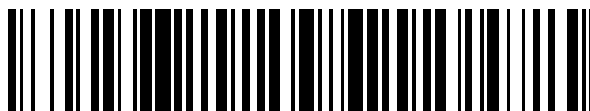


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 452 866**

51 Int. Cl.:

H04N 7/36

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.06.2010** **E 10730948 (6)**

97 Fecha y número de publicación de la concesión europea: **29.01.2014** **EP 2449786**

54 Título: **Puesta en correspondencia de plantillas para codificación de vídeo**

30 Prioridad:

02.07.2009 US 222836 P
08.04.2010 US 756661

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
03.04.2014

73 Titular/es:

QUALCOMM INCORPORATED (100.0%)
International IP Administration 5775 Morehouse
Drive
San Diego, CA 92121-1714, US

72 Inventor/es:

KARCZEWICZ, MARTA y
CHIEN, WEI-JUNG

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 452 866 T3

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

DESCRIPCIÓN

Puesta en correspondencia de plantillas para codificación de vídeo

Campo técnico

La presente divulgación se refiere a la codificación de vídeo y, más en concreto, a las técnicas de codificación de vídeo que utilizan la predicción de movimientos de puesta en correspondencia de plantillas.

Antecedentes

Las capacidades multimedia digitales pueden ser incorporadas en una amplia gama de dispositivos incluyendo televisiones digitales, sistemas digitales de difusión directa, dispositivos de comunicaciones inalámbricas, sistemas de difusión inalámbrica, asistentes personales digitales (PDAs), computadoras portátiles o de escritorio, cámaras digitales, dispositivos de registro digitales, dispositivos de videojuegos, consolas de videojuegos, radiotelefonos celulares o por satélite, reproductores de medios digitales, y similares. Los dispositivos multimedia digitales pueden implementar técnicas de comunicación de vídeo, como por ejemplo MPEG- 2, ITU - H.263, MPEG 4 o ITU-H. 264 / MPEG-4 Parte 10, Codificación de Vídeo Avanzada (AVC), para transmitir y recibir o almacenar y recuperar de manera más eficiente datos de vídeo digitales. Las técnicas de codificación de vídeo pueden llevar a cabo una compresión de vídeo por medio de una predicción espacial y temporal para reducir o eliminar la redundancia inherente en las secuencias de vídeo.

En la codificación de vídeo, la compresión a menudo incluye la predicción espacial, la estimación de movimientos y la compensación de movimientos. La intracodificación se basa en la predicción espacial y en la codificación por transformación, como por ejemplo una transformación cosenoidal discreta (DCT), para reducir o eliminar la redundancia espacial entre los bloques de vídeo dentro de una trama de vídeo determinada. La intercodificación se basa en la predicción temporal y en la codificación por transformación para reducir o eliminar la redundancia temporal entre bloques de vídeo de tramas de vídeo sucesivas de una secuencia de vídeo. Las tramas intracodificadas ("tramas - I") a menudo se utilizan como puntos de acceso aleatorios así como referencias para la intercodificación de otras tramas. Las tramas - I, sin embargo, típicamente muestran una compresión menor que las demás tramas. El término unidades - I puede referirse a las tramas - I, a las rebanadas - I o a otras porciones independientemente descodificables de una trama - I.

Para la intercodificación, un codificador de vídeo lleva a cabo una estimación de movimientos para rastrear el movimiento de los bloques de vídeo en correspondencia entre dos o más tramas adyacentes u otras unidades codificadas, como por ejemplo rebanadas de tramas. Las tramas intercodificadas pueden incluir tramas predictivas ("tramas - P"), las cuales pueden incluir bloques previstos a partir de una trama anterior, y tramas predictivas bidireccionales ("tramas - B") las cuales pueden incluir bloques previstos procedentes de una trama anterior y de una trama posterior de una secuencia de vídeo. Las técnicas de codificación de vídeo convencionales de movimientos compensados comparan un bloque de vídeo con otros bloques de vídeo de una trama de vídeo anterior o posterior con el fin de identificar los datos de vídeo predictivos que pueden ser utilizados para codificar el bloque de vídeo actual. Un bloque de vídeo puede desglosarse en particiones sub-bloque para facilitar una codificación de mayor calidad.

Un bloque de vídeo codificado puede ser representado por una información de predicción que pueda ser utilizada para crear o identificar un bloque predictivo, y un bloque residual de datos indicativo de las diferencias entre el bloque que está siendo codificado y el bloque predictivo. La información de predicción puede comprender los uno o más vectores de movimiento que son utilizados para identificar el bloque de datos predictivo. Dados los vectores de movimiento, el descodificador es capaz de reconstruir los bloques predictivos que fueron utilizados para codificar el residual. Así, dado un conjunto de bloques residuales y un conjunto de vectores de movimiento (y posiblemente alguna sintaxis adicional), el descodificador puede ser capaz de reconstruir una trama de vídeo que fue originalmente codificada. Una secuencia de vídeo codificada puede comprender bloques de datos residuales, vectores de movimiento y posiblemente otros tipos de sintaxis.

La puesta en correspondencia de plantillas es una técnica que puede ser utilizada para eliminar los vectores de movimiento, pero que al mismo tiempo pueda seguir proporcionando las ventajas de la codificación de vídeo de movimientos compensados. En la puesta de correspondencia de plantillas, los píxeles vecinos con respecto al bloque de vídeo que está siendo codificado pueden definir una plantilla, y esta plantilla (más que el bloque de vídeo que está siendo codificado) puede ser comparada con los datos de una trama de vídeo anterior o posterior. Tanto el codificador de vídeo como el descodificador de vídeo pueden llevar a cabo el proceso de puesta en correspondencia de plantillas para identificar el movimiento sin el uso de vectores de movimiento. Así, con la puesta en correspondencia de plantillas, los vectores de movimiento no son codificados en un flujo de bits. Por el contrario, los vectores de movimiento son derivados esencialmente del proceso de puesta en correspondencia de plantillas a medida que la trama es codificada y descodificada.

El documento "Procedimiento multihipótesis que utiliza una derivación de vectores del lado del movimiento del descodificador en una codificación de vídeo intertramas" ["Multhypothesis prediction using decoder side-motion vector derivation in inter-frame vídeo coding"], Kamp S. et al, Visual Communications and Imagen Processing, 20-01-

2009 a 22-01-2009, San José, 20 de enero de 2009, XP030081712 divulga un procedimiento de codificador de vídeo en el que un conjunto de hipótesis son generadas para una codificación predictiva de puesta en correspondencia de plantillas, siendo el bloque de vídeo actual codificado utilizando datos de vídeo procedentes del conjunto de hipótesis.

5 **Sumario**

En general, la presente divulgación describe técnicas de codificación de vídeo aplicables a la codificación de predicción de movimiento de puesta en correspondencia de plantillas de bloques de vídeo dentro de unidades de vídeo. Una unidad de vídeo puede comprender una trama de vídeo o una rebanada de una trama de vídeo. En la predicción de movimiento de la puesta en correspondencia de plantillas, un bloque de vídeo es codificado de manera predictiva y descodificado en base a una o más listas de datos de referencia predictivos sin basarse en vectores de movimiento, proporcionando sin embargo al tiempo las ventajas de la codificación de vídeo de compensación de movimiento compensado. En particular, los píxeles vecinos con respecto al bloque de vídeo que está siendo codificado pueden definir una plantilla y esta plantilla, y no el bloque de vídeo que está siendo codificado, puede ser comparada con los datos almacenados en las listas de los datos de referencia predictivos. Los datos de referencia predictivos son generados en base a una o más tramas que pueden producirse antes o después de una trama destinataria.

De acuerdo con un primer aspecto de la primera invención, se proporciona un procedimiento de codificación de vídeo para la codificación de un bloque de vídeo actual de una unidad de vídeo actual que comprende la generación de un primer conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del primer conjunto se basan en tramas individuales de datos procedentes de al menos una lista que contiene tramas anteriores en orden de representación y una trama procedente de una lista que contiene tramas futuras en orden de representación;

la generación de un segundo conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del segundo conjunto están formadas a partir de una predicción de movimientos bidireccionales entre pares de tramas, en el que cada par de tramas comprende una trama procedente de la lista que contiene las tramas anteriores en orden de representación y una trama procedente de la lista que contiene tramas futuras en orden de representación; en el que cada una de las hipótesis de los primero y segundo conjuntos de hipótesis comprende unos datos de vídeo correspondientes a una forma de plantilla definida con respecto a un emplazamiento del bloque;

la aplicación de técnicas para la aplicación de hipótesis entre los primero y segundo conjuntos de hipótesis; y la codificación, por medio de un codificador de vídeo, del bloque de vídeo actual utilizando los datos de vídeo predictivos procedentes de las hipótesis seleccionadas.

De acuerdo con un segundo aspecto de la presente invención, se proporciona un medio de almacenamiento legible por computadora codificado con instrucciones para determinar que uno o más procesadores programables lleven a cabo el procedimiento de acuerdo con el primer aspecto de la invención.

De acuerdo con un tercer aspecto de la presente invención, se proporciona un aparato de codificación de vídeo para la codificación de un bloque de vídeo actual, comprendiendo el aparato:

una unidad de predicción que genera un primer conjunto de hipótesis y un segundo conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del primer conjunto se basan en las tramas individuales de datos procedentes de al menos una lista que contiene tramas anteriores en orden de representación y una trama procedente de una lista que contiene tramas futuras en orden de representación, y las hipótesis del segundo conjunto están formadas a partir de la predicción de movimientos bidireccionales entre pares de tramas, en el que cada par de tramas comprende una trama procedente de la lista que contiene tramas anteriores en orden de representación y una trama procedente de la lista que contiene tramas futuras en orden de representación, en el que cada una de las hipótesis de los primero y segundo conjuntos de hipótesis comprende datos de vídeo correspondientes a una forma de plantilla definida con respecto a un emplazamiento de bloque de vídeo y aplica técnicas para la selección de hipótesis entre los primero y segundo conjuntos de hipótesis; y

una unidad de codificación de vídeo que codifica los datos de vídeo actual utilizando los datos de vídeo predictivos procedentes de las hipótesis seleccionadas.

Al aplicar la predicción de puesta en correspondencia de plantillas a las tramas - B, se generan dos listas de datos de vídeo predictivos, la primera lista procedente de una trama que se produce antes de la trama destinataria (por ejemplo, la trama anterior), y la segunda lista procedente de una trama posterior a la trama destinataria (por ejemplo, la trama futura). En la aplicación de la puesta en correspondencia de plantillas a las tramas - B, se pueden utilizar uno o dos conjuntos de hipótesis para la trama - B. Si se utiliza un conjunto de hipótesis, la hipótesis establecida puede contener hipótesis encontradas tanto en la primera lista como en la segunda lista. Si se utilizan dos conjuntos de hipótesis, un conjunto de hipótesis puede contener hipótesis procedentes de la primera lista y el otro conjunto de hipótesis puede contener las hipótesis encontradas procedentes de la segunda lista. Un conjunto de hipótesis

adicional se puede formar a partir de la predicción de movimientos bidireccionales, en el que los datos procedentes de la primera lista y los datos procedentes de la segunda lista son combinados y utilizados para la predicción.

Un elemento de sintaxis puede ser utilizado para señalar, al descodificador, el número de hipótesis utilizadas por un codificador al codificar o descodificar datos de vídeo que utilizan la predicción de puesta en correspondencia de plantillas. Por ejemplo, el codificador puede estar configurado para seleccionar uno entre dos números diferentes que especifiquen el número de hipótesis utilizadas durante el proceso de codificación, y el descodificador puede ser configurado para aplicar el número de hipótesis identificados por el codificador. El codificador puede determinar un coste asociado con la utilización de uno u otro número de hipótesis durante la codificación y puede seleccionar el número de hipótesis con el coste más bajo. El codificador puede entonces codificar los datos de vídeo en base al número de hipótesis con el coste más bajo y puede señalar este número al descodificador, utilizando el elemento de sintáctico.

Al determinar qué hipótesis no considerar, un codificador o descodificador compara la hipótesis objeto de consideración (esto es, la hipótesis candidata) con una hipótesis de referencia. Si la diferencia entre la hipótesis candidato y la hipótesis de referencia es mayor que un umbral predefinido, la hipótesis candidato puede no ser tomada en consideración. La hipótesis de referencia puede ser generada promediando, en un ejemplo, todas las hipótesis objeto de consideración o, en otro ejemplo, un subconjunto de las mejoras hipótesis objeto de consideración. En el aspecto de rechazo de las hipótesis de la presente divulgación, no se requiere la señalización al codificador o al descodificador en cuanto a cuáles sean las hipótesis aceptadas o rechazadas para la predicción de movimientos de puesta en correspondencia de plantillas aplicadas a las tramas - P o a las tramas - B.

Se puede prever una técnica para la ponderación de la suma de la diferencia absoluta en la predicción de movimientos de la puesta en correspondencia de las plantillas sobre las tramas - P y las tramas - B.

Una forma de plantilla puede ser definida con respecto al bloque actual y una forma de plantilla correspondiente puede ser definida con respecto a un bloque destinatario. La suma de la diferencia absoluta es el valor absoluto de la diferencia entre los valores de píxeles de cada píxel de la forma de plantilla definida con respecto al bloque actual y a los valores de píxeles de cada píxel correspondiente de cada plantilla definida con respecto al bloque de referencia. Las formas de las plantillas definidas con respecto al bloque actual o al bloque destinatario pueden, cada una, ser divididas en dos o más particiones. A cada partición se le otorga una ponderación que disminuye a medida que aumenta la distancia entre cada partición y el bloque correspondiente. Mediante la ponderación de la suma de la diferencia absoluta, la relación espacial de la plantilla y de los bloques de vídeo puede ser tomada en consideración al llevar a cabo la predicción de la puesta en correspondencia de plantillas, lo que puede traducirse en una puesta en correspondencia de plantillas más precisa.

