto, es posible calcular e introducir, como valores de las componentes U, V (según la disposición mostrada por las flechas delgadas en la figura 16), los errores de predicción realizados por el interpolador. Tales valores se representarán como números enteros sin signo utilizando las mismas técnicas ya descritas. Los errores de predicción transportados por las componentes U, V permitirán finalmente, con la recepción, reconstruir los valores de profundidad en las filas impares aplicando la corrección correspondiente a los valores interpolados desde la componente Y.

Solución encima y debajo 4:2:2 con sumas y diferencias medias.

Además en el caso de 4:2:2, una variante adicional que permite reducir la contribución energética de las componentes U y V proporciona la sustitución de los valores de los mapas de profundidad por sumas y diferencias medias, que se representan utilizando las componentes de luminancia y crominancia, respectivamente. Las ventajas de esta variante son las mismas que las de la solución anterior, porque en este caso también se introducen en U y V señales que representan diferencias. Las muestras de cada bloque de 2x2 del mapa de profundidad se transforman en sumas y diferencias medias tal como se muestra en la figura 10, donde a, b, c, d representan los valores de profundidad en un bloque de 2x2 del mapa de profundidad.

La única diferencia con respecto al caso de 4:2:0 es que con la solución 4:2:2 ya no es necesario descartar un coeficiente de cuatro, puesto que ambos coeficientes de diferencia pueden introducirse en U y V. Por tanto, con la recepción, el mapa se reconstruirá sin errores, excepto una pérdida de precisión de +/-1 inducida por la representación de números enteros sin signo de las sumas y diferencias medias.

Solución con transformada de subbanda, formato 4:2:2.

Aplicando sólo un nivel a cualquier transformada de subbanda, la imagen formada por el mapa de profundidad se divide en cuatro subbandas con resolución reducida a la mitad tanto en las filas como en las columnas, es decir con una resolución $W/2 \times H/2$. Tales subbandas se indican generalmente de derecha a izquierda y de arriba abajo como subbandas LL, LH, HL y HH.

5 Cuando se utiliza el formato YUV 4:2:2, pueden representarse las transformadas de dos mapas de profundidad en una única imagen de color utilizando el procedimiento mostrado en la figura 18, donde DM1' y DM2' indican las transformadas de subbanda de los dos mapas de profundidad DM1 y DM2. A diferencia de la técnica similar propuesta para el formato YUV 4:2:0 (véase la figura 13), en este caso es posible establecer las cuatro subbandas de DM1' y DM2'.

10 En la solución propuesta en la presente memoria a modo de ejemplo, se forma una imagen de color encima y debajo. Las componentes U, V se construyen de tal manera que mantienen la correlación espacial con las imágenes correspondientes de Y, introduciendo las subbandas LH y HH submuestreadas por columna; en particular, las columnas pares se introducen en la componente U, mientras que las columnas impares se introducen en la componente V.

15 Introduciendo sólo un mapa.

Debe observarse que la presente invención también puede aplicarse al caso de la presencia de sólo un mapa de profundidad. En este caso, las técnicas propuestas en la presente memoria permiten construir una imagen de color con una resolución reducida a la mitad (por fila o columna) en comparación con la del mapa de profundidad original.

20 Sistema de generación y reconstrucción completa.

Las soluciones propuestas en la presente memoria pueden utilizarse para enviar un par de secuencias de vídeo estereoscópico, junto con los mapas de profundidad respectivos, a un sistema de visualización, permitiendo así la síntesis de cualquier punto de vista intermedio.

25 Las figuras 19a y 19b muestran un ejemplo de un sistema de generación, reconstrucción y visualización completa que utiliza el formato de empaquetado de fotogramas conocido como formato de baldosa, separándose el lado de generación del lado de reconstrucción y visualización.

En particular, en el sistema propuesto, se fusionan dos mapas de profundidad con resolución $W \times H$, designados en el dibujo como mapa de profundidad izquierdo DL (profundidad izquierdo) y mapa de profundidad derecho DR (profundidad derecho) (obtenidos posiblemente mediante submuestreo a 4:1 de dos mapas de profundidad de las dos vistas con dimensiones de $2W \times 2H$) mediante un dispositivo denominado "fusionador de profundidad" en una única imagen de color YUV 4:2:0 con resolución $W \times H$ adoptando una de las técnicas propuestas en la presente memoria. Entonces, el dispositivo multiplexor de formato de baldosa construye el fotograma compuesto de formato de baldosa partiendo de dos imágenes estereoscópicas con una resolución de $2W \times 2H$ y el par respectivo de mapas de profundidad con resolución $W \times H$. El par de mapas de profundidad ocupan exactamente el espacio que dejó disponible el formato de empaquetado de fotogramas en la esquina derecha inferior (véase la figura 1). Por tanto, se obtiene una única imagen con resolución de $3W \times 3H$, que puede transportar todas las componentes de la señal estereoscópica. Suponiendo que $W=640$ y $H=360$, se producirá una imagen de formato de baldosa con una resolución de 1920×1080 , que es compatible con el formato de vídeo de alta definición conocido como 1080p. Por tanto, la señal de vídeo constituida por las imágenes de formato de baldosa se vuelve compatible con cualquier infraestructura para codificar, transmitir o transportar una señal de vídeo de 1920×1080 .

45 Un codificador conocido (por ejemplo MPEG4 o HEVC) puede comprimir eficazmente el flujo de vídeo estereoscópico producido por el multiplexor de formato de baldosa para la transmisión en un canal de comunicación, posiblemente después de otros tratamientos (multiplexación con otro contenido, modulación, etc.). Alternativamente, puede almacenarse el flujo codificado mediante una unidad de escritura en cualquier medio de almacenamiento (memoria de

semiconductor volátil o no volátil, medio optoelectrónico, etc.) para un disfrute futuro.

En el lado de receptor (véase la figura 19b), se decodifica el flujo de vídeo estereoscópico recibido del canal de comunicación o leído desde un medio de almacenamiento y se envía al demultiplexor de formato de baldosa, que extrae el par de imágenes estereoscópicas L y R con resolución de $2W \times 2H$ y la imagen de color producida con la generación mediante el fusionador de profundidad. Una unidad de división denominada divisor de profundidad extrae de dicha imagen el par de mapas de profundidad de color con resolución $W \times H$ DL y DR, que puede expandirse a su tamaño de $2W \times 2H$ original mediante un dispositivo de interpolación conocido simple denominado muestreador ascendente de profundidad. Un dispositivo sintetizador de vistas calcula las vistas intermedias entre R y L para una visualización a través de un dispositivo autoestereoscópico.

Debe observarse que la descripción de los aparatos para la generación (1500) y la reconstrucción (1550) de un flujo de vídeo estereoscópico obtenido aplicando la invención a mapas de profundidad, puede someterse a muchas variaciones. Por ejemplo, las funciones del fusionador de profundidad y del multiplexor de formato de baldosa pueden llevarse a cabo en realidad por el mismo dispositivo físico; en el lado de reconstrucción, la síntesis y la visualización de las imágenes estereoscópicas pueden llevarse a cabo mediante el mismo aparato o mediante dos aparatos distintos.

Operaciones de reconstrucción.

Para reconstruir los mapas de profundidad según la invención, se llevan a cabo operaciones que son inversas a aquéllas para generar las imágenes de color según las variantes descritas anteriormente.

En particular, el procedimiento para reconstruir al menos un mapa de profundidad (DM1, DM2) partiendo de una imagen de color compuesta por una pluralidad de componentes (Y, U, V) proporciona introducir en el mapa de profundidad un primer conjunto de píxeles tomados de la componente de luminancia (Y), un segundo conjunto de píxeles tomados de una de las componentes de crominancia (U) y un tercer conjunto de píxeles tomados de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V).

Preferiblemente, copiando de nuevo los conjuntos de píxeles primero, segundo y tercero, se obtienen todos los píxeles de dicho al menos un mapa de profundidad; o, copiando de nuevo los conjuntos de píxeles primero, segundo y tercero, se obtienen algunos píxeles del mapa de profundidad, y se obtienen los píxeles que faltan sometiendo a los píxeles copiados de nuevo a operaciones de interpolación.

En algunas posibles variantes, los conjuntos segundo y tercero (U y V) contienen diferencias con respecto a los valores interpolados, de modo que con la reconstrucción será necesario calcular en primer lugar los valores interpolados y luego aplicar las correcciones obtenidas de U y V.

En algunas posibles variantes, el procedimiento comprende la etapa de sustituir los valores (w_0 , w_1 , w_d) tomados de los conjuntos de píxeles primero, segundo y tercero de la imagen de color por sumas y diferencias de los mismos que permiten reconstruir los valores (a, b, c, d) que van a copiarse en un bloque de píxeles del mapa de profundidad (DM1, DM2).

En posibles variantes adicionales, el procedimiento proporciona realizar una transformación de dominio que es inversa a la llevada a cabo en la fase de generación con el fin de obtener dicho mapa de profundidad (DM1, DM2), después de copiar de nuevo los conjuntos de píxeles de las componentes (Y, U, V) de la imagen de color.

El aparato para reconstruir al menos un mapa de profundidad (DM1, DM2) partiendo de una imagen de color compuesta por las componentes Y, U, V según la invención comprende

medios para introducir en el al menos un mapa de profundidad o disparidad un primer conjunto de píxeles tomados de la componente de luminancia (Y), un segundo conjunto de píxeles tomados de una de las componentes de crominancia (U) y un tercer conjunto de píxeles tomados de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V).

- 5 En dicho aparato, preferiblemente, los conjuntos de píxeles primero, segundo y tercero cubren todos los píxeles de por lo menos un mapa de profundidad o disparidad.

Como alternativa, los conjuntos de píxeles primero, segundo y tercero cubren algunos píxeles de por lo menos un mapa de profundidad o disparidad, y se introducen píxeles obtenidos sometiendo a los píxeles copiados de nuevo a operaciones de interpolación en los píxeles
10 restantes.

El aparato puede reconstruir dos mapas de profundidad DM1, DM2 que presentan el mismo tamaño WxH, y en la imagen de color Y, U, V la componente de luminancia Y presenta el mismo tamaño WxH que los dos mapas de profundidad o disparidad DM1, DM2.

- 15 Preferiblemente, el aparato comprende medios para realizar una suma o una diferencia de los valores de los píxeles seleccionados de un grupo de píxeles con respecto a un valor de referencia predicho o interpolado, obteniéndose este último valor de valores de píxeles que pertenecen al primer conjunto, con el fin de obtener los píxeles de los conjuntos segundo y tercero.

- 20 Preferiblemente, el aparato comprende medios para sustituir los valores w0, w1, wd tomados de los conjuntos de píxeles primero, segundo y tercero de la imagen de color por sumas y diferencias de los mismos, para reconstruir los valores a, b, c, d que van a copiarse en un bloque de píxeles de por lo menos un mapa de profundidad DM1, DM2.

- 25 Preferiblemente, el aparato comprende medios para realizar una transformación de dominio que es inversa a la llevada a cabo en la fase de generación con el fin de obtener dicho al menos un mapa de profundidad DM1, DM2, después de copiar de nuevo los conjuntos de píxeles de dichas componentes Y, U, V de la imagen de color.

- 30 En el caso en el que los mapas de profundidad se introducen en un fotograma compuesto de formato de baldosa, en todas las variantes descritas anteriormente se obtiene un formato universal para generar, transportar y reproducir contenido en 3D en todos los tipos de visualizaciones presentes y futuras.