Un desplazamiento de plantillas calculado puede ser aplicado a la predicción de movimientos de puesta en correspondencia de plantillas según se aplica a las tramas - P y a las tramas - B. El desplazamiento de las plantillas puede ser utilizado para predecir el desplazamiento entre el bloque de referencia y el bloque destinatario. Al buscar la plantilla de puesta en correspondencia, cada píxel del bloque de referencia puede incorporar un conjunto de correspondientes coordenadas que definan donde está situado cada píxel en el bloque de referencia. Las coordenadas indican una posición de búsqueda. Para cada posición de búsqueda, se puede calcular la diferencia media entre el valor de píxel de cada píxel de la plantilla de la trama actual y el valor de píxel de cada píxel de la trama de referencia. Esta diferencia media en valores de píxel para la entera plantilla se puede definir como el desplazamiento de plantillas. Para cada plantilla, un desplazamiento puede ser calculado mediante este técnica, una vez que se ha calculado el desplazamiento de la plantilla, las técnicas de esta divulgación pueden calcular la suma de las diferencias absolutas para el bloque destinatario tomando cada valor del píxel del bloque destinatario, restando el valor de píxel del correspondiente píxel del bloque de referencia y restando después el desplazamiento. Mediante el ajuste de la suma de la diferencia absoluta del bloque destinatario mediante el desplazamiento, puede ser más probable que la hipótesis seleccionada sea la mejor hipótesis, reduciendo con ello los residuales y mejorando la compresión del vídeo.

Un residual de luma puede ser incorporado en la predicción de puesta en correspondencia de plantillas aplicado a las tramas - P y a las tramas - B. En este caso, el bloque destinatario puede ser un bloque de vídeo luma y el bloque de vídeo luma se puede dividir en sub-bloques. Cuando un primer sub-bloque de un bloque luma no incorpora unos píxeles reconstruidos disponibles, la codificación de un segundo sub-bloque que se base en el primer sub-bloque puede o bien retrasarse hasta que la codificación del primer sub-bloque se haya completado, o bien la codificación del segundo sub-bloque se puede basar en el contenido del primer sub-bloque como ha sido generado mediante la combinación de la predicción de compensación del movimiento con el residual luma codificado. Mediante la combinación de la predicción de movimientos compensados y del residual luma codificado, el contenido generado del primer sub-bloque puede ser más próximo a los datos originales y la predicción de movimientos de la puesta en correspondencia de plantillas del segundo sub-bloque puede ser más precisa.

Las técnicas descritas en la presente divulgación pueden ser implementadas en hardware, software, firmware o cualquier combinación de estos. Si se implementan en software, el software puede ser ejecutado en uno o más procesadores, como por ejemplo un microprocesador, un circuito integrado específico de la aplicación (ASIC), una matriz programable de puertas sobre el terreno (FPGA), o un procesador digital de la señal (DSP). El software que

ejecuta las técnicas puede estar inicialmente almacenado en un medio legible por computadora y cargado y ejecutado en el procesador.

Los detalles de una o más formas de realización de la invención se definen en los dibujos que se acompañan y en la descripción posterior. Otras características, objetos y ventajas de la invención se pondrán de manifiesto a partir de la descripción los dibujos y las reivindicaciones.

Breve descripción de los dibujos

La FIG. 1 es un diagrama de bloques ejemplar que ilustra un sistema de codificación y descodificación de vídeo.

La FIG. 2 es un diagrama de bloques que ilustra un ejemplo de un codificador de vídeo de acuerdo con la presente divulgación.

La FIG. 3 es un diagrama de bloques conceptual que ilustra una unidad de vídeo destinataria ejemplar y una unidad de vídeo de referencia de acuerdo con la presente divulgación.

La FIG. 4 es un diagrama de bloques conceptual que ilustra un ejemplo de una predicción de movimiento de puesta de correspondencia de plantillas en el que el contenido de un bloque de vídeo se determina de acuerdo con la presente divulgación.

La FIG. 5 es un diagrama de bloques que ilustra un ejemplo de un descodificador de vídeo de acuerdo con la presente divulgación.

Las FIGS. 6A - 6B son diagramas de flujo que ilustran un proceso ejemplar llevado a cabo por un codificador de vídeo de acuerdo con la presente divulgación.

Las FIGS. 7A - 7B son diagramas de flujo que ilustran un proceso ejemplar llevado a cabo por un descodificador de vídeo de acuerdo con la presente divulgación.

Descripción detallada

La presente divulgación describe unas técnicas de codificación de vídeo aplicables a la codificación de predicción de movimientos de puesta en correspondencia de plantillas de unidades de vídeo. En la presente divulgación, el término "codificación" se refiere a la codificación o a la descodificación. De modo similar, el término "codificador" se refiere, en términos generales, a cualquier codificador de vídeo, descodificador de vídeo o codificador / descodificador combinados (códec). De acuerdo con ello, el término "codificador" se utiliza en la presente memoria para referirse a un dispositivo o aparato informático especializado que llevar a cabo una codificación de vídeo o una descodificación de vídeo o tanto una codificación como una descodificación de vídeo.

La puesta en correspondencia de plantillas es una técnica que puede ser utilizada para eliminar vectores de movimiento pero que al mismo tiempo proporciona las ventajas de la codificación de vídeo de movimiento compensado. En la puesta en práctica de la correspondencia de plantillas, una plantilla define una forma de píxeles vecinos con respecto a un bloque de vídeo que está siendo codificado. La forma de plantilla se define con respecto al bloque de vídeo y puede lindar con o rodear el bloque de vídeo pero no se requiere que esté situado inmediatamente en posición adyacente al bloque de vídeo. La forma de plantilla puede ser aplicada a uno o más bloques de vídeo dentro de una unidad de vídeo de referencia. Una hipótesis se puede basar en los datos de vídeo predictivos contenidos dentro de una forma de plantilla, en base a un primer conjunto de datos de vídeo predictivos procedente de al menos una porción de una primera unidad de vídeo predictiva, o en base a un primer conjunto de datos de vídeo predictivo procedentes de al menos una porción de una primera unidad de vídeo predictiva y un segundo conjunto de datos de vídeo predictivo procedente de al menos una porción de una segunda unidad de vídeo predictiva. Las hipótesis que sean generadas y tomadas en consideración durante la predicción de movimientos de puesta en correspondencia de plantillas pueden ser designadas como hipótesis candidata. Pueden existir múltiples hipótesis candidatas para cada bloque de vídeo que está siendo codificado.

Las técnicas de la presente divulgación incluyen la aplicación de la puesta en correspondencia de plantillas a las tramas - B así como a las diversas técnicas que pueden ser aplicadas para mejorar la puesta en correspondencia de plantillas cuando se aplican a las tramas - B o a las tramas - P. En un aspecto de la presente divulgación, la selección y rechazo de hipótesis mejoradas se aplica a la predicción de movimiento de puesta en correspondencia de plantillas. En este contexto, un elemento sintáctico es introducido para señalar, al descodificador, el número de hipótesis utilizadas a un codificador al codificar o descodificar los datos de vídeo que utilizan la predicción de puesta en correspondencia de plantillas. Al determinar cuáles son las hipótesis que hay que descartar, un codificador o descodificador compara la hipótesis candidata con una hipótesis de referencia. Si la diferencia entre la hipótesis candidata y la hipótesis de referencia es mayor que un umbral predefinido, la hipótesis candidata puede no ser tomada en cuenta.

En aspectos adicionales de la presente divulgación, se prevén técnicas para la ponderación de la suma de la diferencia absoluta, aplicando un desplazamiento de plantillas calculado, e incorporando un residual luma en la predicción de movimiento de puesta en correspondencia de plantillas. La relación espacial de la plantilla y de los bloques de vídeo puede ser tomada en consideración al llevar a cabo la predicción de puesta en correspondencia en plantillas mediante la ponderación de la suma de la diferencia absoluta, que puede traducirse en una puesta en correspondencia de plantillas más precisa. El desplazamiento de plantillas puede ser utilizado para predecir el desplazamiento entre el bloque de referencia y el bloque destinatario. Una vez que el desplazamiento de plantillas se ha calculado, las técnicas de la presente divulgación pueden calcular la suma de las diferencias absolutas para el bloque destinatario tomando cada valor de píxel del bloque destinatario, restando el valor de píxel del píxel correspondiente del bloque de referencia y restando también el desplazamiento. Mediante el ajuste de la suma de la diferencia absoluta del bloque destinatario mediante el desplazamiento, puede ser más probable que la hipótesis seleccionada sea la mejor hipótesis, reduciendo con ello los residuales y mejorando la compresión del vídeo. El contenido generado de un primer sub-bloque puede ser más próximo a los datos originales si el contenido es generado mediante la combinación de la predicción de compensación de movimiento y del residual luma codificado y la predicción de movimiento de puesta en correspondencia de plantillas de un segundo sub-bloque que se base en el primer sub-bloque puede ser más precisa. De esta manera, las técnicas de la presente divulgación pueden mejorar la predicción de movimiento de puesta en correspondencia de plantillas mediante la generación de un bloque de vídeo codificado que contenga menos valores residuales.

La FIG. 1 es un diagrama de bloques que ilustra un sistema 10 de codificación y descodificación de vídeo ejemplar que puede implementar técnicas de la presente divulgación. Como se muestra en la FIG. 1, el sistema 10 incluye un dispositivo 12 de fuente que transmite un vídeo codificado a un dispositivo 16 de destino por medio de un canal 15 de codificación. El dispositivo 12 de fuente y el dispositivo 16 de destino pueden comprender cualquier dispositivo entre una amplia variedad de dispositivos. En algunos casos, el dispositivo 12 de fuente y el dispositivo 16 de destino comprenden dispositivos de comunicaciones inalámbricos, como por ejemplo teléfonos inalámbricos, los llamados radiotelefonos celulares o por satélite, o cualquier dispositivo que pueda comunicar información de vídeo a través de un canal 15 de comunicación, que puede o puede no ser inalámbrico. Las técnicas de la presente divulgación, sin embargo, las cuales se refieren a las técnicas de predicción de la puesta en correspondencia de plantillas, no están necesariamente limitadas a aplicaciones o configuraciones inalámbricas.

En el ejemplo de la FIG. 1, el dispositivo 12 de fuente puede incluir una fuente 20 de vídeo, un codificador 22 de vídeo, un modulador / desmodulador (módem) 23 y un transmisor 24. El dispositivo 16 de destino puede incluir un receptor 26, un módem 27, un descodificador 28 de vídeo y un dispositivo 30 de representación. De acuerdo con la presente divulgación, el codificador 22 de vídeo del dispositivo 12 de fuente puede estar configurado para aplicar la predicción de puesta en correspondencia de plantillas a tramas - B así como a tramas - P. El codificador 22 de vídeo puede también estar configurado para aplicar aspectos concretos de la predicción en movimiento de la puesta en correspondencia de plantillas de la presente divulgación. Por ejemplo, el codificador 22 de vídeo puede estar configurado para identificar el número de hipótesis utilizadas al codificar datos de vídeo. El número de hipótesis puede ser señalado al codificador 28 de vídeo, utilizando un elemento de sintaxis como parte de un flujo de bits codificado. El codificador 28 de vídeo puede estar configurado para interpretar y aplicar el elemento de sintaxis que identifica el número de hipótesis que hay que utilizar al descodificar los datos de vídeo.

El sistema 10 ilustrado de la FIG. 1 es meramente ejemplar. Las técnicas de predicción de puesta en correspondencia de plantillas de la presente divulgación pueden ser llevadas a cabo mediante cualquier dispositivo de codificación que soporte la codificación de vídeo de movimiento compensado utilizando la puesta en correspondencia de plantillas. El dispositivo 12 de fuente y el dispositivo 16 de destino son simplemente ejemplos de dichos dispositivos de codificación en los cuales el dispositivo 12 de fuente genera datos de vídeo codificados desde la transmisión hasta el dispositivo 16 de destino. En algunos casos, los dispositivos 12, 16 pueden operar de una manera sustancialmente simétrica de forma que, cada uno de los dispositivos 12, 16 incluya componentes de codificación y descodificación de vídeo. Por tanto, el sistema 10 puede soportar una transmisión de vídeo unidireccional o bidireccional entre los dispositivos 12, 16 de vídeo, por ejemplo, para la transmisión de flujos de vídeo, reproducción de vídeo, difusión de vídeo o telefonía de vídeo.

La fuente 20 de vídeo del dispositivo 12 de fuente puede incluir un dispositivo de captación de vídeo, como por ejemplo una videocámara, un archivo de vídeo que contenga el vídeo anteriormente captado o una alimentación de vídeo procedente de un proveedor de contenidos de vídeo. Como una alternativa adicional, la fuente 20 de vídeo puede generar datos a base de gráficos informáticos como el vídeo de fuente, o una combinación de vídeo en vivo, vídeo archivado y vídeo generado por computadora. En algunos casos, si la fuente 20 de vídeo es una videocámara, el dispositivo 12 de fuente y el dispositivo 16 de destino pueden formar unos llamados teléfonos con cámara o videotelefonos. En cada caso, el vídeo captado, precaptado o generado por computadora puede ser codificado por el codificador 22 de vídeo. La información de vídeo codificada puede, a continuación, ser modulada por el módem 23 de acuerdo con un estándar de comunicación, por ejemplo, el acceso múltiple por división de código (CDMA) u otro estándar de comunicación, y transmitida al dispositivo 16 de destino por medio del transmisor 24. El módem 23 puede incluir varios mezcladores, filtros, amplificadores u otros componentes diseñados para la modulación de la señal. El transmisor 24 puede incluir circuitos diseñados para transmitir datos, incluyendo amplificadores, filtros y una o más antenas.

El receptor 26 del dispositivo 16 de destino recibe información a través del canal 15, y el módem 27 desmodula la información. De nuevo aquí, el proceso de codificación de vídeo puede implementar una o más de las técnicas descritas en la presente memoria. La información comunicada a través del canal 15 puede incluir la información definida por el codificador 22 de vídeo, la cual puede ser utilizada por el descodificador 28 de vídeo de acuerdo con la presente divulgación. El dispositivo 30 de pantalla presenta los datos de vídeo descodificados a un usuario, y puede comprender cualquier gama de dispositivos de pantalla, como por ejemplo un tubo de rayos catódicos, una pantalla de cristal líquido (LCD), una pantalla de plasma, una pantalla de diodos fotoemisores orgánicos (OLED), o cualquier otro tipo de dispositivo de representación.

En el ejemplo de la FIG. 1, el canal 15 de comunicación puede comprender cualquier medio de comunicación inalámbrico o cableado, como por ejemplo un espectro de radiofrecuencia (RF) o una o más líneas de transmisión físicas, o cualquier combinación de medios inalámbricos o cableados. De acuerdo con ello, el módem 23 y el transmisor 24 pueden soportar muchos posibles protocolos inalámbricos, protocolos cableados o protocolos cableados e inalámbricos. El canal 15 de comunicación puede formar parte de una red basada en paquetes, como por ejemplo una red de área local (LAN), una red de área extendida (WAN), o una red global, como por ejemplo Internet, comprendiendo una interconexión de una o más redes. El canal 15 de comunicación representa, en términos generales, cualquier medio de comunicación apropiado, o colección de diferentes medios de comunicación, para transmitir datos de vídeo desde el dispositivo 12 de fuente hasta el dispositivo 16 de destino. El canal 15 de comunicación puede incluir encaminadores, conmutadores, estaciones de base o cualquier otro equipamiento que pueda ser de utilidad para facilitar la comunicación entre el dispositivo 12 de fuente hasta el dispositivo 16 de destino.

El codificador 22 de vídeo y el descodificador 28 de vídeo pueden operar de acuerdo con un estándar de compresión de vídeo, como por ejemplo el estándar ITU-T H.264, descrito como alternativa como Codificación de Vídeo Avanzada (AVC), el MPEG-4, Parte 10. Las técnicas de la presente divulgación, sin embargo, no están limitadas a ningún estándar de codificación concreto. Aunque no se muestra en la FIG. 1, en algunos aspectos, el codificador 22 de vídeo y el descodificador 28 de vídeo pueden estar cada uno integrados con un codificador y descodificador de audio, y pueden incluir unidades MUX-DE-MUX apropiadas, u otro hardware o software para gestionar la codificación tanto de audio como de vídeo en un flujo de datos común o en flujos de datos separados. Si son aplicables, las unidades MUX-DE-MUX pueden adaptarse al protocolo de multiplexación ITU H.223 o a otros protocolos como por ejemplo el protocolo de datagramas de usuario (UDP).

El estándar ITU-T H.264 / MPEG-4 (AVC) fue formulado por el Grupo de Expertos de Codificación de Vídeo ITU-T (VCEG) junto con el Grupo de Expertos de Imágenes Móviles (MPEG) ISO / IEC como producto de una participación colectiva conocida como Equipo Mixto de Vídeo (JVT). En algunos aspectos, las técnicas descritas en la presente divulgación pueden ser aplicadas a dispositivos que en general se adaptan al estándar H.264, se describe la Recomendación H.264 ITU-T. La Codificación de Vídeo Avanzada para servicios audiovisuales genéricos, por el Grupo de Estudio ITU-T, y fechada en marzo de 2005, la cual será designada en la presente memoria como el estándar H.264 o la especificación H.264, o el estándar o especificación H.264 / AVC. El Equipo Mixto de Vídeo (JVT) continúa trabajando sobre extensiones hasta H.264 / MPEG-4 AVC.

El codificador 22 de vídeo y el codificador 28 de vídeo pueden cada uno ser implementados como uno o más microprocesadores, procesadores digitales de la señal (DSPs), circuitos integrados específicos de la aplicación (ASICs), matrices de puertas programables sobre el terreno (FPGAs), lógica discreta, software, hardware, firmware o cualquier combinación de estos. Cada codificador 22 de vídeo y codificador 28 de vídeo puede ser incluido en uno o más codificadores o descodificadores, pudiendo cualquiera de ellos estar integrados como parte de un codificador / descodificador combinado (CODEC) en un dispositivo móvil respectivo, dispositivo de abonado, dispositivo de difusión, servidor, o similares.

Una secuencia de vídeo típicamente incluye una serie de tramas de vídeo. El codificador 22 de vídeo y el codificador 28 de vídeo puede operar sobre bloques de vídeo dentro de tramas de vídeo individuales con el fin de codificar y descodificar los datos de vídeo. Los bloques de vídeo pueden presentar tamaños fijos o variables, y pueden diferir en cuanto a su tamaño de acuerdo con un estándar de codificación específico. Cada trama de vídeo puede incluir una serie de rebanadas u otras unidades descodificables de manera independiente. Cada rebanada puede incluir una serie de macrobloques, los cuales pueden estar dispuestos en sub-bloques. A modo de ejemplo, el estándar ITU-T H.264 soporta una intrapredicción en bloques de diversos tamaños, como por ejemplo 16 por 16, 8 por 8, o 4 por 4 para componentes luma, y 8 x 8 para componentes croma, así como una interpredicción en diversos tamaños de bloque como por ejemplo 16 por 16, 16 por 8, 8 por 16, 8 por 8, 8 por 4, 4 por 8 y 4 por 4 para componentes luma y los correspondientes tamaños a escala para componentes croma. Los bloques de vídeo pueden comprender bloques de datos de píxeles o bloques de coeficientes de transformación, por ejemplo, después de un proceso de transformación, como por ejemplo una transformación cosenoidal discreta o un proceso de transformación conceptualmente similar. Algunas técnicas de la presente divulgación pueden ser concretamente aplicadas a bloques luma (u otros tipos de bloques) que sean divididos en sub-bloques.

Los bloques de vídeo de menor tamaño pueden proporcionar una mejor resolución, y pueden ser utilizados para emplazamientos de una trama de vídeo que incluyan altos niveles de detalle. En general, los macrobloques y los diversos sub-bloques pueden ser considerados como bloques de vídeo. Así mismo, una rebanada puede ser

considerada como una serie de bloques de vídeo como por ejemplo macrobloques y / o sub-bloques. Cada rebanada puede ser una unidad descodificable de manera independiente de una trama de vídeo. Como alternativa, las propias tramas pueden ser unidades descodificables u otras porciones de una trama pueden ser definidas como unidades descodificables. El término "unidad codificada" se refiere a cualquier unidad descodificable de manera independiente de una trama de vídeo, como por ejemplo una rebanada entera, una rebanada de una trama, un grupo de imágenes (GOPs), u otra unidad descodificable de manera independiente definida de acuerdo con las técnicas de codificación utilizadas.

Después de la intercodificación predictiva (la cual puede incluir las técnicas de puesta en correspondencia de plantillas de la presente divulgación) y después de cualquier transformación (como por ejemplo la transformación de números enteros 4×4 u 8×8 utilizada en el H.264 / AVC o una transformación cosenoidal discreta o DCT), se puede llevar a cabo una cuantificación. La cuantificación se refiere en general a un proceso en el que los coeficientes son cuantificados para posiblemente reducir la cantidad de datos utilizados para representar los coeficientes. El proceso de cuantificación puede reducir la profundidad de bits asociada con algunos o todos los coeficientes. Por ejemplo, un valor de 16 bits puede ser redondeado a la baja hasta un valor de 15 bits durante la cuantificación. Después de la cuantificación, se puede llevar a cabo una codificación de entropía, por ejemplo, de acuerdo con una codificación de longitud variable de contenido adaptativo (CAVLC), una codificación aritmética binaria de contexto adaptativa (CABAC), u otra metodología de codificación de entropía.

Las técnicas de la presente divulgación son específicamente aplicables a la predicción de la puesta en correspondencia de plantillas. La predicción bidireccional es la predicción de los llamados "bloques de vídeo - B" en base a dos listas diferentes de datos. Los bloques de vídeo - B pueden ser previstos a partir de dos listas de datos procedentes de dos tramas anteriores, dos listas de datos procedentes de tramas posteriores, o una lista de datos procedente de una trama anterior y una procedente de una trama posterior. Por contra, los bloques de vídeo - P son previstos en base a una lista, la cual puede corresponder a una trama predictiva, por ejemplo una trama anterior o una trama posterior. Las tramas - B, y las tramas - P pueden ser designadas en términos más generales como unidades - P y unidades - B. Las unidades - P y las unidades - B pueden también materializarse en unidades codificadas de menor tamaño, como por ejemplo rebanadas de tramas o porciones de tramas. Las unidades - P pueden incluir unidades de vídeo - B, bloques de vídeo - P o bloques de vídeo - I. Las unidades - P pueden incluir bloques de vídeo - P o bloques de vídeo - I. Las unidades - I pueden incluir solo bloques de vídeo - I.

La puesta en correspondencia de plantillas es una técnica que puede ser utilizada para eliminar vectores de movimiento pero al mismo tiempo proporcionar las ventajas de la codificación de vídeo de movimiento compensado. En la puesta en correspondencia de plantillas, una plantilla define una forma de píxeles vecinos con respecto a un bloque de vídeo que está siendo codificado. La forma de plantilla se define con respecto al bloque de vídeo. La forma de plantilla puede lindar con el bloque de vídeo o rodear el bloque de vídeo. Sin embargo, no se requiere que la plantilla esté situada inmediatamente en posición adyacente al bloque de vídeo. Esto es, uno o más píxeles del bloque de vídeo pueden existir entre el bloque más próximo de la forma de plantilla y el bloque de vídeo.

Una hipótesis, en un ejemplo, se puede basar en los datos de vídeo predictivos contenidos dentro de una forma de plantilla. La forma de plantilla puede ser aplicada a uno o más bloques de vídeo dentro de una unidad de vídeo de referencia. Una hipótesis se refiere a una forma de plantilla correspondiente de una unidad de vídeo de referencia, la cual se define con respecto a un bloque de los datos de vídeo predictivos dentro de la unidad de vídeo de referencia. Una hipótesis se puede basar en un primer conjunto de datos predictivos procedentes de al menos una porción de una primera unidad de vídeo predictiva. Una hipótesis se puede también basar en un primer conjunto de datos de vídeo predictivos procedente de al menos una porción de una unidad de vídeo predictiva y en base a un segundo conjunto de datos de vídeo predictivos procedente de al menos una porción de una segunda unidad de vídeo predictiva. Las hipótesis que son generadas y consideradas durante la predicción de movimiento de puesta en correspondencia de plantillas pueden ser designadas como hipótesis candidatas. Pueden existir múltiples hipótesis candidatas para cada bloque de vídeo que está siendo codificado.

Los datos definidos por la plantilla asociada con el bloque de vídeo que está siendo codificado (más que por el bloque de vídeo propiamente dicho que está siendo codificado) pueden ser comparados con los datos de una correspondiente plantilla dentro de una unidad de vídeo anterior o posterior. En base a la comparación, el algoritmo de predicción de movimiento de puesta en correspondencia de plantillas identifica la mejor hipótesis a partir del conjunto de hipótesis candidato. Se pueden considerar múltiples hipótesis en la unidad de vídeo de referencia durante el proceso de codificación para crear una probabilidad mayor de que se encuentre la hipótesis puesta en correspondencia más próxima en la unidad de vídeo de referencia. El bloque de vídeo correspondiente a la hipótesis puesta en correspondencia más próxima puede ser el bloque de vídeo más próximo al bloque de vídeo actual y, por tanto, el bloque de vídeo codificado resultante puede contener menos valores residuales. El bloque de vídeo actual es el bloque de vídeo que está siendo codificado y se puede designar como el bloque de vídeo destinatario. El codificador 22 de vídeo y el codificador 28 de vídeo pueden llevar a cabo el proceso de puesta en correspondencia de plantillas para identificar el movimiento sin el uso de vectores de movimiento. De esta manera, con la puesta en correspondencia de plantillas, los vectores de movimiento no son codificados en el flujo de bits. Por el contrario, los vectores de movimiento son esencialmente derivados del proceso de puesta en correspondencia de plantillas cuando la unidad de vídeo es codificada y descodificada.

La FIG. 2 es un diagrama de bloques que ilustra un ejemplo de un codificador 40 de vídeo que puede llevar a cabo tanto técnicas de predicción de movimiento de puesta en correspondencia de plantillas de acuerdo con la presente divulgación, como técnicas de predicción de movimiento basadas en vectores. El codificador 40 de vídeo es un ejemplo de un dispositivo o aparato informático de vídeo especializado designado en la presente memoria como "codificador". El codificador 40 de vídeo puede corresponderse con el codificador 22 de vídeo del dispositivo 12 de fuente, o con un codificador de vídeo de un dispositivo diferente. El codificador 40 de vídeo puede llevar a cabo intra e intercodificaciones de bloques dentro de las tramas de vídeo, aunque los componentes de intracodificación no se muestran en la FIG. 2 para facilitar la ilustración. La intracodificación se basa en la predicción espacial para reducir o eliminar la redundancia espacial en vídeo dentro de una trama de vídeo determinada. La intercodificación se basa en la predicción temporal para reducir o eliminar la redundancia temporal en vídeo dentro de tramas adyacentes de una secuencia de vídeo. Un intramodo (modo - I) puede referirse a un modo basado en compresión espacial, y los Intramodos - I como por ejemplo una predicción (modo - P) o uno bidireccional (modo - B) se pueden referir a los modos de compresión de base temporal. Las unidades codificadas que utilizan el modo - B son codificadas con referencia a dos tramas. En el H.264 y en otros estándares, la trama - B, u otra unidad, puede ser codificada con referencia a dos tramas temporalmente adyacentes. Por ejemplo, si la unidad actual que va a ser codificada se aloja dentro de una trama N, entonces las tramas N - 1 y N - 2 pueden ser utilizadas como tramas de referencia para codificar la unidad actual. En otro ejemplo, las tramas N + 1 y N + 2 pueden ser utilizadas como tramas de referencia para codificar la unidad actual. En algunos casos, al menos una trama temporalmente futura (por ejemplo, la trama N + 1) y al menos una trama temporalmente anterior (por ejemplo la trama N - 1) pueden ser utilizadas para codificar la unidad actual que se aloja en la trama N. Muchas veces, solo una porción de las tramas anteriores o posteriores son de hecho tomadas en consideraciones o utilizadas en la codificación.