En el caso de un dispositivo de reproducción en 2D, el procesador de vídeo del dispositivo descartará simplemente las imágenes R y los mapas de profundidad DM1 y DM2 que pueden estar presentes en el fotograma de vídeo decodificado, y visualizará, sujeto a ajuste a escala, sólo la secuencia de imágenes L en un dispositivo de visualización asociado.

- 35 Lo mismo se aplica al caso de un dispositivo de reproducción en 3D en el que el usuario ha activado el modo de visualización en 2D.

- Un dispositivo de reproducción en 3D en el que se ha activado el modo de visualización en 3D puede mostrar dos comportamientos diferentes, dependiendo de si el reproductor ofrece o no al usuario la posibilidad de ajustar (disminuir) la profundidad de la escena. En este último caso, el
40 procesador de vídeo utilizará las dos secuencias de imágenes L (izquierda) y R (derecha) para generar el efecto tridimensional. En el primer caso, el procesador de vídeo utilizará los mapas de profundidad (uno o dos) incluidos en los fotogramas compuestos asociados con cada par de imágenes estereoscópicas R y L para generar vistas intermedias entre L y R, obteniendo de ese modo imágenes tridimensionales que presentan una profundidad variable, menor de la que
45 puede alcanzarse con L y R.

El último caso de aplicación se representa mediante reproductores autoestereoscópicos, que

requieren un número de vistas muy grande (algunas decenas de las mismas) para generar el efecto tridimensional para los espectadores situados en diferentes puntos en el espacio delante de la pantalla. En este caso, el procesador de vídeo utilizará los mapas de profundidad (uno o dos) incluidos en los fotogramas compuestos, junto con las propias imágenes L y R, para sintetizar una serie de otras imágenes.

En conclusión, el procesador de vídeo del dispositivo de reproducción puede comprender medios adaptados para enviar a la pantalla dos secuencias de imágenes, consistiendo al menos una en imágenes sintetizadas partiendo de al menos una de las vistas transmitidas y de al menos un mapa de profundidad. En este caso, también comprende preferiblemente medios adaptados para dar al espectador la posibilidad de elegir secuencias de imágenes en relación con puntos de vista más o menos cercanos, para poder variar la percepción de profundidad.

El procesador de vídeo del dispositivo de reproducción también puede comprender medios adaptados para generar imágenes adicionales correspondientes a vistas adicionales, de modo que los espectadores situados en diferentes puntos en el espacio puedan ver diferentes secuencias de imágenes a través de una pantalla autoestereoscópica asociada.

Ninguno de los formatos propuestos hasta ahora ofrece tal flexibilidad y variedad de utilización, mientras que al mismo tiempo todavía garantiza una calidad de reproducción muy buena en términos de equilibrio de la resolución horizontal y vertical y de asignación de resolución apropiada a las imágenes estereoscópicas y a los mapas de profundidad asociados.

Las operaciones de reconstrucción descritas anteriormente pueden llevarse a cabo en parte en el dispositivo receptor y en parte en el dispositivo de visualización (reproductor). La presente invención puede implementarse ventajosamente a través de un programa informático que comprende medios de codificación para ejecutar una o más etapas del procedimiento. Por tanto, se entiende que el alcance de protección se extiende a dicho programa informático así como a medios legibles por ordenador que comprenden un mensaje grabado, comprendiendo dichos medios legibles por ordenador medios de codificación de programa para implementar una o más etapas del procedimiento cuando dicho programa se ejecuta por un ordenador.

El ejemplo de realización descrito anteriormente puede someterse a variaciones sin apartarse del alcance de protección de la presente invención, incluyendo todos los diseños equivalentes conocidos para un experto en la materia.

La presente descripción ha abordado formas de realización de la invención en las que el sistema de muestreo empleado para la imagen de color es 4:2:0 o 4:2:2, que son los más extendidos. No obstante, puede aplicarse de la misma manera a cualquier otro sistema de muestreo, tal como 4:4:4, 4:1:1, 4:1:0, por ejemplo, realizando, siempre que sea necesario, operaciones para submuestrear la imagen a niveles de escala de grises iniciales.

Sin embargo, los elementos y características mostrados en las diversas formas de realización preferidas pueden combinarse entre sí sin apartarse del alcance de protección de la presente invención.

A partir de la descripción anterior, los expertos en la materia podrán producir el objetivo de la invención sin introducir ningún detalle de construcción adicional.

REIVINDICACIONES

1. Procedimiento para generar una imagen de color compuesta por una pluralidad de componentes (Y, U, V), partiendo de al menos un mapa de profundidad o disparidad (DM1, DM2), caracterizado porque se introduce un primer conjunto de píxeles de dicho al menos un
5 mapa de profundidad o disparidad (DM1, DM2) en la componente de luminancia (Y) de dicha imagen de color, y porque se introducen un segundo y tercer conjunto de píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) en las dos componentes de crominancia (U, V) de dicha imagen de color.
2. Procedimiento según la reivindicación 1, caracterizado porque la elección de las posiciones
10 en las que los píxeles de dichos conjuntos se introducirán en dichas componentes de luminancia (Y) y crominancia (U, V) se realiza, de manera que se garantice la correlación espacial entre la componente de luminancia (Y) y las componentes de crominancia (U, V) de dicha imagen de color.
3. Procedimiento según la reivindicación 1 o 2, caracterizado porque la unión de dicho primer,
15 segundo y tercer conjuntos comprende todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).
4. Procedimiento según la reivindicación 1 o 2, caracterizado porque dicho primer, segundo y tercer conjuntos comprenden sólo una parte de todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).
- 20 5. Procedimiento según cualquiera de las reivindicaciones anteriores, caracterizado porque se introducen dos mapas de profundidad o disparidad (DM1, DM2) de tamaño WxH en una imagen de color del mismo tamaño utilizando una técnica de lado a lado o encima y debajo.
6. Procedimiento según cualquiera de las reivindicaciones anteriores, caracterizado porque los
25 píxeles de dicho primer y tercer conjuntos que pertenecen a un grupo de píxeles se sustituyen por valores de diferencia en relación con un valor predicho o interpolado, obteniéndose este último a partir de valores de píxeles que pertenecen a dicho primer conjunto.
7. Procedimiento según cualquiera de las reivindicaciones 1 a 4, caracterizado porque
30 comprende la etapa de sustituir los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) que pertenece a un grupo de píxeles contiguos (a, b, c, d) por sumas o diferencias de los mismos (w0, w1, w2, w3), colocando las sumas en la componente de luminancia (Y) y eligiendo sólo una de las diferencias (wd) que se colocará de manera alterna en las componentes de crominancia (U, V).
8. Procedimiento según cualquiera de las reivindicaciones 1 a 4, caracterizado porque
35 comprende la etapa de sustituir los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) que pertenece a un grupo de píxeles contiguos (a, b, c, d) por sumas o diferencias de los mismos (w0, w1, w2, w3), colocando las sumas en la componente de luminancia (Y), la primera de las dos diferencias en una componente de crominancia (U), y la segunda en la otra componente de crominancia (V).
9. Procedimiento según cualquiera de las reivindicaciones anteriores, caracterizado porque
40 comprende la etapa de someter a dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) a una transformación de dominio antes de introducir dichos conjuntos de píxeles en dichas componentes de la imagen de color (Y, U, V).
10. Procedimiento para generar un flujo de vídeo, caracterizado porque comprende una
45 secuencia de imágenes de color obtenidas utilizando un procedimiento según una de las reivindicaciones anteriores.
11. Procedimiento para generar un flujo de vídeo según la reivindicación 10, caracterizado

porque se utiliza el formato de empaquetado de fotogramas de formato de baldosa, y porque se introduce dicha imagen de color en la parte de fotograma que está libre de píxeles de imágenes relacionados con las dos vistas estereoscópicas.

5 12. Aparato, caracterizado porque comprende unos medios de procesamiento de imagen adaptados para generar un flujo de vídeo que comprende una secuencia de imágenes de color obtenidas utilizando un procedimiento según una de las reivindicaciones anteriores.

10 13. Procedimiento para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por una pluralidad de componentes (Y, U, V), caracterizado porque se introducen un primer conjunto de píxeles obtenidos partiendo de la componente de luminancia (Y), un segundo conjunto de píxeles obtenidos partiendo de una de las componentes de crominancia (U) y un tercer conjunto de píxeles obtenidos partiendo de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V) en dicho por lo menos un mapa de profundidad o disparidad.

15 14. Procedimiento según la reivindicación 13, caracterizado porque se toma dicho primer conjunto de píxeles de la componente de luminancia, se toma dicho segundo conjunto de píxeles de una de las componentes de crominancia (U) y se toma dicho tercer conjunto de píxeles de la otra componente de crominancia (V) de dicha imagen de color (Y, U, V).

20 15. Procedimiento según la reivindicación 13, caracterizado porque se toma dicho primer conjunto de píxeles de la componente de luminancia, mientras se obtienen dicho segundo o tercer conjuntos realizando una suma de los valores de los píxeles tomados de una o, respectivamente, de la otra de las componentes de crominancia, representando la diferencia en relación con un valor de referencia obtenido mediante interpolación a partir de valores de píxeles que pertenecen a dicho primer conjunto, y dicho valor de referencia.

25 16. Procedimiento según las reivindicaciones 13, 14, 15, caracterizado porque se obtienen todos los píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) obteniendo dicho primer, segundo y tercer conjuntos de píxeles.

30 17. Procedimiento según las reivindicaciones 13, 14, 15, caracterizado porque se obtienen algunos píxeles de dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2) obteniendo dicho primer, segundo y tercer conjuntos de píxeles, y se obtienen los píxeles que faltan de los píxeles copiados de nuevo por medio de operaciones de interpolación.

18. Procedimiento según cualquiera de las reivindicaciones 14 a 17, caracterizado porque se obtienen dos mapas de profundidad o disparidad (DM1, DM2), que presentan el mismo tamaño (WxH) que dicha imagen de color (Y, U, V), obteniendo dicho primer, segundo y tercer conjuntos de píxeles.

35 19. Procedimiento para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por una pluralidad de componentes (Y, U, V), caracterizado porque comprende la etapa de combinar linealmente unos valores (w_0 , w_1 , w_d ; w_0 , w_1 , w_2 , w_3) tomados de dicha pluralidad de componentes (Y, U, V), en posiciones contiguas, para reconstruir unos valores (a, b, c, d) que van a ser copiados en dicho por lo menos un mapa de profundidad o disparidad (DM1, DM2).

40 20. Procedimiento según cualquiera de las reivindicaciones 13 a 19, caracterizado porque, después de que se han obtenido dichos conjuntos de píxeles de dichas componentes (Y, U, V) de la imagen de color, se realiza una transformación de dominio inversa a la llevada a cabo en la fase de generación con el fin de obtener dicho por lo menos un mapa de profundidad (DM1, DM2).

21. Procedimiento para reconstruir un flujo de vídeo utilizando el formato de empaquetado de fotogramas denominado formato de baldosa, caracterizado porque comprende unas etapas

destinadas a reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color presente en la parte de fotograma que está libre de píxeles de imágenes que se relacionan con las dos vistas estereoscópicas, según cualquiera de las reivindicaciones 13 a 20.