Como se muestra en la FIG. 2, el codificador 40 de vídeo recibe un bloque de vídeo dentro de una unidad de vídeo destinada a ser codificada. En el ejemplo de la FIG. 2, el codificador 40 de vídeo incluye una unidad 42 de predicción de puesta en correspondencia de plantillas ("Unidad 42 de predicción de TM") que comprende la unidad 43 de estimación de movimiento de puesta en correspondencia de plantillas (ME de TM) y la unidad 45 de compensación de movimiento de puesta en correspondencia de plantillas (MC de TM). El codificador de vídeo incluye, así mismo, una memoria 44, un sumador 46, una unidad 48 de transformación, una unidad 50 de cuantificación, y una unidad 52 de codificación de entropía. Para una reconstrucción de bloques de vídeo, el codificador 40 de vídeo incluye también una unidad 54 de cuantificación inversa, una unidad 56 de transformación inversa y un sumador 58. El codificador 40 de vídeo puede, así mismo, incluir un filtro desbloqueante (no mostrado) para filtrar las fronteras de los bloques para eliminar los artefactos de bloqueo del vídeo reconstruido. Si se desea el filtro desbloqueante típicamente filtraría la salida del sumador 58. Aunque no se muestra en la FIG. 2, el codificador 40 de vídeo puede también incluir una unidad de predicción para llevar a cabo técnicas de predicción de movimiento basadas en vectores. La unidad de predicción puede generar unos modos de predicción y unos vectores de movimiento, los cuales comprendan unos elementos sintácticos que puedan ser utilizados para identificar los bloques destinatarios utilizados para codificar el bloque de vídeo actual. La unidad de predicción puede incluir unidades de estimación de movimiento y de compensación de movimiento.

Durante el proceso de codificación, el codificador de vídeo recibe un bloque de vídeo destinado a ser codificado, y la unidad 43 de ME de TM y la unidad 45 de MC de TM llevan a cabo una codificación interpredictiva. La unidad 43 de ME de TM y la unidad 45 de MC de TM pueden estar muy integradas, pero se ilustran por separados con fines conceptuales. La estimación de movimiento se considera típicamente como el proceso de generar vectores de movimiento, los cuales estiman el movimiento del bloque de vídeo. Un vector de movimiento, por ejemplo, puede indicar el desplazamiento del bloque predictivo dentro de una trama predictiva (u otra unidad codificada) con respecto al bloque actual que está siendo codificado dentro de la trama actual (u otra unidad codificada). La compensación de movimiento se considera típicamente como el proceso de extracción o generación del bloque predictivo en base al vector de movimiento determinado por la estimación de movimiento. De nuevo aquí, la unidad 43 de la ME de TM y la unidad 45 de la MC de TM pueden estar funcionalmente integradas. Con fines demostrativos, las técnicas descritas en la presente divulgación se describen como llevadas a cabo por la unidad 42 de predicción de la TM.

En un ejemplo, la unidad 42 de predicción de la TM aplica la predicción de movimiento de puesta en correspondencia de plantillas para codificar el bloque de vídeo actual, por ejemplo, el bloque de vídeo destinatario. La unidad 42 de predicción de TM puede, por ejemplo, comparar los píxeles existentes dentro de una forma de plantilla definida con respecto al bloque destinatario que está siendo codificado con los datos correspondientes de una unidad de vídeo anterior y / o posterior. Al suministrar la predicción de movimiento de puesta en correspondencia de plantillas, la unidad 42 de predicción de TM puede seleccionar una unidad de vídeo anterior o una unidad de vídeo futura procedente de un primer conjunto de unidades de vídeo (el cual puede ser designado como lista 0) y determinar un conjunto de hipótesis que utilicen solo esta unidad de vídeo anterior o futura procedente de la lista 0. Como alternativa la unidad 42 de predicción de TM puede seleccionar una unidad de vídeo anterior o futura procedente de un conjunto de unidades de vídeo (el cual puede ser designado como lista 1) y determinar un conjunto de hipótesis utilizando solo esta unidad de vídeo anterior o futura procedente de la lista 1.

Para hacer posible la puesta en correspondencia de plantillas sobre las tramas - B, pueden ser utilizados dos conjuntos de hipótesis para la trama - B. Si solo se utiliza un conjunto de hipótesis, el conjunto de hipótesis puede contener hipótesis encontradas tanto en la lista 0 como en la lista 1. Si se utilizan dos conjuntos de hipótesis, un

conjunto de hipótesis puede contener las hipótesis encontradas en la lista 1 y otro conjunto de hipótesis puede contener las hipótesis encontradas en la lista 0. Así mismo, puede formarse un tercer conjunto de hipótesis a partir de la predicción de movimientos bidireccionales, en la que los datos procedentes de la lista 0 y los datos procedentes de la lista 1 sean utilizados con fines de predicción, posiblemente de una forma ponderada para definir los datos del tercer conjunto de hipótesis. El conjunto de hipótesis para cualquier lista determinada puede incluir una hipótesis concreta que identifique un bloque de vídeo predictivo deseable que sea el más parecido al bloque de vídeo que está siendo codificado. Una métrica, como por ejemplo la suma de las diferencias absolutas (SAD) o la suma de la diferencia cuadrática (SSD), pueden ser utilizadas para encontrar la mejor hipótesis, la cual, a su vez, identifique el bloque de vídeo predictivo que debe ser utilizado para el codificado.

La unidad 42 de predicción de TM selecciona una o más hipótesis procedentes del conjunto generado de hipótesis durante el proceso de codificación. En un ejemplo, un elemento sintáctico es introducido con el fin de señalar el número de hipótesis utilizado en la predicción de movimiento de la puesta en correspondencia de plantillas. En este ejemplo, dos números de hipótesis diferentes, por ejemplo, K_1 y K_2 pueden ser disponibles. K_1 representa un primer número de hipótesis que puede ser utilizado al codificar y descodificar el bloque destinatario y K_2 representa un segundo número, por ejemplo un número máximo de hipótesis que puede ser utilizado al codificar y descodificar el bloque destinatario. Por ejemplo, K_1 puede ser cuatro y K_2 puede ser ocho. Para determinar el número de hipótesis que hay que utilizar para codificar el bloque destinatario, la unidad 42 de predicción de TM puede generar un primer conjunto de datos de vídeo predictivos en base a las hipótesis K_1 y un segundo conjunto de datos de vídeo predictivos en base a las hipótesis K_2 . La unidad 42 de predicción de TM puede generar el primer conjunto de datos de vídeo predictivos mediante la identificación de un subconjunto de hipótesis que contenga el número K_1 de las mejores hipótesis procedentes del conjunto de todas las hipótesis posibles. Las mejores hipótesis dentro de un subconjunto determinado pueden ser las hipótesis que se correspondan con más exactitud a los datos existentes dentro de la forma de plantilla definida con respecto al bloque de vídeo destinatario que está siendo codificado. La suma de la diferencia absoluta (SAD) o la suma de la diferencia cuadrática (SSD) por ejemplo, pueden ser aplicadas para determinar las mejores hipótesis para definir un subconjunto determinado. La unidad 42 de predicción de TM puede identificar los bloques de vídeo de referencia correspondientes a cada una de las hipótesis del subconjunto de hipótesis y promediar los valores de píxeles para cada píxel que presente el mismo emplazamiento relativo dentro de los bloques de vídeo de referencia identificados. La unidad 42 de predicción de TM puede generar un segundo conjunto de datos de vídeo predictivos en base a las hipótesis K_2 sustancialmente de la misma manera. Aunque en este ejemplo, el valor de píxel medio para cada píxel de los bloques de referencia se calcula para generar los conjuntos de datos predictivos asociados con las hipótesis K_1 y / o K_2 , pueden ser utilizados otros cálculos y métricas matemáticas para combinar los diferentes conjuntos de datos predictivos asociados con las hipótesis K_1 y / o K_2 .

Después de que la unidad 42 de predicción de TM genera al menos uno de los primero y segundo conjuntos de datos de vídeo predictivos, la unidad 42 de predicción de TM puede aplicar un multiplicador de Lagrange para determinar el coste de la codificación del bloque destinatario al utilizar los primero y segundo conjuntos de datos de vídeo predictivos. Para determinar el coste, la unidad 42 de predicción de TM, utiliza una función del coste que se define como $C = D + \lambda R$, en la que C es el coste, D es la distorsión, λ es el multiplicador de Lagrange, y R es la tasa de codificación. La unidad 42 de predicción de TM selecciona el número de hipótesis que producirá en último término el coste de la tasa de distorsión más baja y puede identificar el número de hipótesis que fue utilizado por la unidad 42 de predicción de TM al codificar la trama destinataria a través del nuevo elemento sintáctico. Cuando se toman en consideración demasiadas hipótesis, la eficiencia de la codificación de la puesta en correspondencia de plantillas puede resentirse. Mediante la incorporación de un elemento sintáctico para señalar el número de hipótesis, lo que produce el coste menor al codificar un bloque destinatario, se puede tomar en consideración un número más conveniente de hipótesis durante la codificación y la descodificación, mejorando de esta manera la eficiencia de la codificación global de la puesta en correspondencia de plantillas.

En un aspecto de la presente divulgación, la unidad 42 de predicción de TM puede también rechazar las hipótesis que utilizan una técnica de rechazo de hipótesis de mejora. En esta técnica, cada hipótesis objeto de consideración es comparada con una hipótesis de referencia. La hipótesis de referencia puede ser generada promediando los valores de píxeles de todas las hipótesis o de un subconjunto de las mejoras hipótesis sometidas a consideración. Si la diferencia entre las hipótesis sometidas a consideración y las hipótesis de referencia es mayor que un umbral, entonces la hipótesis sometida a consideración puede ser eliminada de la consideración. La diferencia puede ser calculada en base a la suma ponderada de las diferencias absolutas (WSAD), descritas más adelante, o en base a la SAD de los píxeles de la hipótesis sometida a consideración y la hipótesis de referencia, por ejemplo. Las hipótesis pueden ser rechazadas o seleccionadas sin señalar a un descodificador cuáles fueron las hipótesis rechazadas o seleccionadas.

De acuerdo con otro aspecto de la presente divulgación, la SAD puede ser ponderada para mejorar la selección de hipótesis. La SAD es la suma del valor absoluto de la diferencia entre el valor de píxel de cada píxel del bloque actual y del correspondiente píxel del bloque de referencia. La plantilla del bloque actual y la plantilla del bloque de referencia pueden ser divididas en K partes. Como se muestra en la FIG. 3, la unidad 62 de vídeo destinataria incluye un bloque 64 destinatario y una plantilla 66 de bloque destinatario y una unidad 68 de vídeo de referencia incluye un bloque 70 de referencia y la plantilla 72 del bloque de referencia. En este ejemplo, la plantilla 66 del bloque destinatario y la plantilla 73 del bloque de referencia están cada una divididas en cuatro particiones de plantilla. La ponderación otorgada a cada partición de plantilla concreta de la plantilla 66 de bloque destinatario o de

la plantilla 72 de bloque de referencia puede disminuir cuando la distancia entre cada partición de plantilla y el bloque 64 destinatario o el bloque 70 de referencia aumenta. La suma ponderada de las referencias absolutas (WSAD) se puede definir como:

$$WSAD = \sum_{k=1}^K \sum_{n_k} w_k \cdot |T_{Pk} - T_{Bk}|$$

- 5 donde w_k es el valor ponderado de la partición de plantilla K ésima, T_{Pk} es la partición de plantilla K ésima asociada con el bloque de referencia, T_{Bk} es la partición de plantilla k ésima asociada con el bloque destinatario, n_k es el número de píxeles de la partición k de plantilla y K es el número de particiones en la que está dividida cada plantilla asociada con el bloque destinatario y la plantilla asociada con el bloque de referencia.

- 10 La unidad 42 de predicción de TM puede, así mismo, aplicar un desplazamiento calculado de plantilla para predecir el desplazamiento entre el bloque de referencia y el bloque destinatario para mejorar la calidad de la codificación de vídeo. Al efectuar la búsqueda de la mejor hipótesis de puesta en correspondencia definida por la forma de plantilla, la unidad 42 de predicción de TM puede calcular la diferencia media entre el valor de píxel de cada píxel dentro de la plantilla definida con respecto al bloque destinatario y el valor de píxel de la hipótesis candidato utilizando la siguiente fórmula:

$$\text{offset} = \frac{1}{K} \sum_{K=1}^K (T_{Bk} - T_{Pk})$$

- 15 donde T_{Bk} es un valor de píxel de un píxel de la hipótesis candidata, T_{Pk} es un valor de píxel de un píxel en la plantilla definida con respecto al bloque destinatario, K es el número de píxeles de la plantilla definida con respecto al bloque destinatario y **offset** es la diferencia media en el valor de píxel entre los píxeles de la hipótesis candidata y la plantilla definida con respecto al bloque destinatario. Esta diferencia media en valores de píxel es conocida como desplazamiento de plantilla. Para cada hipótesis candidata del conjunto de hipótesis potenciales, la unidad 42 de predicción de TM puede calcular un desplazamiento de este tipo.

- 20 Una vez que el desplazamiento de plantilla es calculado para una hipótesis candidata concreta, la unidad 42 de predicción de TM puede calcular la suma de las diferencias absolutas respecto de la hipótesis candidata tomando cada valor de píxel de la hipótesis candidata, restando el valor de píxel del píxel correspondiente de la plantilla de referencia y después restando el desplazamiento. La suma de las diferencias absolutas de la hipótesis candidata se define como:

$$SAD = \sum_{K=1}^K |T_{Pk} + \text{offset} - T_{Bk}|$$

- 30 donde T_{Bk} es un valor de píxel de un píxel en la hipótesis candidata, T_{Pk} es un valor de píxel de un píxel en la plantilla definida con respecto al bloque destinatario, K es el número de píxeles en la hipótesis candidato y **offset** es la diferencia media en el valor de píxel entre los píxeles en la hipótesis candidata y los píxeles de la plantilla definida con respecto al bloque destinatario. El desplazamiento de plantilla obtenido se supone que es el desplazamiento objeto de consideración, la predicción de movimiento compensado del bloque destinatario puede ser representada como:

$$B' = P + \text{offset}$$

- 35 donde B' es la predicción de movimiento compensado del bloque destinatario, P es el valor de píxel de un píxel del bloque de referencia, y **offset** es el desplazamiento medio de la plantilla según se ha calculado con anterioridad. Mediante el ajuste de la suma de la diferencia absoluta del bloque destinatario por el desplazamiento de plantilla, es probable que las hipótesis seleccionadas sean de hecho las mejores hipótesis, reduciendo con ello las residuales y disminuyendo la cantidad de información que debe ser almacenada.

- 40 En la FIG. 4, la plantilla 74 se define con respecto a un sub-bloque 76 destinatario y se superpone a una porción del sub-bloque 78 donde el sub-bloque 76 y el sub-bloque 78 son particiones del mismo bloque de vídeo. Si el sub-bloque 78 no presenta píxeles reconstruidos disponibles el contenido del sub-bloque 78 se puede determinar mediante el resultado de predicción de movimiento compensado, por ejemplo. Si el contenido del sub-bloque 78 se determina mediante la utilización del resultado de predicción del movimiento compensado, la precisión de la predicción de movimiento del sub-bloque 76 se puede resentir debido a que el resultado de la predicción de la compensación del movimiento del sub-bloque 78 puede diferir significativamente de los datos originales. Para mejorar la precisión de la predicción de movimiento del sub-bloque 76 destinatario cuando el sub-bloque 78 no incorpora píxeles reconstruidos disponibles, una unidad 42 de predicción puede esperar hasta que el sub-bloque 78

haya sido codificado antes de la codificación del sub-bloque 76 destinatario. Mediante la espera hasta que el sub-bloque 78 haya sido codificado, el contenido del sub-bloque 78 está más próximo a los datos originales y la predicción de movimiento de puesta en correspondencia de plantillas del sub-bloque 76 destinatario puede ser más precisa.

En un ejemplo, el sub-bloque 76 y el sub-bloque 78 son sub-bloques de un bloque de vídeo luma. En general, un bloque de vídeo luma es un bloque de 16 por 16 valores luma para píxeles dentro del bloque de vídeo. Un bloque de vídeo luma puede ser dividido en múltiples sub-bloques de 8 por 8, 8 por 4, o 4 por 4 sub-bloques de píxel, por ejemplo. Para determinar el contenido del sub-bloque 78, la unidad 42 de predicción de TM, en primer término, calcula la predicción de la compensación de movimiento y el residual luma para el sub-bloque 78. La predicción de movimiento compensado incluye los valores de píxel del correspondiente sub-bloque en la unidad de vídeo de referencia. El residual luma es la diferencia en valores luma del sub-bloque 78 y del correspondiente sub-bloque en la unidad de vídeo de referencia. La unidad 42 de predicción de TM combina entonces la predicción de la compensación de movimiento y el residual luma codificado en resultados luma para determinar el contenido del sub-bloque 78. Una vez que la unidad 42 de predicción de TM ha determinado el contenido del sub-bloque 78, los valores de píxel están disponibles para la plantilla 74 y la unidad 42 de predicción de TM puede proceder con la ejecución de la predicción de movimiento de puesta en correspondencia de plantillas respecto del sub-bloque 76.

La FIG. 5 es un diagrama de bloques que ilustra un ejemplo de un descodificador 80 de vídeo, el cual lleva a cabo las técnicas de descodificación recíprocas a las técnicas de codificación descritas en las líneas anteriores, como por ejemplo la aplicación de predicción de movimiento de puesta en correspondencia de plantillas a las tramas - P y a la tramas - B, la aplicación de un número de hipótesis definido por un elemento sintáctico, el cálculo y la aplicación de una WSAD, y / o el cálculo y la aplicación de un desplazamiento de plantilla. El descodificador 80 de vídeo puede incluir una unidad 82 de descodificación de entropía, una unidad 84 de predicción de puesta en correspondencia de plantillas (TM), una unidad 86 de cuantificación inversa, una unidad 88 de transformación inversa, una memoria 90 y un sumador 92. La unidad 84 de predicción puede incluir una unidad 85 de estimación de movimiento de puesta en correspondencia de plantillas (ME de TM), una unidad 87 de compensación de movimiento de puesta en correspondencia de plantillas (MC de TM), así como unos componentes de predicción espaciales, los cuales no se muestran para simplificar y facilitar la ilustración.

En general, la unidad 82 de descodificación de entropía recibe un flujo de bits codificado y descodifica el flujo de bits para generar unos coeficientes cuantificados y otros elementos sintácticos. La unidad 84 de predicción puede llevar a cabo una predicción de movimiento de puesta en correspondencia de plantillas de acuerdo con la presente divulgación. Los elementos sintácticos pueden incluir una información que especifique el número de hipótesis consideradas por un codificador de vídeo al codificar la unidad de vídeo codificada. La unidad 84 de predicción de TM puede considerar el número codificado de hipótesis al descodificar la unidad de vídeo mediante la limitación del número de hipótesis utilizado durante el proceso de descodificación al número de hipótesis identificado por el elemento sintáctico. Al descodificar una unidad de vídeo codificada utilizando la pluralidad de hipótesis identificada por el elemento sintáctico, la unidad 84 de predicción de TM puede generar un conjunto de datos de vídeo predictivos mediante la identificación de un subconjunto de hipótesis que contiene la pluralidad de mejores hipótesis identificada por el elemento sintáctico procedente del conjunto de todas las hipótesis posibles. Las mejores hipótesis incluidas en un conjunto determinado pueden ser las hipótesis que más estrechamente se correspondan con los datos de vídeo incluidos en la forma de plantilla definida con respecto al bloque destinatario que está siendo codificado. La SAD o SSD, por ejemplo, pueden ser aplicadas para determinar las mejores hipótesis para definir un subconjunto determinado. La unidad 84 de predicción de TM puede identificar los bloques de vídeo de referencia correspondientes a cada una de las hipótesis del subconjunto de hipótesis y promediar los valores de píxel para cada píxel que presente el mismo emplazamiento relativo dentro de los bloques de vídeo de referencia identificados. Aunque, en el presente ejemplo, el valor de píxel medio para cada píxel de los bloques de referencia se calcula para generar un conjunto de datos de vídeo predictivos asociados con el número de hipótesis identificado por el elemento sintáctico, se pueden utilizar otros cálculos o métricas matemáticas. En general, el mismo cálculo matemático utilizado por el codificador es también utilizado por el descodificador. De esta manera, la unidad 84 de predicción de TM genera un conjunto de datos de vídeo predictivos para la descodificación del bloque de vídeo codificado.

Los coeficientes cuantificados son enviados desde la unidad 82 de descodificación de entropía hasta la unidad 86 de cuantificación inversa, la cual lleva a cabo la cuantificación inversa. La unidad 88 de transformación inversa transforma a la inversa los coeficientes descuantificados hasta el dominio de píxeles para generar un bloque residual. El sumador 92 combina el conjunto de datos predictivos generado por la unidad 84 de predicción de TM con el bloque residual a partir de la unidad 88 de transformación inversa para crear un bloque de vídeo reconstruido, el cual puede ser almacenado en la memoria 90 y / o emitido desde el descodificador 80 de vídeo como salida de vídeo descodificada.

Las FIGS. 6A - 6B son diagramas de flujo que ilustran un proceso ejemplar llevado a cabo por un codificador de vídeo de acuerdo con la presente divulgación. Las FIGS. 6A - 6B se describirán desde la perspectiva del codificador 40 de vídeo de la FIG. 2. Como se muestra en la FIG. 6A, la unidad 42 de predicción de TM genera unas listas (100) de referencia. Las listas de referencia pueden incluir información de vídeo procedente de una o más unidades de vídeo de referencia. Las unidades de vídeo de referencia pueden producirse con anterioridad en el tiempo o más tarde en el tiempo con respecto a una unidad de vídeo que incluya el bloque de vídeo destinatario. La unidad 42 de

predicción selecciona, a continuación, un bloque (102) de vídeo destinatario y define una forma de plantilla con respecto al bloque destinatario. Si el bloque destinatario es un sub-bloque de un bloque luma y alguna porción de la forma de plantilla no presenta píxeles reconstruidos disponibles (NO en la etapa 104), la unidad 42 de predicción de TM genera el contenido de la participación de sub-bloques asociado con la porción de la forma de plantilla que no incorpora píxeles reconstruidos disponibles (106). En concreto, el contenido de la participación en sub-bloques puede ser generado mediante la combinación del residual luma con la predicción del movimiento compensado de la partición de sub-bloques. Una vez que el contenido de la partición de sub-bloques se ha generado (106) o si los píxeles reconstruidos se encuentran disponibles respecto de la entera plantilla asociada con el bloque destinatario (SI en la etapa 104), entonces la unidad 42 de predicción de TM genera un o más conjuntos de hipótesis en base a las listas de referencia y a la plantilla asociada con el bloque destinatario (108).

Para cada hipótesis del conjunto de hipótesis, la unidad 42 de predicción de TM calcula el desplazamiento de plantilla (110) mediante el cálculo de la diferencia media entre cada valor de píxel de la forma de plantilla definida con respecto al bloque destinatario y el valor de píxel de cada píxel de una hipótesis candidata procedente del conjunto de hipótesis. La unidad 42 de predicción de TM puede, a continuación, calcular la suma ponderada de las diferencias absolutas para cada hipótesis (112). En un ejemplo, la unidad 42 de predicción de TM calcula la WSAD utilizando los valores de píxel calculados en base al desplazamiento de plantilla mediante la multiplicación de los valores de píxel calculados por un factor de ponderación, de manera que el factor de ponderación se reduzca a medida que la distancia entre el bloque de vídeo y la partición de plantilla aumente. En otro ejemplo, la unidad 42 de predicción de TM calcula la WSAD mediante la multiplicación de los valores de píxel calculados mediante un valor de ponderación, de manera que, el factor de ponderación disminuya a medida que la distancia entre el bloque de vídeo y la partición de plantilla aumente, pero sin ajustar los valores de píxel basados en el desplazamiento de plantilla. La unidad 42 de predicción de TM puede calcular la WSAD mejor que la WSAD mediante la eliminación del factor de ponderación. En el cálculo de la WSAD o de la SAD, la unidad 42 de predicción de TM puede incorporar el desplazamiento de plantilla calculado. Una vez que el desplazamiento de plantilla y o bien la WSAD o la SAD son calculadas, la unidad 42 de predicción de TM puede generar una hipótesis de referencia (114), en un ejemplo, mediante la promediación de todas las hipótesis sometidas a consideración o, en otro ejemplo, mediante la promediación de un subconjunto de las mejoras hipótesis sometidas a consideración.

Con respecto a la FIG. 6B, la unidad 42 de predicción de TM puede, a continuación, seleccionar una de las hipótesis disponibles para ser objeto de consideración (116) y calcular la diferencia entre la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R y comparar la diferencia de valor con un valor de umbral T (118). Este valor de diferencia entre la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R , puede comprender la suma de la diferencia absoluta entre cada píxel de la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R . Si esta diferencia (esto es, si la SAD de los píxeles de la hipótesis objeto de consideración y la hipótesis de referencia) es mayor que un umbral T (sí de la etapa 118), entonces esa hipótesis puede ser eliminada de la consideración (120). Sin embargo, si la diferencia entre los píxeles de H_K y los píxeles de H_R son inferiores a T (NO en la etapa 118) o después de que H_K es eliminada de la consideración (120), la unidad 42 de predicción de TM determina si hay hipótesis que restan en el conjunto de hipótesis objeto de consideración (122). Si hay hipótesis que restan, la unidad 42 de predicción de TM selecciona una nueva hipótesis de consideración (116), y determina si la diferencia de valor entre la hipótesis que está siendo considerada y la hipótesis de referencia anteriormente generada es mayor que el valor de umbral (118). Una vez que no hay más hipótesis disponibles para ser consideradas que no hayan sido ya consideradas, (NO en la etapa 122), la unidad 42 de predicción de TM selecciona el número de hipótesis (124) mediante el cálculo del coste asociado con la codificación del bloque de vídeo destinatario cuando un primer número de hipótesis y el coste asociado con la codificación del bloque de vídeo destinatario cuando el segundo número de hipótesis procedente del conjunto de hipótesis son utilizados en la codificación.