- 5 22. Aparato para reconstruir por lo menos un mapa de profundidad o disparidad (DM1, DM2) partiendo de una imagen de color compuesta por las componentes Y, U, V, caracterizado porque comprende unos medios para poner en práctica el procedimiento según cualquiera de las reivindicaciones 13 a 21.

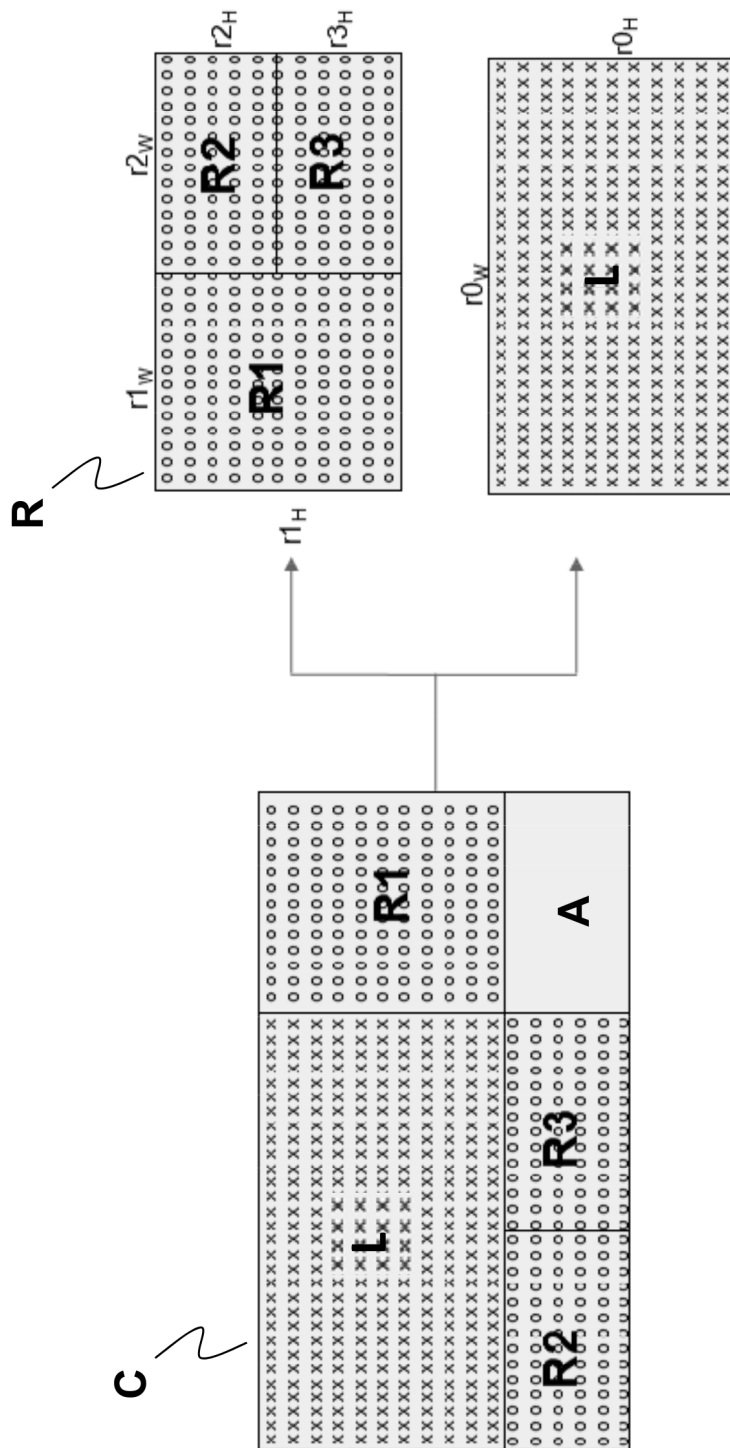


FIG. 1

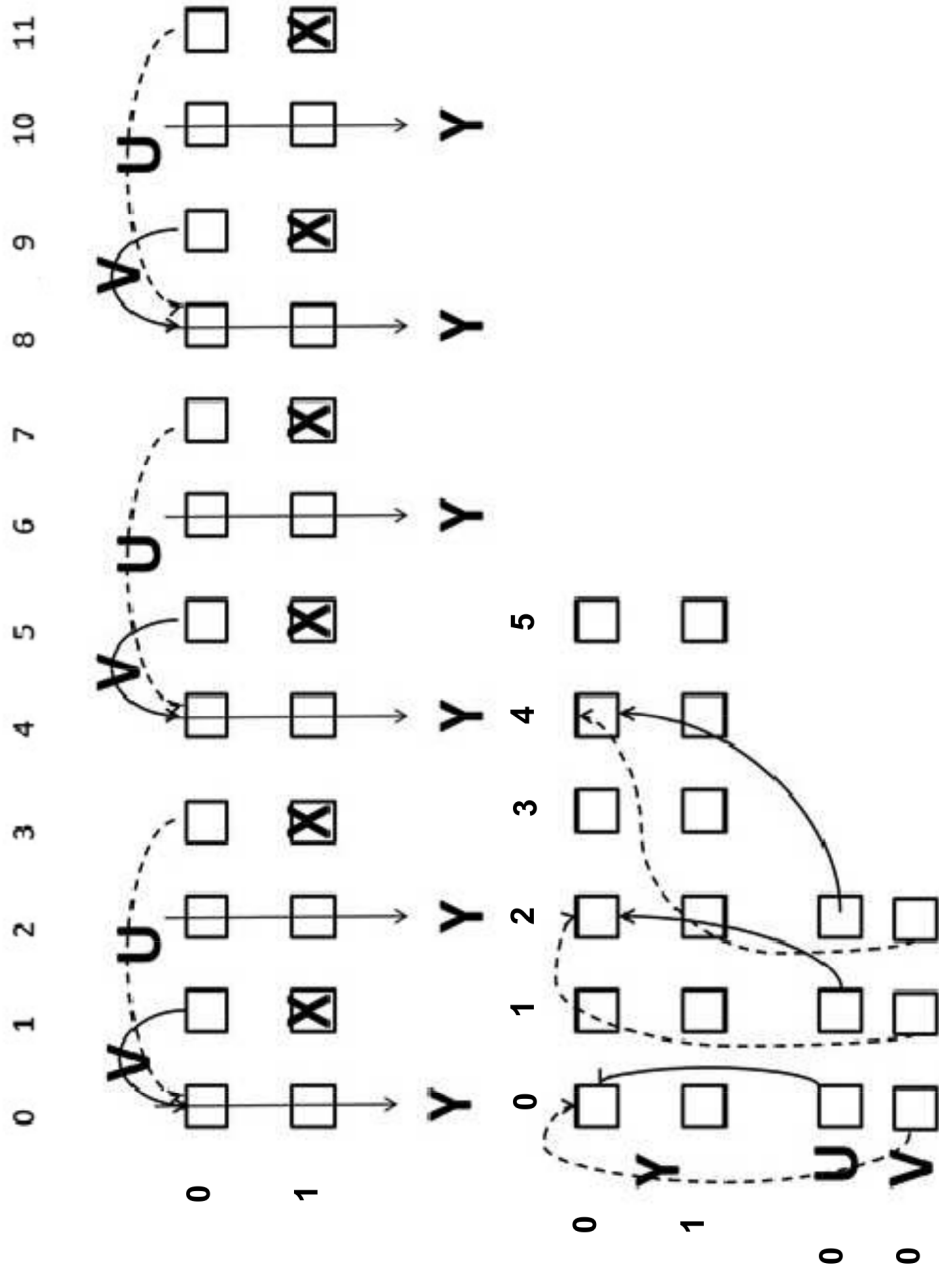


FIG. 2a

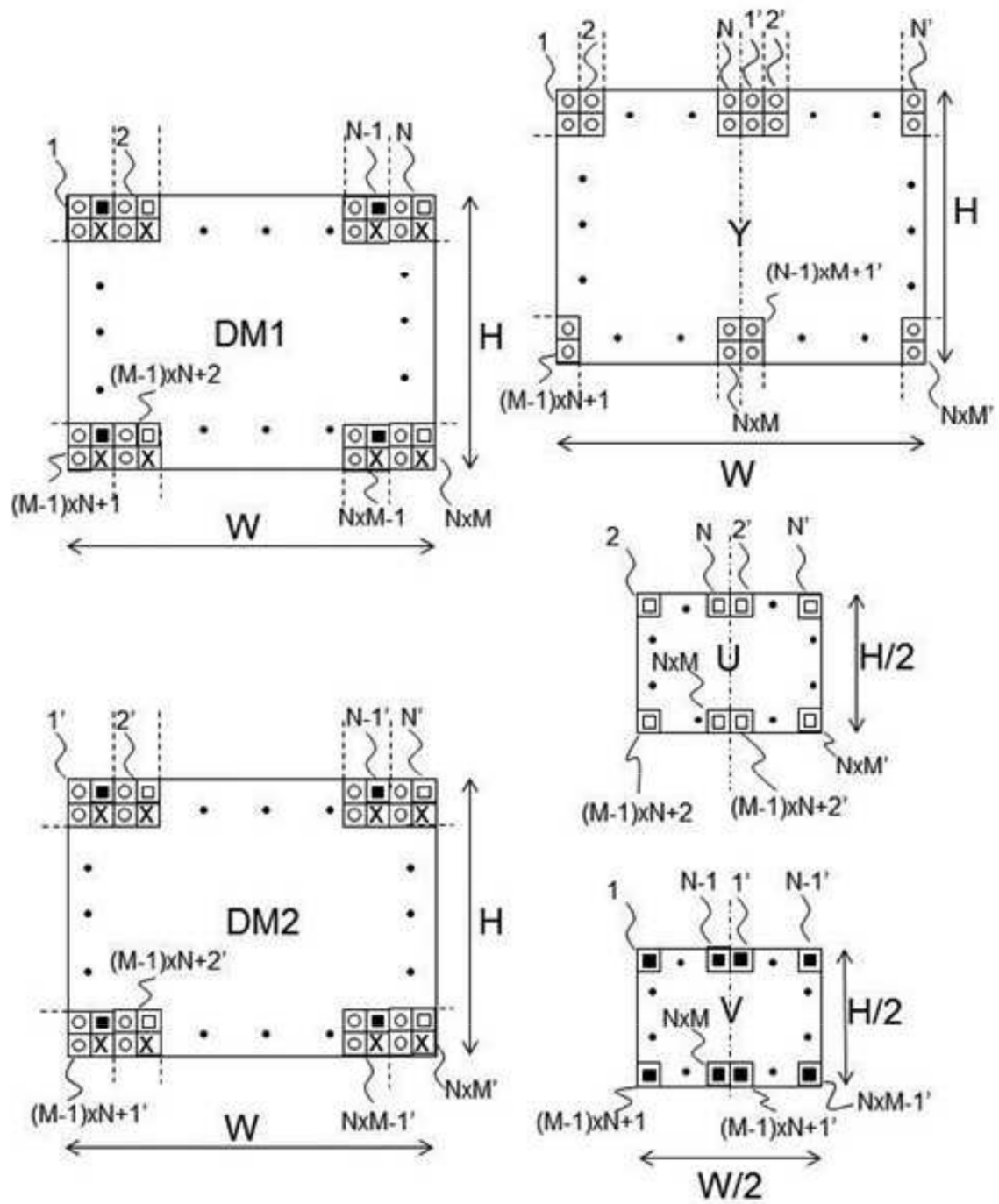


FIG. 2b

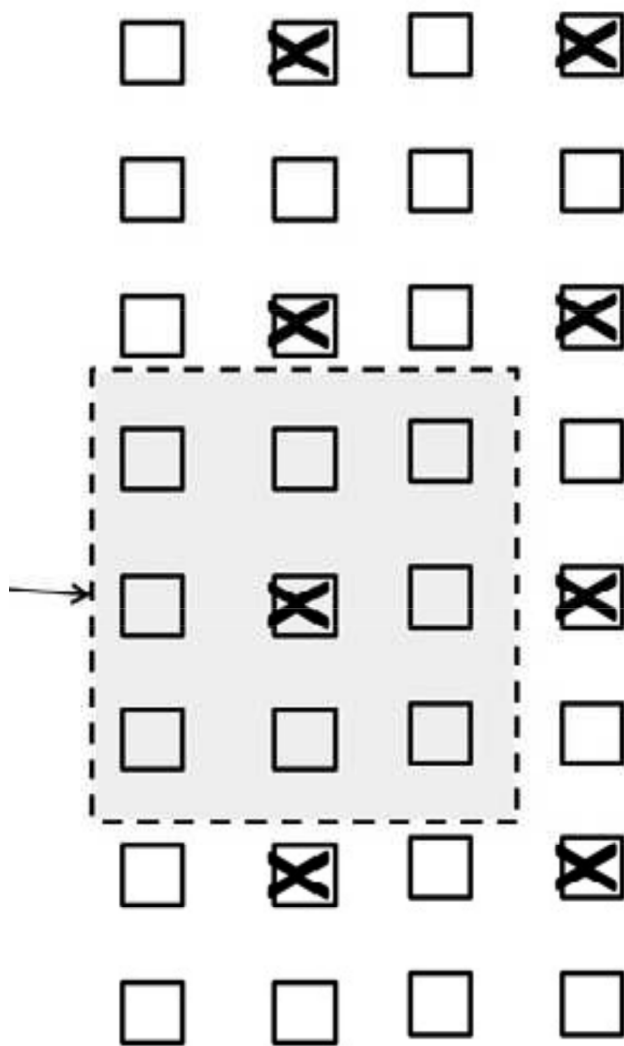


FIG. 3

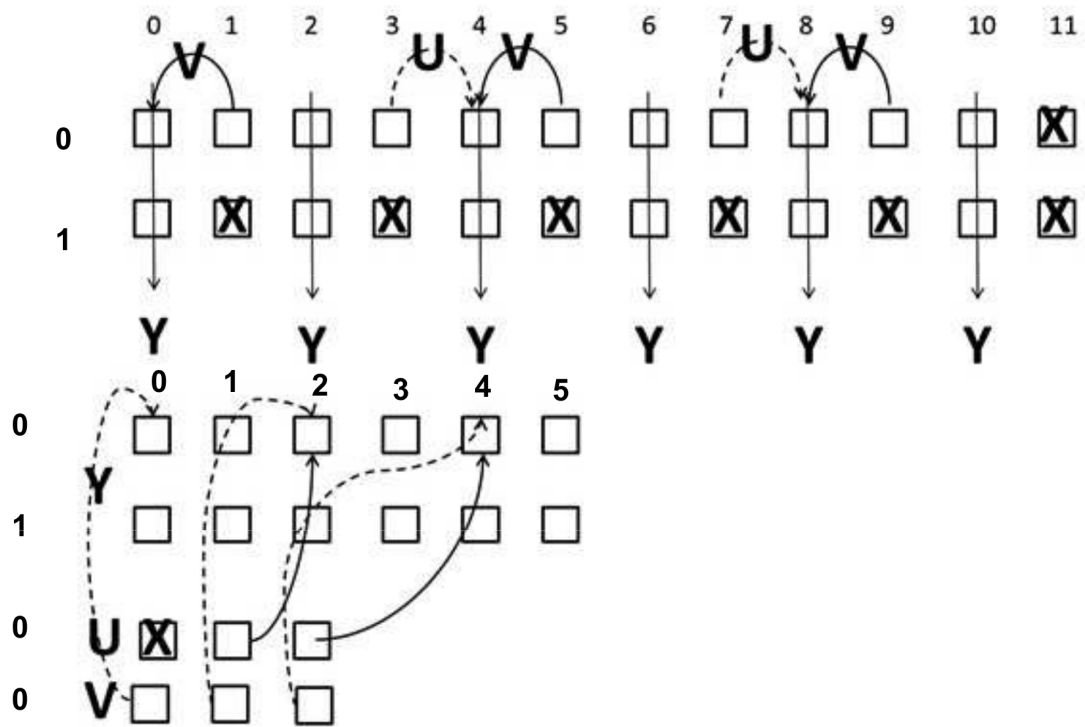