Para seleccionar el número de hipótesis que hay que utilizar para la codificación del bloque destinatario (124) la unidad 42 de predicción de TM puede generar un primer conjunto de datos de vídeo predictivos en base a un primer número de hipótesis y un segundo conjunto de datos de vídeo predictivos en base a un segundo número de hipótesis. La unidad 42 de predicción de TM puede generar el primer conjunto de datos de vídeo predictivos mediante la identificación de un subconjunto de hipótesis que contenga el primer número de mejores hipótesis procedentes del conjunto de todas las posibles hipótesis. La unidad 42 de predicción de TM puede, a continuación, identificar los bloques de vídeo de referencia correspondientes a cada una de las hipótesis del subconjunto de hipótesis y promediar los valores de píxel para cada píxel que presente el mismo emplazamiento relativo en los bloques de vídeo de referencia identificados. La unidad 42 de predicción de TM puede generar un segundo conjunto de datos de vídeo predictivos en base al segundo número de hipótesis sustancialmente de la misma manera. Después de que la unidad 42 de predicción de TM genera en los primero y segundo conjuntos de datos de vídeo predictivos, la unidad 42 de predicción de TM determina el coste de la tasa de distorsión para la codificación del bloque destinatario al utilizar los primero o segundo conjuntos de datos de vídeo predictivos. La unidad 42 de predicción de TM selecciona el número de hipótesis que en último término producirán el coste de la tasa de distorsión más baja.

La unidad 42 de predicción de TM genera, a continuación, el elemento sintáctico para el número de hipótesis seleccionado (126). El elemento sintáctico señala al descodificador si el primer número o el segundo número de las hipótesis seleccionadas fueron utilizados al codificar el bloque destinatario (126). El codificador 40 de vídeo puede, a

continuación, codificar el bloque de vídeo (128) en base a un conjunto de datos de vídeo predictivos. En un ejemplo, el codificador 40 de vídeo codifica el bloque de vídeo en base al conjunto de datos de vídeo predictivo asociado con el número de hipótesis que presenta el coste más bajo. En otro ejemplo, la unidad 42 de predicción selecciona la mejor hipótesis y el codificador 40 de vídeo codifica el bloque de vídeo en base a los datos de vídeo predictivos contenidos dentro del bloque de vídeo de referencia asociado con la mejor hipótesis. En otro ejemplo, la unidad 42 de predicción de TM puede generar datos de vídeo predictivos en base a todos los bloques de vídeo de referencia asociados con todas las hipótesis restantes disponibles susceptibles de consideración, y el codificador 40 de vídeo codifica el bloque de vídeo en base a estos datos de vídeo predictivos.

Las FIGS. 7A - 7B son diagramas de flujo que ilustran un proceso ejemplar llevado a cabo por un codificador de vídeo de acuerdo con la presente divulgación. La FIGS. 7A - 7B se describirán desde la perspectiva del descodificador 80 de vídeo de la FIG. 5. Como se muestra en la FIG. 7A, el descodificador 80 de vídeo recibe los datos de vídeo codificados (132), y recibe uno o más elementos sintácticos que identifican el número de hipótesis utilizado por el codificador de vídeo al codificar los datos de vídeo codificados. La unidad 84 de predicción de TM selecciona un bloque de vídeo destinatario, identifica la plantilla asociada con el bloque destinatario, y genera las listas de referencia (134). Las listas de referencia incluyen la información de vídeo procedente de una o más unidades de vídeo de referencia. Las unidades de vídeo de referencia, pueden encontrarse con carácter previo o con posterioridad respecto de una unidad de vídeo que incluya el bloque de vídeo destinatario. Si porciones de la forma de plantilla definida con respecto al bloque destinatario no han reconstruido píxeles disponibles (NO en la etapa 136) y el bloque destinatario es un sub-bloque de un bloque luma, la unidad 84 de predicción de TM puede generar el contenido de la partición en sub-bloques asociado con la porción de la forma de plantilla que no presente píxeles reconstruidos disponibles (138). En particular, el contenido de la partición en sub-bloques puede ser generado mediante la combinación del residual luma con la predicción de movimiento compensado. Una vez que el contenido de la partición de sub-bloques es generado (138) o si los píxeles reconstruidos están disponibles para la entera forma de plantilla definida con respecto al bloque destinatario (SI en la etapa 136), la unidad 42 de predicción de TM genera uno o más conjuntos de hipótesis en base a las listas de referencia y a la plantilla asociada con el bloque destinatario (140).

Para cada hipótesis del conjunto de hipótesis, la unidad 84 de predicción de TM puede calcular un desplazamiento de plantilla (142) mediante el cálculo de la diferencia media entre los valores de píxel de cada píxel de la forma de plantilla definida con respecto al bloque destinatario y el valor de píxel de cada píxel correspondiente de una hipótesis candidato del conjunto de hipótesis. La unidad 84 de predicción de TM puede calcular la suma ponderada de las diferencias absolutas para cada hipótesis (144). En un ejemplo, la unidad 84 de predicción de TM calcula la WSAD utilizando los valores de píxel calculados en base al desplazamiento de plantilla mediante la multiplicación de los valores de píxel calculados por un factor de ponderación, de manera que el factor de ponderación disminuya a medida que la distancia entre el vídeo y la partición de plantilla aumente. En otro ejemplo, la unidad 84 de predicción de TM calcula la WSAD mediante la multiplicación de los valores de píxel calculados por un factor de ponderación, de manera que el factor de ponderación disminuya a medida que la distancia entre el bloque de vídeo y la partición de plantilla aumente, pero sin ajustar los valores de píxel en base al desplazamiento de plantilla. La unidad 84 de predicción de TM puede calcular la SAD mejor que la WSAD mediante la eliminación del factor de ponderación. En el cálculo de la SAD, la unidad 84 de predicción de TM puede incorporar el desplazamiento de plantilla calculado.

Con respecto a la FIG. 7B, la unidad 84 de predicción de TM puede calcular una hipótesis de referencia (146), en un ejemplo, promediando todas las hipótesis sometidas a consideración o, en otro ejemplo, un conjunto de las mejores hipótesis objeto de consideración. La unidad 84 de predicción de TM puede, a continuación, seleccionar una de las hipótesis disponibles para su consideración (148) y calcular la diferencia entre la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R , y comparar la diferencia de valor con un valor de umbral, T (150). Este valor de diferencia entre la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R , puede comprender la suma de la diferencia absoluta entre cada píxel de la hipótesis que está siendo considerada, H_K , y la hipótesis de referencia, H_R . Si esta diferencia (esto es, si la SAD de los píxeles de la hipótesis sometida a consideración y la hipótesis de referencia) es mayor que un umbral T (SI en la etapa 150), entonces esa hipótesis puede ser eliminada de la consideración. Si la diferencia de valor es inferior a T (NO en la etapa 150) o después de que H_K sea eliminada de la consideración (152), la unidad 84 de predicción de TM determina si hay cualquier hipótesis restante en el conjunto de hipótesis sometido a consideración (154). Si hay hipótesis que restan, la unidad de predicción selecciona una nueva hipótesis de consideración (148), y determina si la diferencia de valor entre la hipótesis que está siendo considerada y la hipótesis de referencia anteriormente generada es mayor que un valor de umbral (150). Una vez que no hay más hipótesis disponibles como consideración que no hayan sido ya consideradas, (NO en la etapa 154), la unidad 84 de predicción de TM puede descodificar un elemento sintáctico que identifique el número de hipótesis consideradas al codificar el bloque destinatario y aplicar el número de hipótesis (156).

Para aplicar el número de hipótesis (156), la unidad 84 de predicción de TM genera un conjunto de datos de vídeo predictivos mediante la identificación de un subconjunto de hipótesis que contiene el número de mejores hipótesis identificados por el elemento sintáctico procedente del conjunto de todas las hipótesis posibles. La unidad 84 de predicción de TM puede entonces identificar los bloques de vídeo de referencia correspondientes a cada una de las hipótesis del subconjunto de hipótesis y promediar los valores de píxel para cada píxel que presente el mismo emplazamiento relativo en los bloques de vídeo de referencia identificados. El descodificador 80 de vídeo puede

entonces descodificar el vídeo en base a los datos de vídeo predictivos (158). En un ejemplo, un descodificador 80 de vídeo descodifica el bloque de vídeo en base al conjunto de datos de vídeo predictivos asociado con el número de hipótesis que presente el coste más bajo. En otro ejemplo, la unidad 84 de predicción de TM selecciona la mejor hipótesis y el descodificador 80 de vídeo descodifica el bloque de vídeo en base a los datos de vídeo predictivos contenidos dentro del bloque de vídeo de referencia asociado con la mejor hipótesis. En otro ejemplo, la unidad 84 de predicción de TM puede generar datos de vídeo predictivos en base a todos los bloques de vídeo de referencia asociados con todas las hipótesis restantes disponibles de consideración y el descodificador 80 de vídeo codifica el bloque de vídeo en base a estos datos de vídeo predictivos.

Un codificador de vídeo que opere de acuerdo con la presente divulgación puede omitir una o más etapas de los procedimientos ejemplares mostrados en las FIGS. 6A - 6B y 7A - 7B. Por ejemplo, un codificador de vídeo puede no calcular un desplazamiento de plantilla, calcular una suma ponderada de las diferencias absolutas, generar una hipótesis de referencia o utilizar una sintaxis que especifique un número de hipótesis. En general, un codificador de vídeo llevará a cabo las técnicas de predicción de movimiento de puesta en correspondencia de plantillas descritas en la presente divulgación seleccionando al menos un bloque destinatario, generando un conjunto de hipótesis y codificando una unidad de vídeo en base al conjunto de hipótesis.

Las técnicas de la presente divulgación pueden ser implementadas en una amplia variedad de dispositivos o aparatos, incluyendo un aparato telefónico inalámbrico y un circuito integrado (CI) o un conjunto de CIs (esto es, un conjunto de chips). Cualquier componente, módulo o unidad se ha descrito para subrayar los aspectos funcionales y no requiere necesariamente la puesta en práctica por parte de unidades de hardware diferentes. Las técnicas descritas en la presente memoria pueden, así mismo, ser implementadas en hardware, software, firmware o cualquier combinación de estos. Cualquier característica descrita como módulo, unidad o componente puede ser implementada de manera conjunta en un dispositivo lógico integrado o por separado como dispositivo lógico discreto pero interoperable. En algunos casos, diversas características pueden ser implementadas como un dispositivo de circuito integrado, como por ejemplo un chip de circuito integrado o un conjunto de chips.

Si se implementan en software, las técnicas pueden llevarse a cabo, al menos en parte, por un medio legible por computadora que comprenda unas instrucciones que, una vez ejecutadas en un procesador, lleven a cabo uno o más de los procedimientos descritos con anterioridad. El medio legible por computadora puede comprender un medio de almacenamiento legible por computadora y puede formar parte de un producto de programa informático, el cual puede incluir materiales en paquetes. El medio legible de almacenamiento por computadora puede comprender una memoria de acceso aleatorio (RAM) como por ejemplo de una memoria de acceso aleatorio dinámica síncrona (SDRAM), una memoria de solo lectura (ROM), una memoria de acceso aleatorio no volátil (NVRAM), una memoria de solo lectura programable eléctricamente borrrable (EEPROM), una memoria FLASH, medios de almacenamiento de datos magnéticos u ópticos, y similares. Las técnicas, adicionalmente o como alternativa, pueden llevarse a cabo, al menos en parte, por un medio de comunicación legible por computadora que incorpore o comunique un código bajo la forma de instrucciones o estructuras de datos y al que se pueda acceder, que se pueda leer y / o que pueda ser ejecutado por una computadora.

El código o las instrucciones pueden ser ejecutadas por uno o más procesadores, como por ejemplo uno o más procesadores digitales de la señal (DSPs), microprocesadores de propósito general, circuitos integrados específicos de la aplicación (ASICs), matrices lógicas programables sobre el terreno (FPGAs), u otro conjunto de circuitos lógicos equivalentes integrados o discretos. De acuerdo con ello, el término "procesador", según se utiliza en la presente memoria, puede referirse a cualquier estructura precedente o a cualquier estructura apropiada para la implementación de las técnicas descritas en la presente memoria. Así mismo, en algunos aspectos, la funcionalidad descrita en la presente memoria puede estar dispuesta dentro de módulos de software dedicados o de módulos de hardware configurados para la codificación y descodificación, o incorporados en un códec combinado. Así mismo, las técnicas podrían ser completamente implementadas en uno o más circuitos o elementos lógicos.

Así mismo, la divulgación contempla cualquier combinación de dispositivos de circuito integrado que incluyan un conjunto de circuitos para implementar una o más de las técnicas descritas en la presente divulgación. Dicho conjunto de circuitos puede estar dispuesto en un chip único de circuito integrado o en múltiples chips de circuito integrado interoperables dentro de un llamado conjunto de chips. Dichos dispositivos de circuito integrado pueden ser utilizados en una diversidad de aplicaciones, algunas de las cuales pueden incluir su uso en dispositivos de comunicaciones inalámbricas, como por ejemplo aparatos telefónicos móviles.

Se han descrito diversas formas de realización de la invención. Estas y otras formas de realización se incluyen en el alcance de las reivindicaciones subsecuentes.