FIG. 4

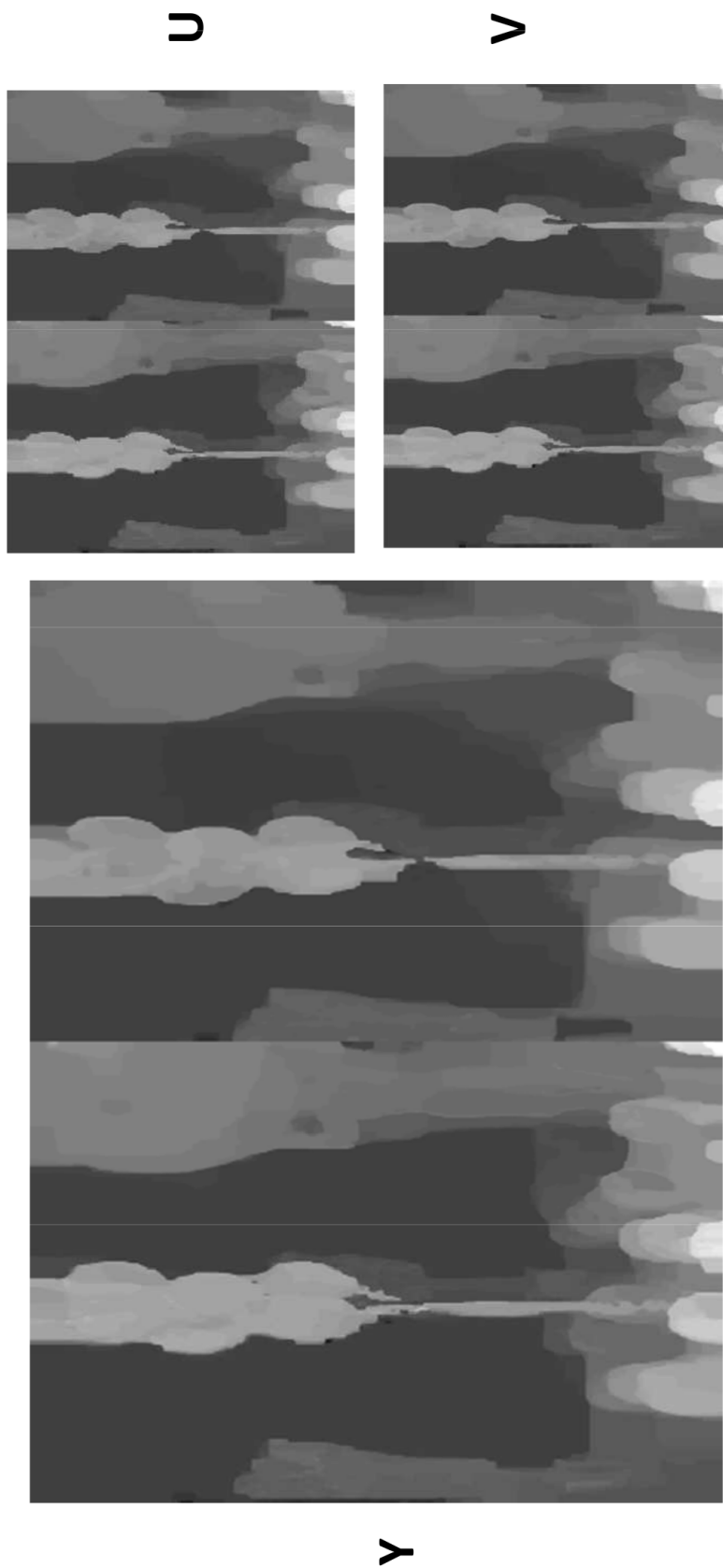


FIG. 5

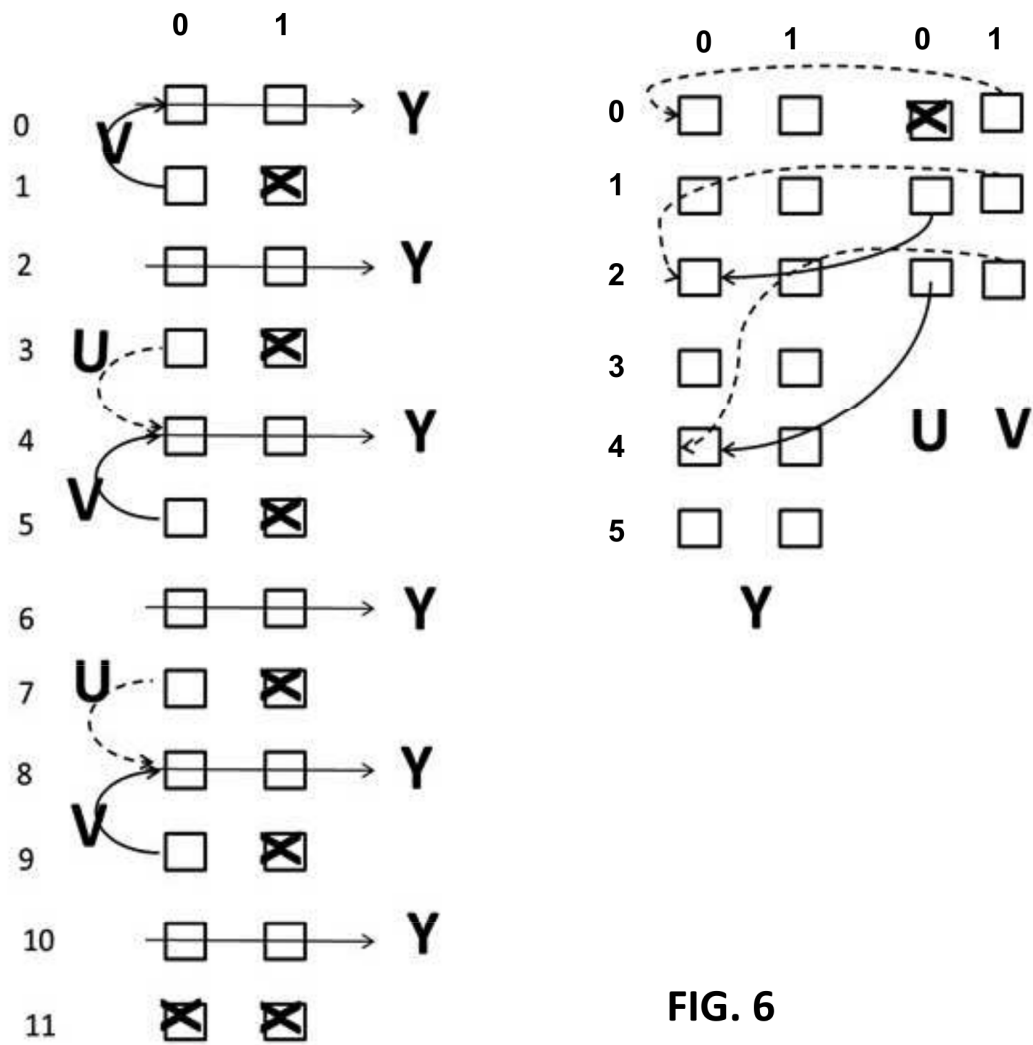


FIG. 6

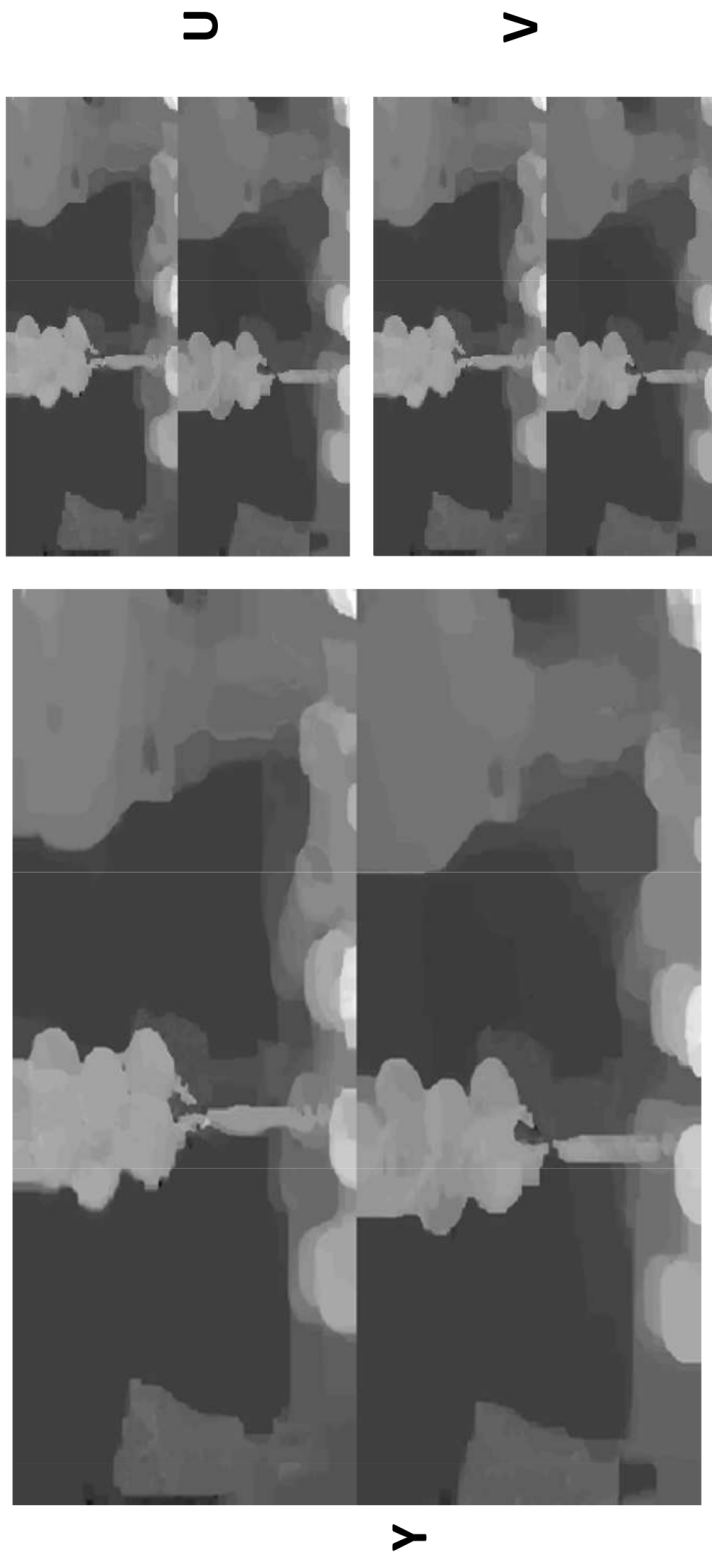


FIG. 7

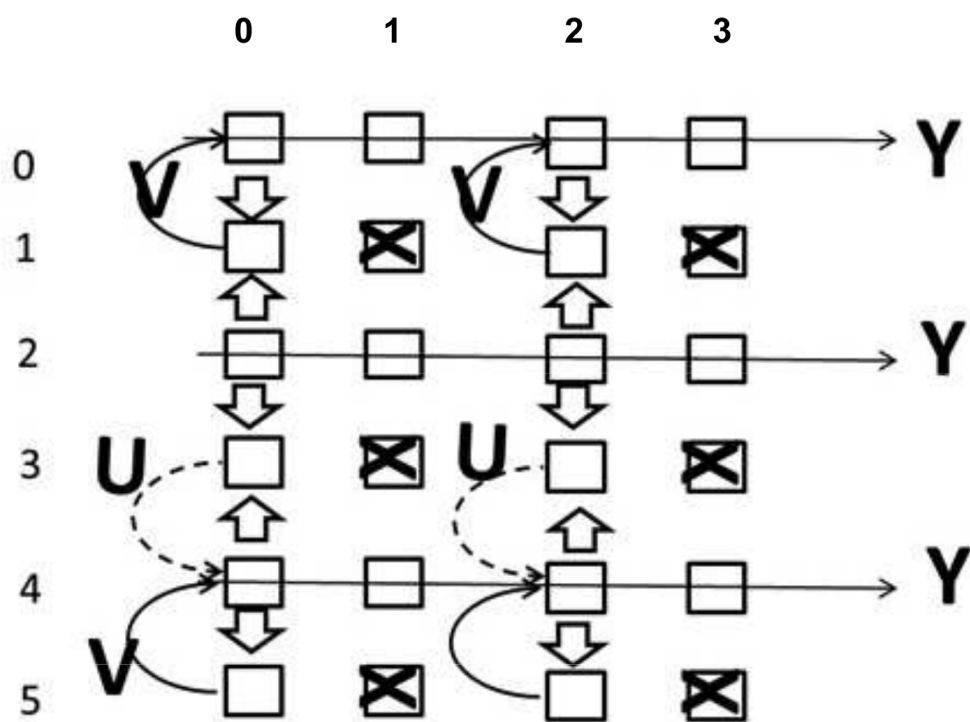


FIG. 8

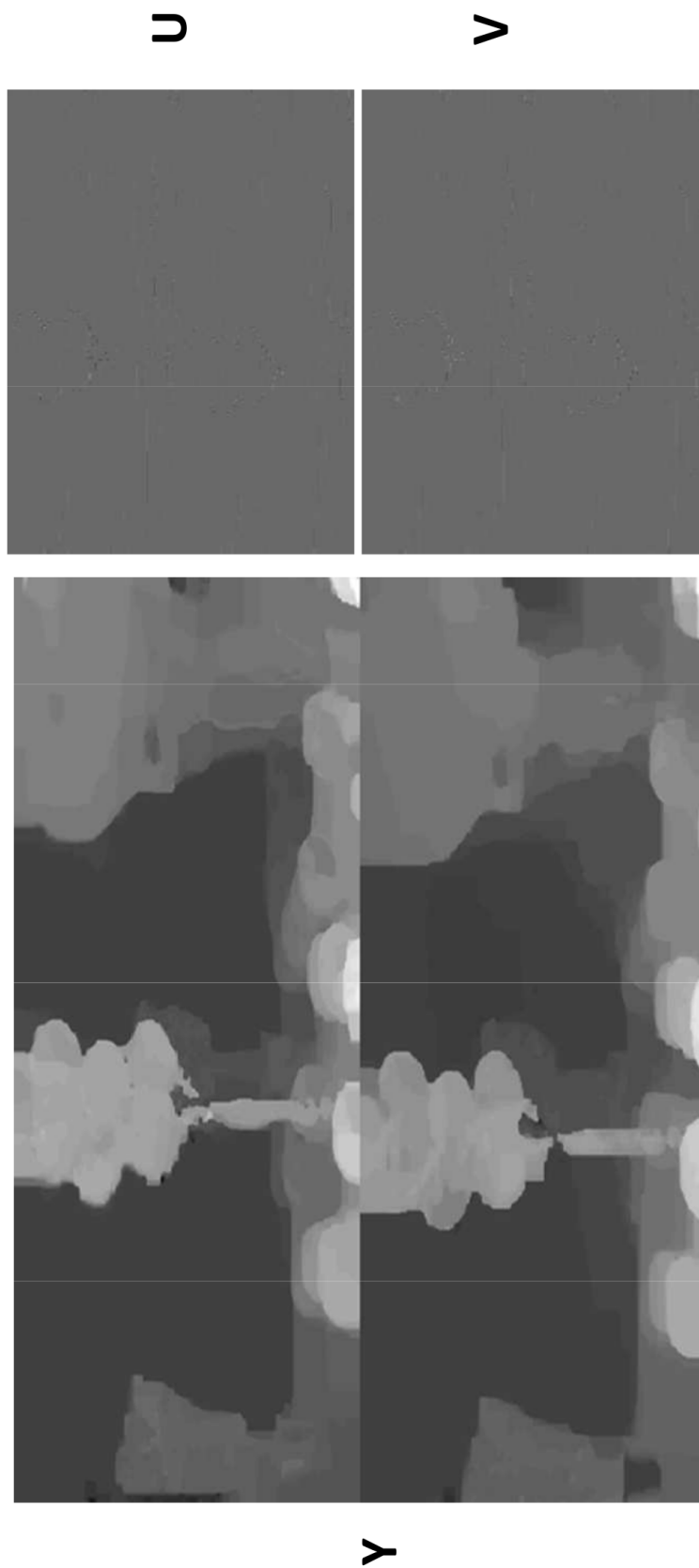


FIG. 9