REIVINDICACIONES

1.- Un procedimiento de codificación de vídeo para la codificación de un bloque de vídeo actual de una unidad de vídeo actual, comprendiendo el procedimiento:

5 la generación de un primer conjunto de hipótesis para una codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del primer conjunto se basan en tramas individuales de datos procedentes de al menos una lista que contiene tramas anteriores en orden de representación y una lista que contiene tramas futuras en orden de representación;

10 la generación de un segundo conjunto de hipótesis para una codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del segundo conjunto están formadas a partir de la predicción de movimientos bidireccionales entre pares de tramas, en el que cada par de tramas comprende una trama procedente de la lista que contiene las tramas anteriores en orden de representación y una trama de la lista que contiene tramas futuras en orden de representación;

15 en el que cada una de las hipótesis de los primero y segundo conjuntos de hipótesis comprende datos de vídeo correspondientes a una forma de plantilla definida con respecto a un emplazamiento de bloque de vídeo;

la aplicación de las técnicas para la selección de las hipótesis entre los primero y segundo conjuntos de hipótesis; y

la codificación, por medio de un codificador de vídeo, del bloque de vídeo actual utilizando los datos de vídeo predictivos procedentes de las hipótesis seleccionadas.

20 2.- El procedimiento de codificación de vídeo de la reivindicación 1, en el que el codificador de vídeo comprende un descodificador de vídeo, en el que el procedimiento comprende también:

el cálculo de un primer coste asociado con la codificación del bloque de vídeo actual, en el que el primer coste se calcula mediante la adición de una distorsión más un producto de un multiplicador de Lagrange por una tasa de codificación para el primer conjunto de hipótesis;

25 el cálculo de un segundo coste asociado con la codificación de bloque de vídeo actual, en el que el segundo coste se calcula mediante la adición de una distorsión más un producto de un multiplicador de Lagrange por una tasa de codificación para el segundo conjunto de hipótesis;

la determinación del coste más bajo asociado con la codificación del bloque de vídeo actual entre el primer coste y el segundo coste; y

30 el establecimiento de un elemento sintáctico para señalar si el primer conjunto de hipótesis o el segundo conjunto de hipótesis deben ser utilizados por un descodificador.

35 3.- El procedimiento de codificación de vídeo de la reivindicación 1, en el que la generación de los conjuntos de hipótesis comprende el cálculo de una suma ponderada de diferencias absolutas, en el que el cálculo de la suma ponderada incluye la división de una plantilla asociada con el bloque de vídeo actual en una pluralidad de particiones y la multiplicación de una suma de las diferencias absolutas de cada partición de la pluralidad de particiones por un valor que disminuye a medida que aumenta una distancia de la partición sometida a consideración a partir del bloque de vídeo actual.

4.- El procedimiento de codificación de vídeo de la reivindicación 1, en el que la generación de los conjuntos de hipótesis comprende:

40 el cálculo de un desplazamiento de plantilla, en el que el cálculo comprende el cálculo de la diferencia media en valores de píxeles de una plantilla definida con respecto al bloque de vídeo actual y en valores de píxeles de una primera hipótesis del conjunto de hipótesis; y

45 la aplicación del desplazamiento de la plantilla a la diferencia entre cada uno de los valores de píxeles de la primera hipótesis y cada uno de los valores de píxeles de la plantilla definida con respecto al bloque de vídeo actual.

5.- Un aparato de codificación de vídeo que codifica un bloque de vídeo actual, comprendiendo el aparato:

50 una unidad de predicción que genera un primer conjunto de hipótesis y un segundo conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del primer conjunto se basan en tramas individuales de datos procedentes de al menos una lista que contiene tramas anteriores en orden de representación y una lista que contiene tramas futuras en orden de representación, y las hipótesis del segundo conjunto se forman a partir de una predicción de movimientos bidireccionales entre partes de tramas, en el que cada par de tramas comprende una trama procedente de

- 5 una lista que contiene tramas anteriores en orden de representación y una trama procedente de la lista que contiene tramas futuras en orden de representación, en el que cada una de las hipótesis de los primero y segundo conjuntos de hipótesis comprende datos de vídeo correspondientes a una forma de plantilla con respecto a un emplazamiento de bloque de vídeo y aplica unas técnicas para la selección de hipótesis entre los primero y segundo conjuntos de hipótesis; y
- una unidad de codificación de vídeo que codifica el bloque de vídeo actual utilizando los datos de vídeo predictivos procedentes de las hipótesis seleccionadas.
- 10 6.- El aparato de codificación de vídeo de la reivindicación 5, en el que la unidad de codificación de vídeo es un codificador de vídeo, y en el que la unidad de predicción calcula un primer coste asociado con la codificación del bloque de vídeo actual, el primer coste es calculado mediante la adición de una distorsión más un producto de un multiplicador de Lagrange por una tasa de codificación para el primer conjunto de hipótesis, calcula un segundo coste asociado con la codificación del bloque de vídeo actual, el segundo coste es calculado mediante la adición de una distorsión más un producto de un multiplicador de Lagrange por una tasa de codificación del segundo conjunto de hipótesis, determina el coste más bajo asociado con la codificación del bloque de vídeo actual entre el primer coste y el segundo coste, y establece un elemento sintáctico para señalar si el primer conjunto de hipótesis o el segundo conjunto de hipótesis deben ser utilizados por un descodificador.
- 15 7.- El aparato de codificación de vídeo de la reivindicación 5, en el que la unidad de predicción genera los conjuntos de hipótesis calculando, al menos, una suma ponderada de diferencias absolutas, en el que la suma ponderada de diferencias absolutas se calcula al menos dividiendo una plantilla asociada con el bloque de vídeo actual en una pluralidad de particiones y multiplicando una suma de las diferencias absolutas de cada partición de la pluralidad de particiones por un valor que disminuye a medida que aumenta una distancia de la partición respecto del bloque de vídeo actual
- 20 8.- El aparato de codificación de vídeo de la reivindicación 5, en el que la unidad de predicción genera los conjuntos de hipótesis calculando al menos un desplazamiento de plantilla mediante el cálculo de al menos la diferencia media de valores de píxeles de una plantilla definida con respecto al bloque de vídeo actual y en valores de píxeles de una primera hipótesis del conjunto de hipótesis, y la aplicación de un desplazamiento de plantilla a la diferencia entre cada uno de los valores de píxeles de la primera hipótesis y cada uno de los valores de píxeles de la plantilla definida con respecto al bloque de vídeo actual.
- 25 9.- El aparato de codificación de vídeo de la reivindicación 5, en el que el aparato de codificación de vídeo comprende un circuito integrado.
- 30 10.- El aparato de codificación de vídeo de la reivindicación 5, en el que el aparato de codificación de vídeo comprende un microprocesador.
- 11.- El aparato de codificación de vídeo de la reivindicación 5, en el que el aparato de codificación de vídeo comprende un dispositivo de comunicación inalámbrico que incluye un codificador de vídeo.
- 35 12.- Un medio de almacenamiento legible por computadora codificado con instrucciones para hacer que uno o más procesadores programables:
- 40 generen un primer conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del primer conjunto se basan en tramas individuales de datos procedentes de al menos una lista que contiene tramas anteriores en orden de representación y una lista que contiene tramas futuras en orden de representación;
- 45 generen un segundo conjunto de hipótesis para la codificación de vídeo predictiva de puesta en correspondencia de plantillas, en el que las hipótesis del segundo conjunto están formadas a partir de la predicción de movimientos bidireccionales entre pares de tramas, en el que cada par de tramas comprende una trama procedente de la lista que contiene tramas anteriores en orden de representación y una trama procedente de la lista que contiene tramas futuras en orden de representación;
- en el que cada una de las hipótesis de los primero y segundo conjuntos de hipótesis comprende unos datos de vídeo correspondientes a una forma de plantilla definida con respecto a un emplazamiento de bloque de vídeo;
- apliquen unas técnicas para la selección de hipótesis entre los primero y segundo conjuntos de hipótesis; y
- 50 codifiquen el bloque de vídeo actual utilizando datos de vídeo predictivos procedentes de las hipótesis seleccionadas.
- 13.- El medio de almacenamiento legible por computadora de la reivindicación 12, en el que las instrucciones que hacen los uno o más procesadores programables codifiquen el bloque de vídeo actual comprenden también instrucciones para hacer que los uno o más procesadores programables:

- calculen un primer coste asociado con la codificación del bloque de vídeo actual, en el que el primer coste se calcula mediante la adición de una distorsión más un producto de un multiplicador de Lagrange por una tasa de codificación para el primer conjunto de hipótesis;
- 5 calculen un segundo coste asociado con la codificación de bloque de vídeo actual, en el que el segundo coste se calcula mediante la adición de una distorsión más un producto de un multiplicador de Lagrange con una tasa de codificación para el subconjunto del conjunto de hipótesis;
- determinen el coste más bajo asociado con la codificación del bloque de vídeo actual entre el primer coste y el segundo coste; y
- 10 establezcan un valor sintáctico para señalar si el primer subconjunto de hipótesis o el segundo conjunto de hipótesis deben ser utilizados por un descodificador.