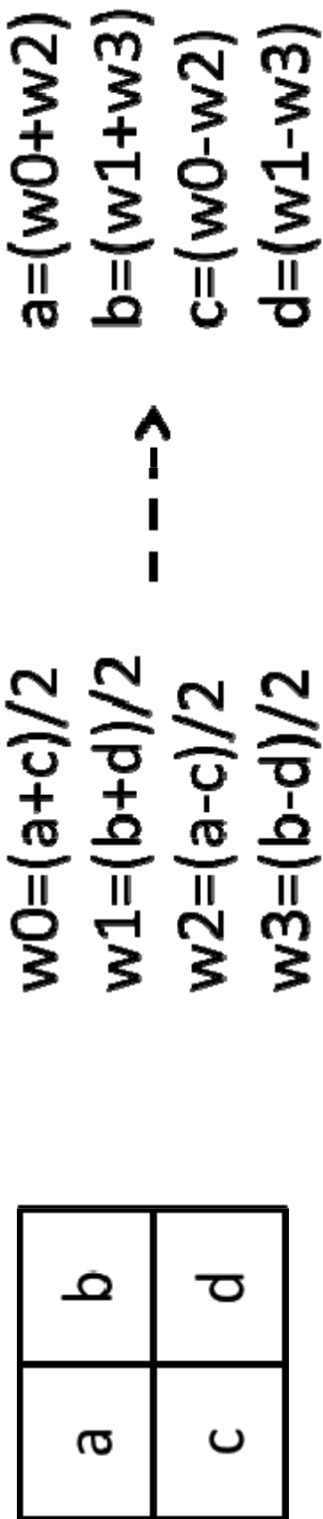


FIG. 10

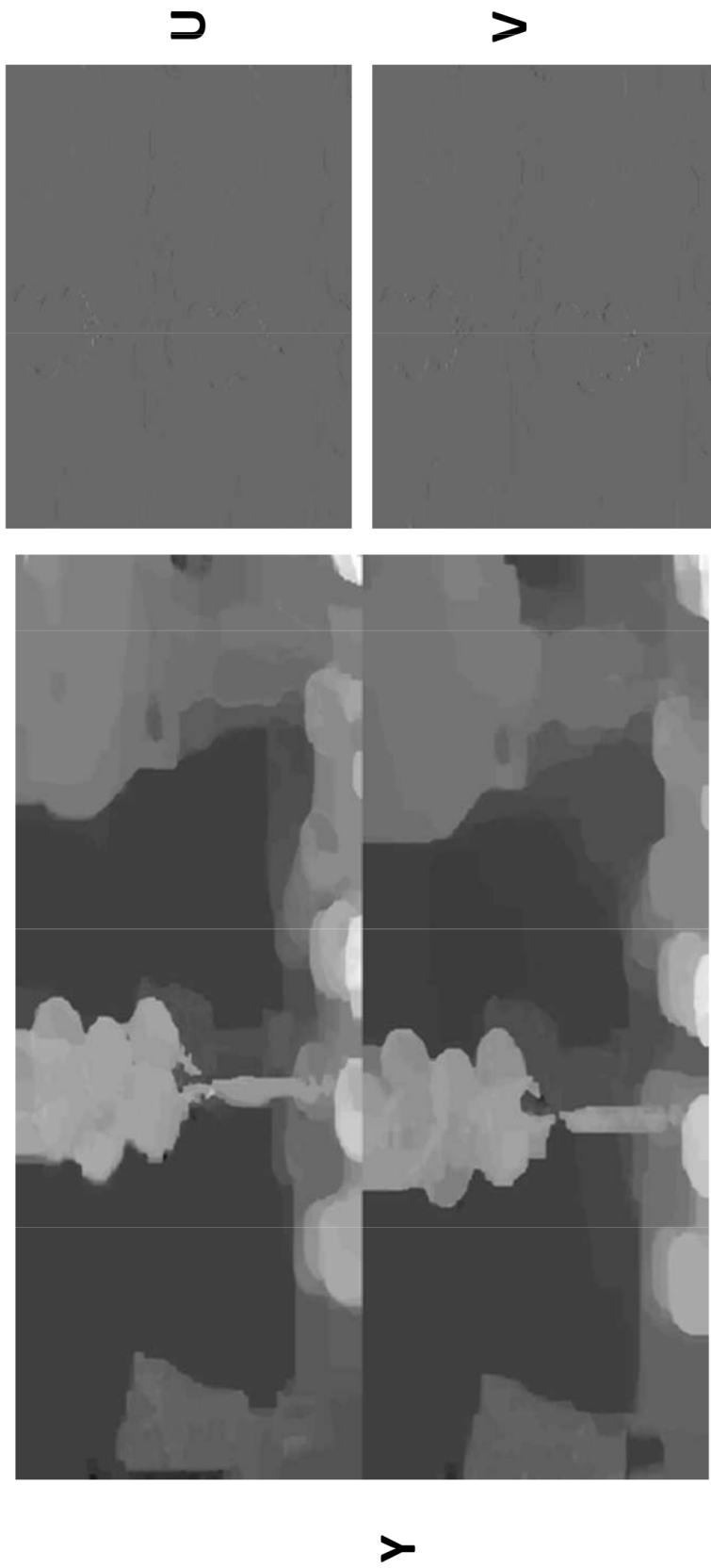


FIG. 11

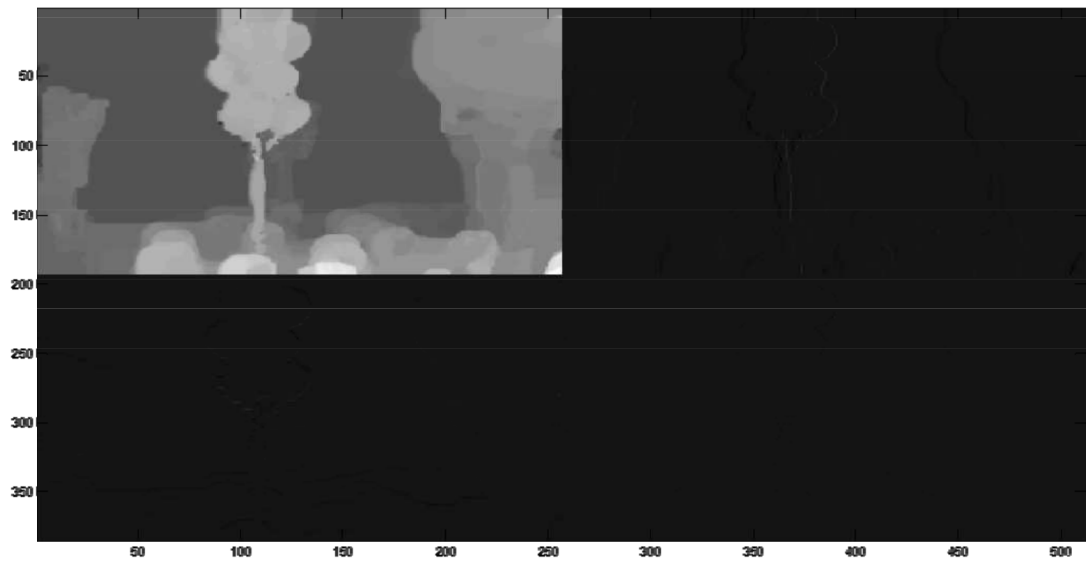
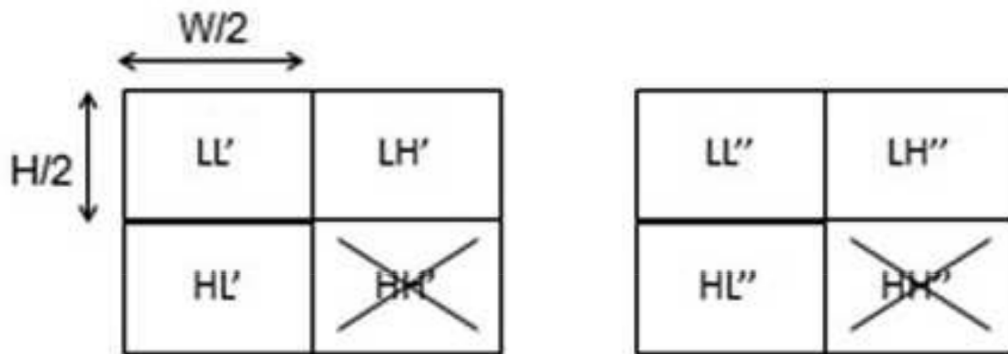


FIG.12

Desensamblaje de subbanda de dos mapas de profundidad



Submuestreo de HL'

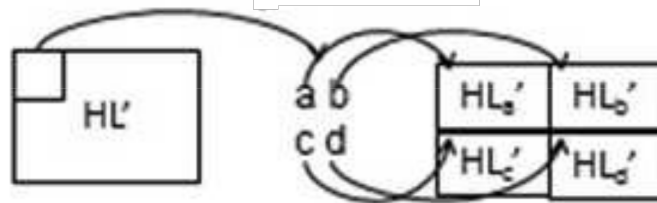


Imagen YUV 4:2:0

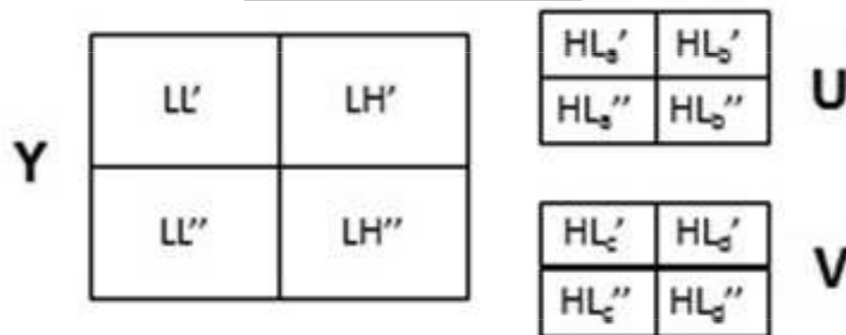


FIG. 13

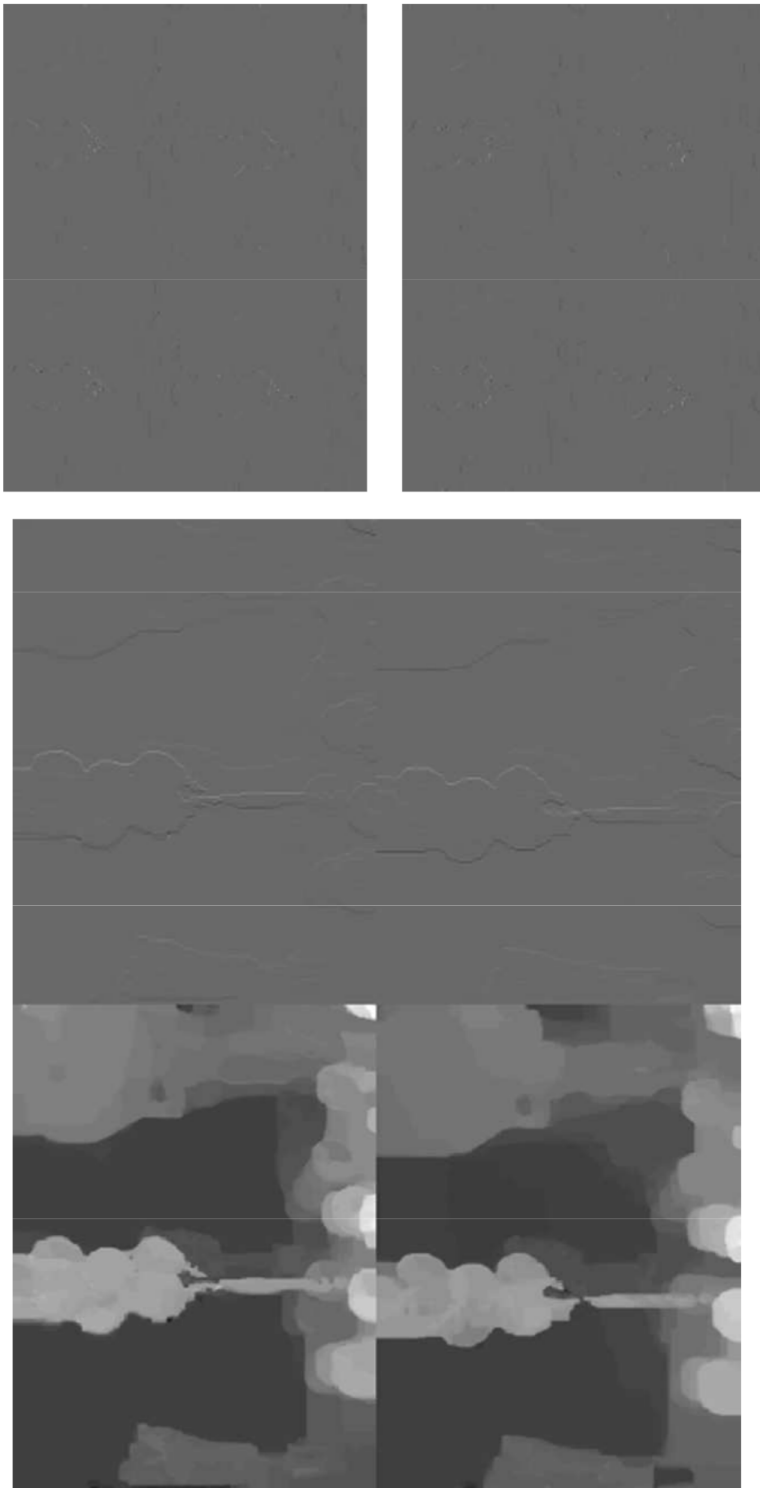


FIG. 14

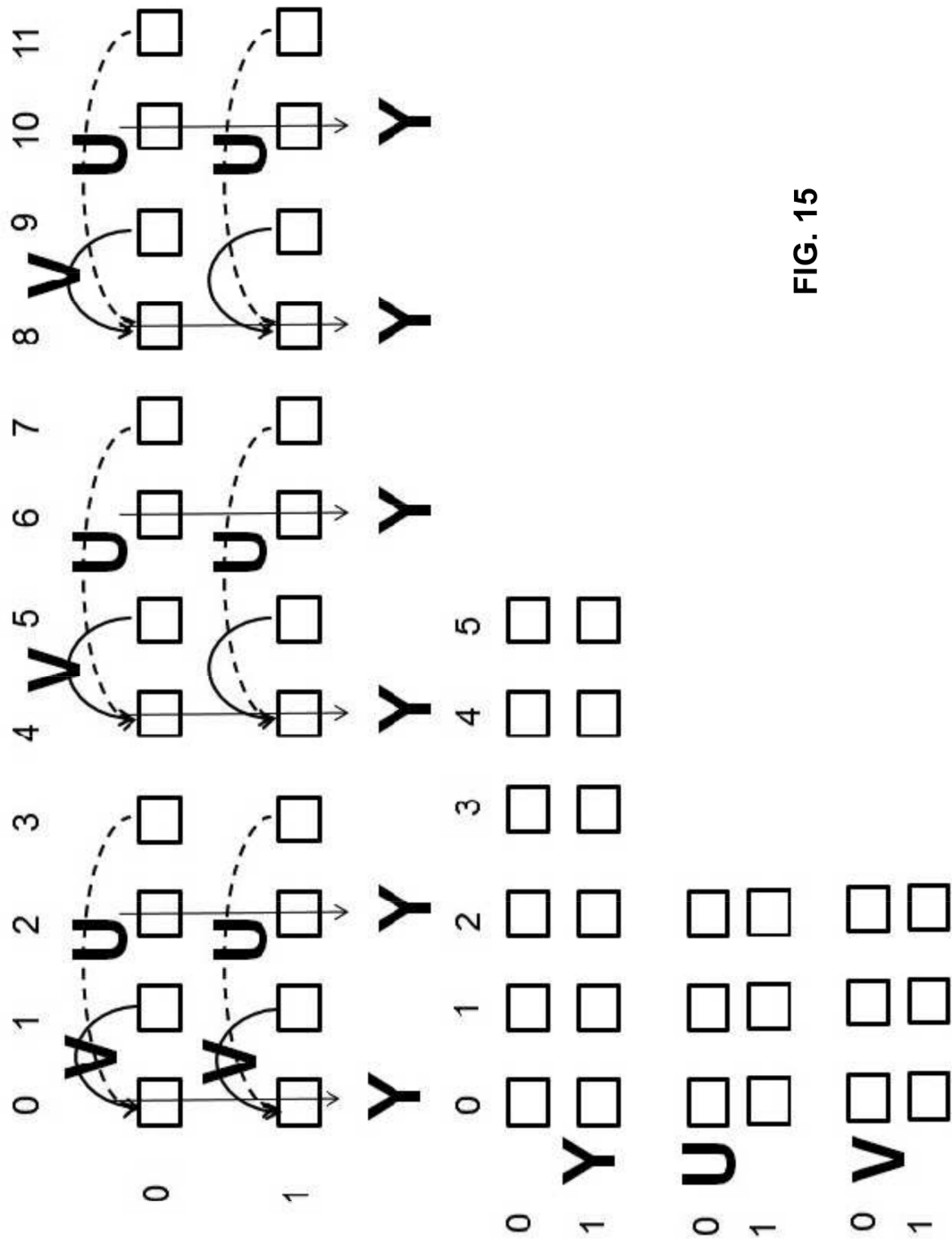


FIG. 15

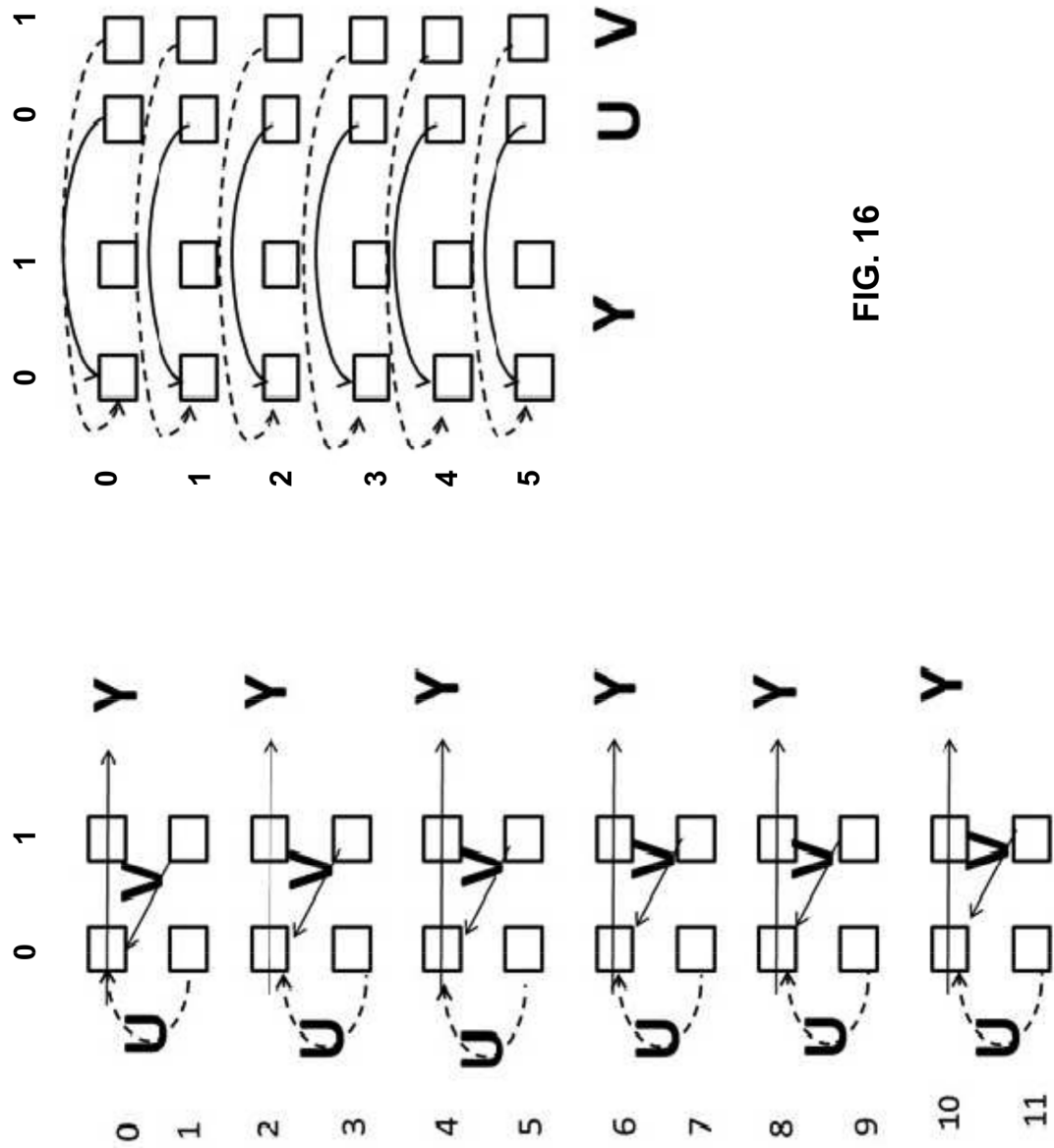


FIG. 16

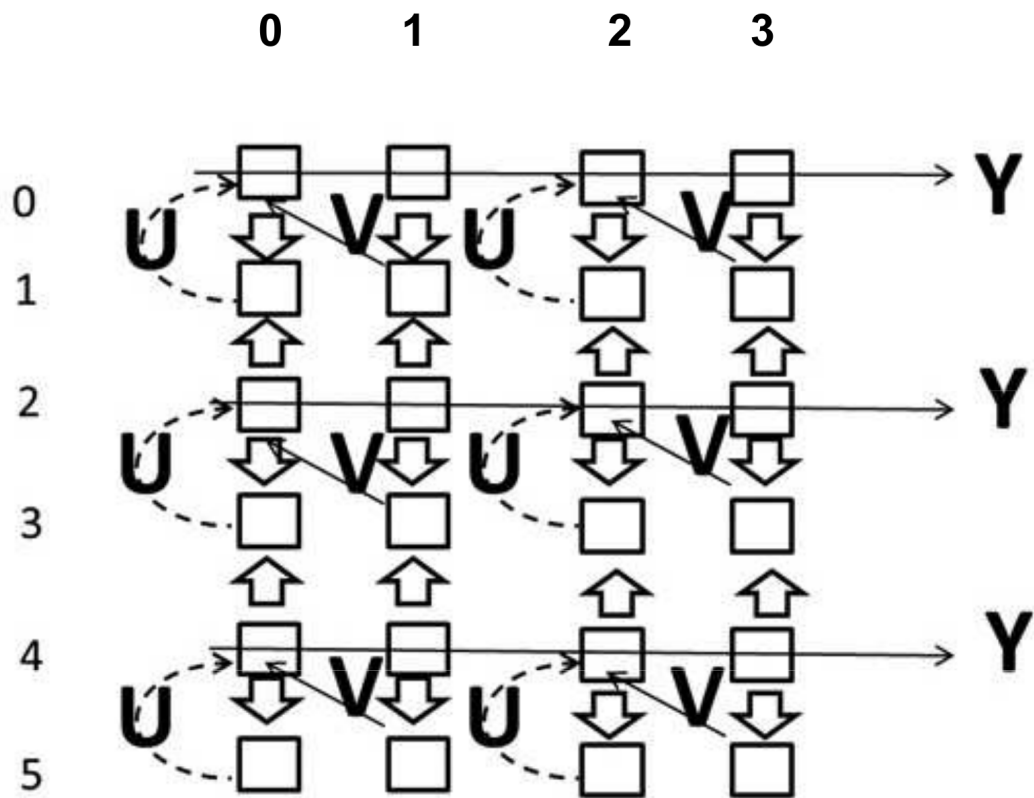
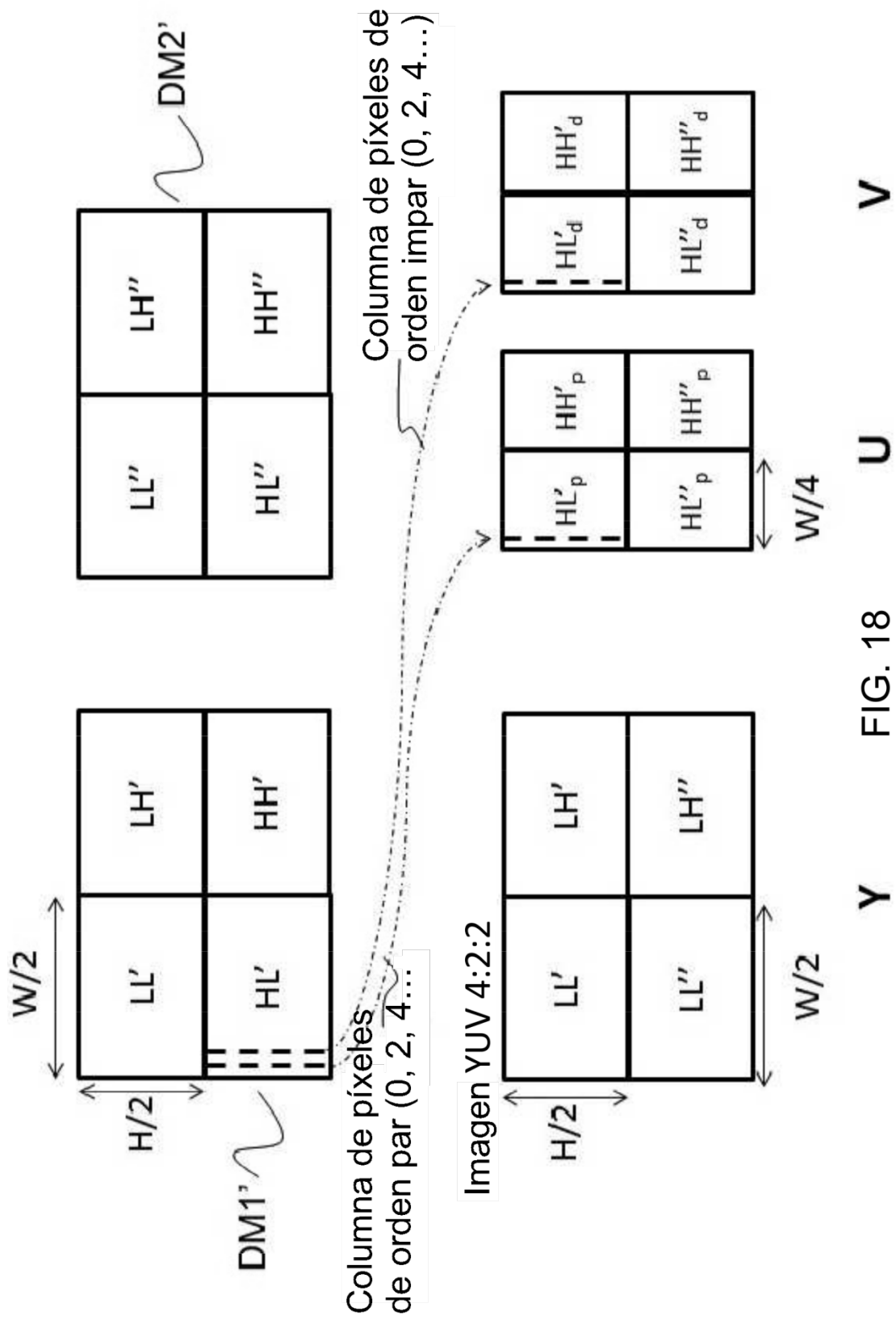


FIG. 17



Y U V

FIG. 18

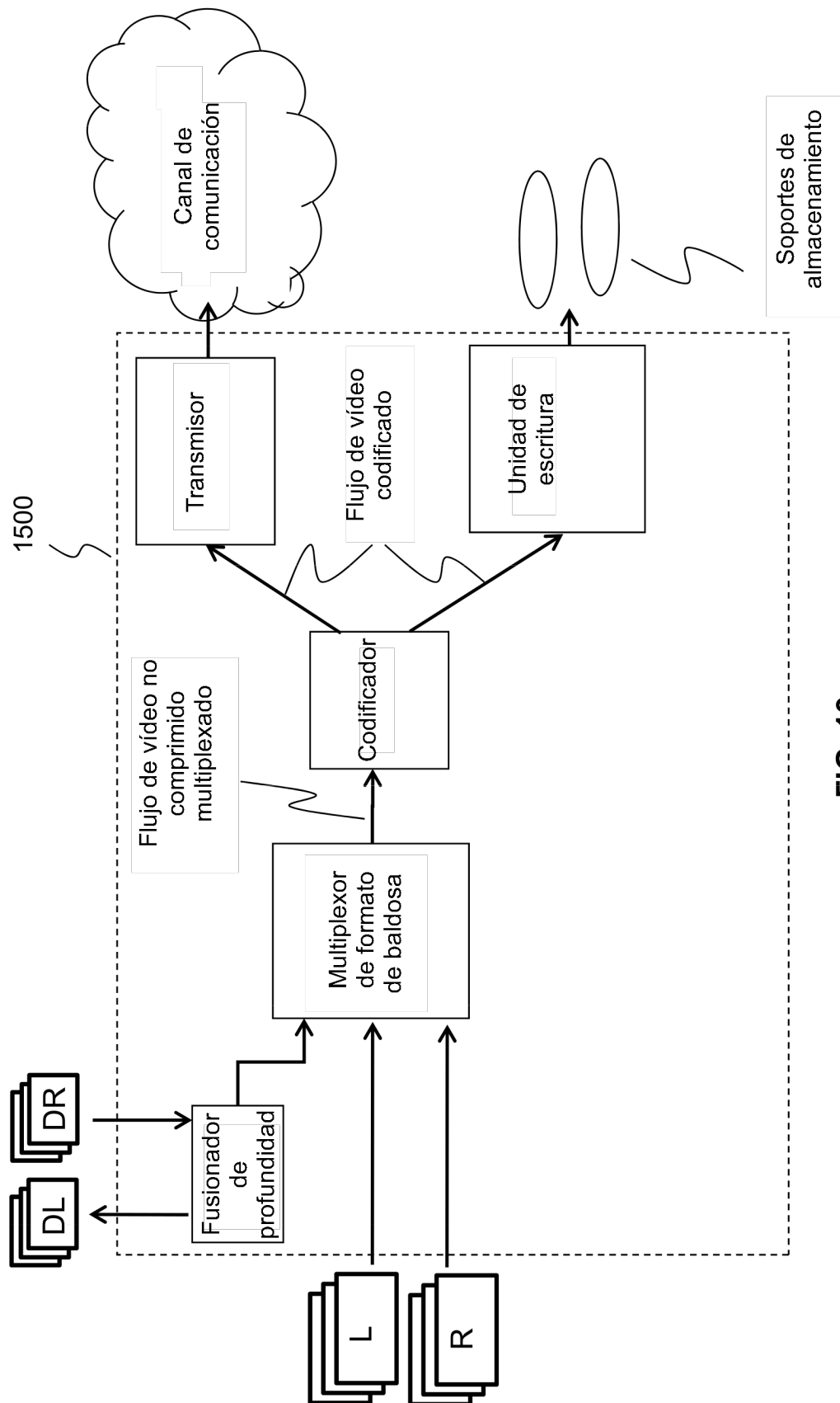


FIG. 19a

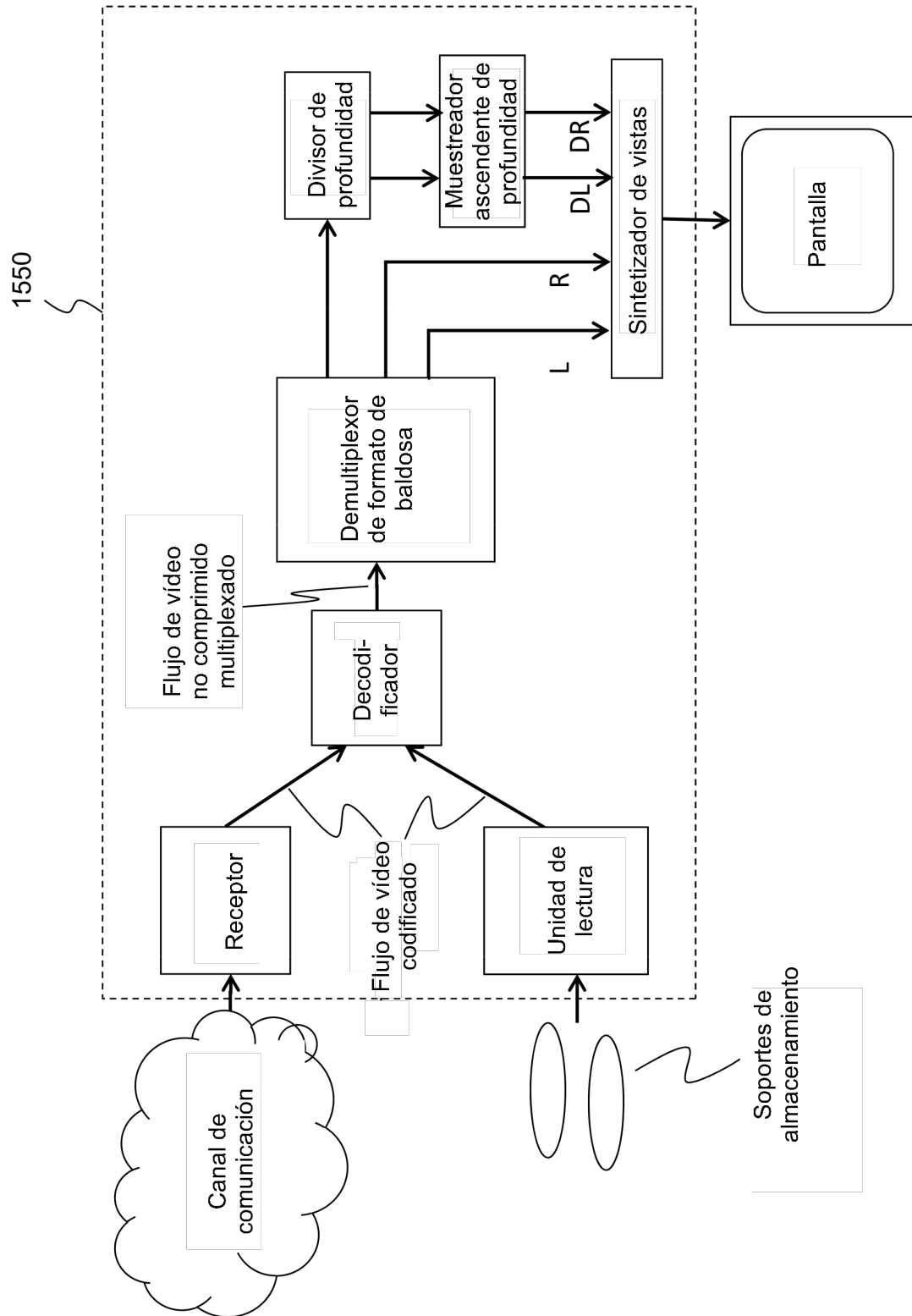


FIG. 19b