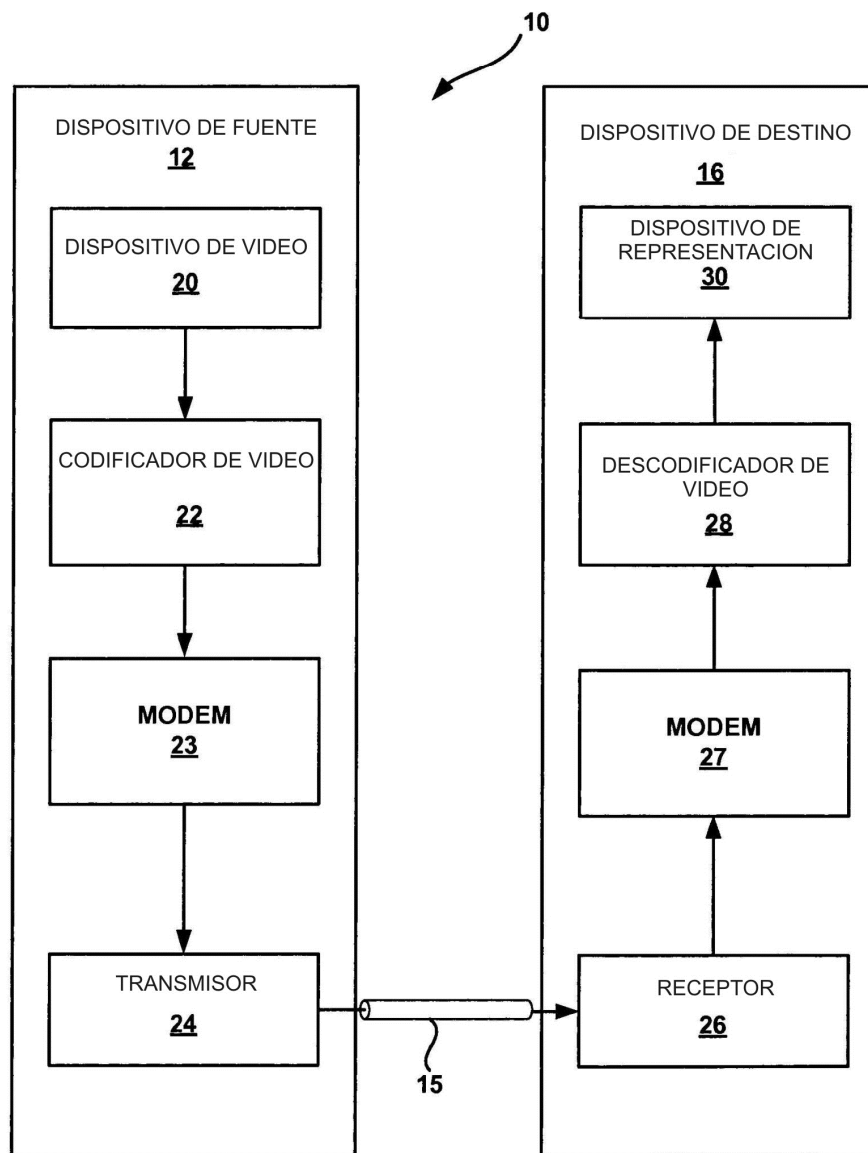


FIG. 1

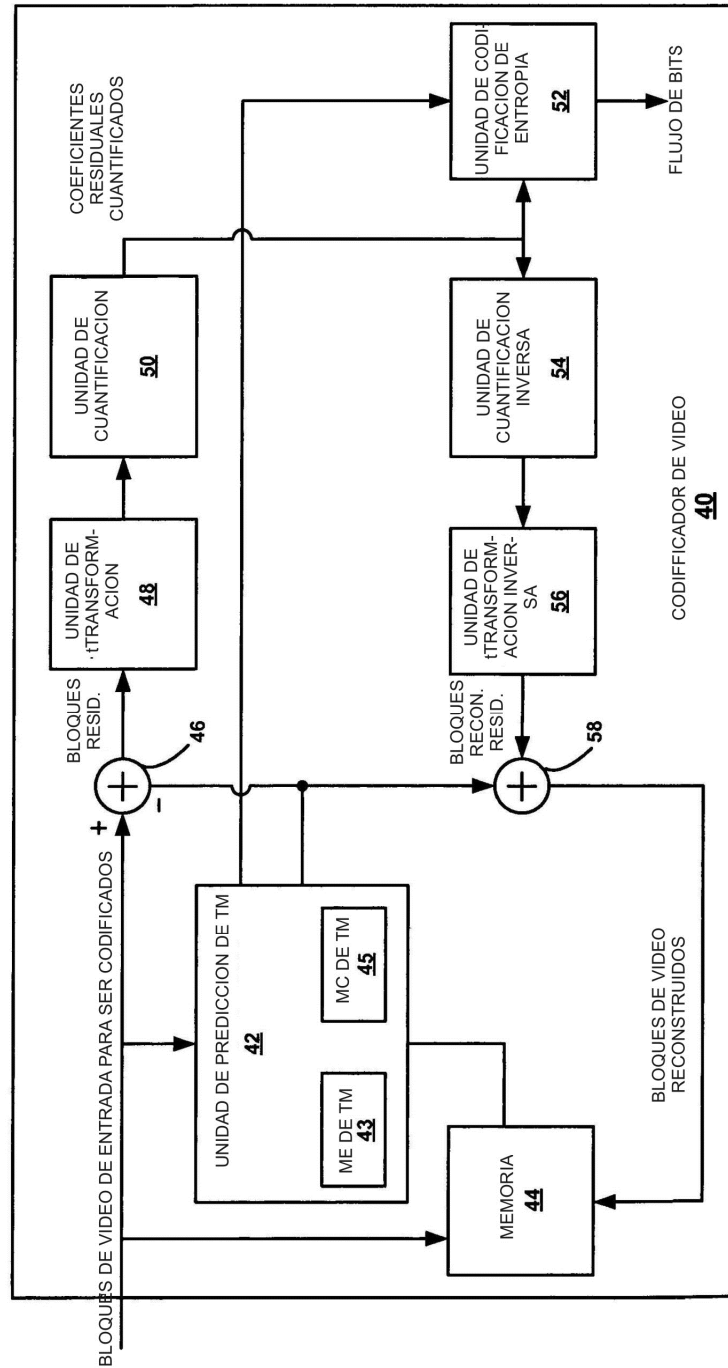


FIG. 2

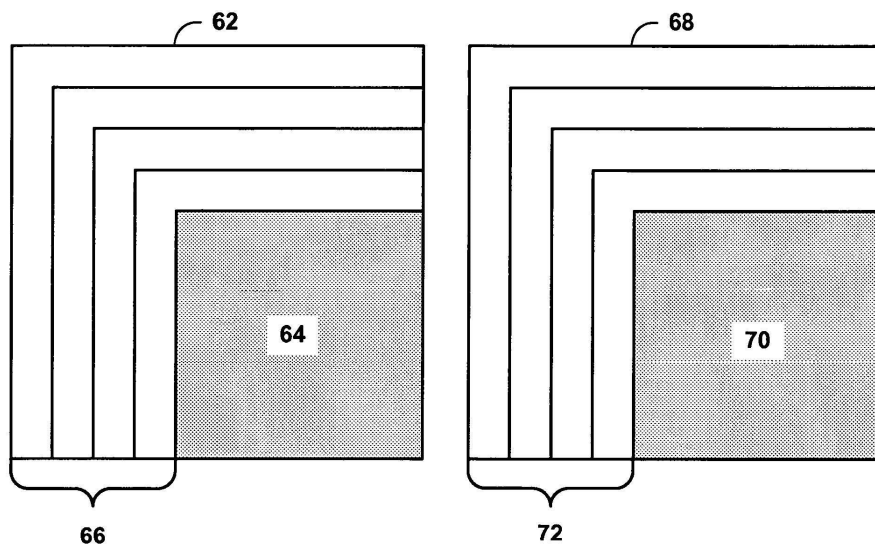


FIG. 3

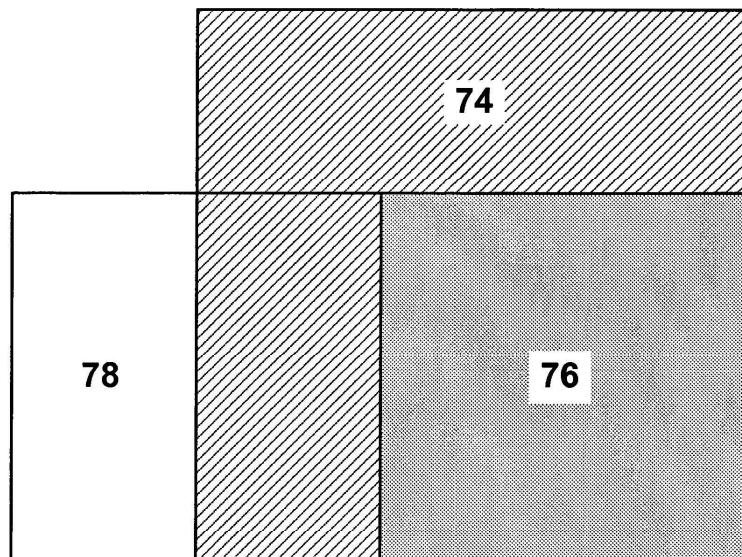


FIG. 4

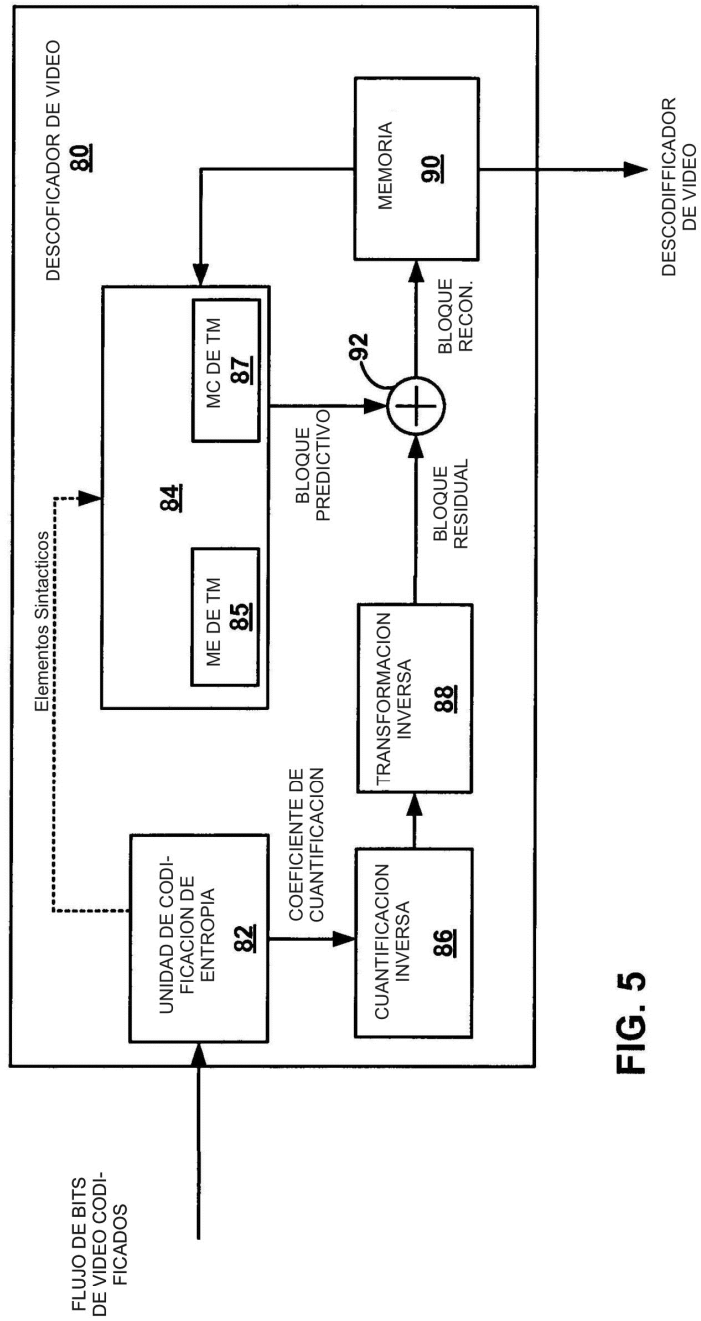


FIG. 5

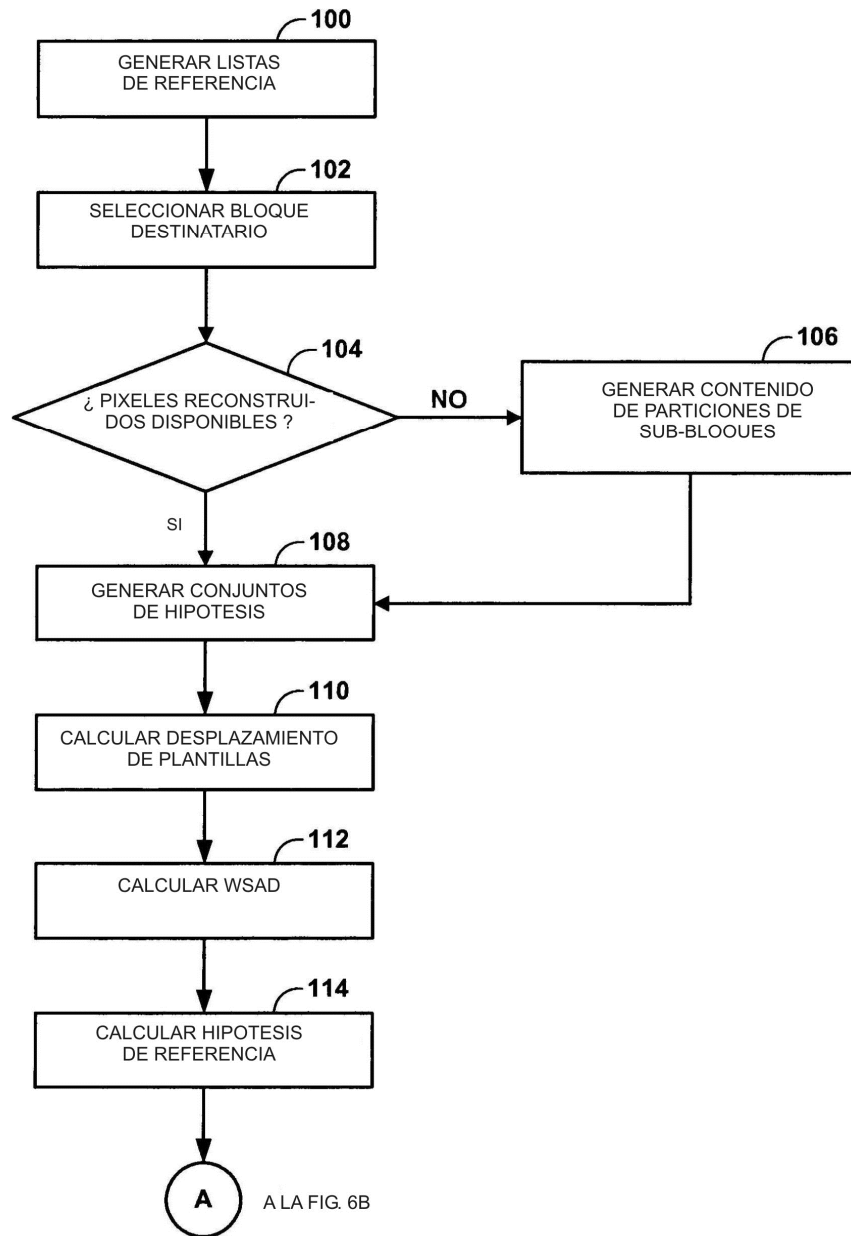


FIG. 6A

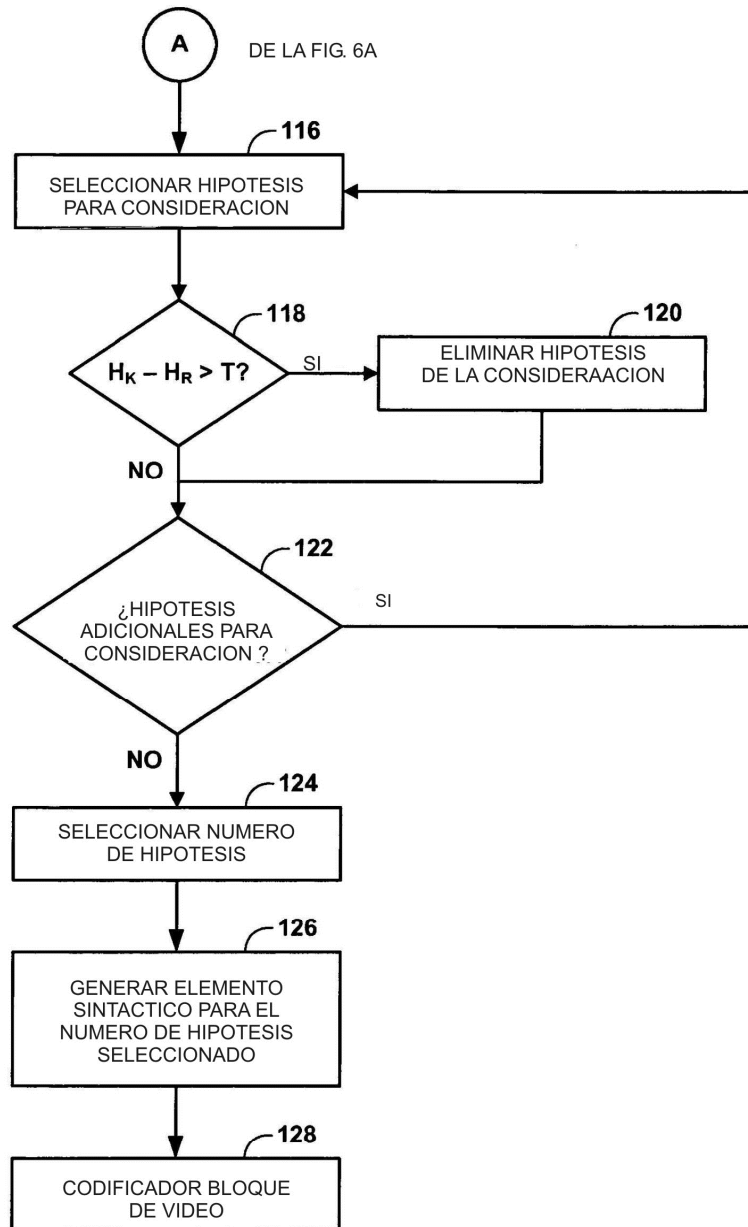
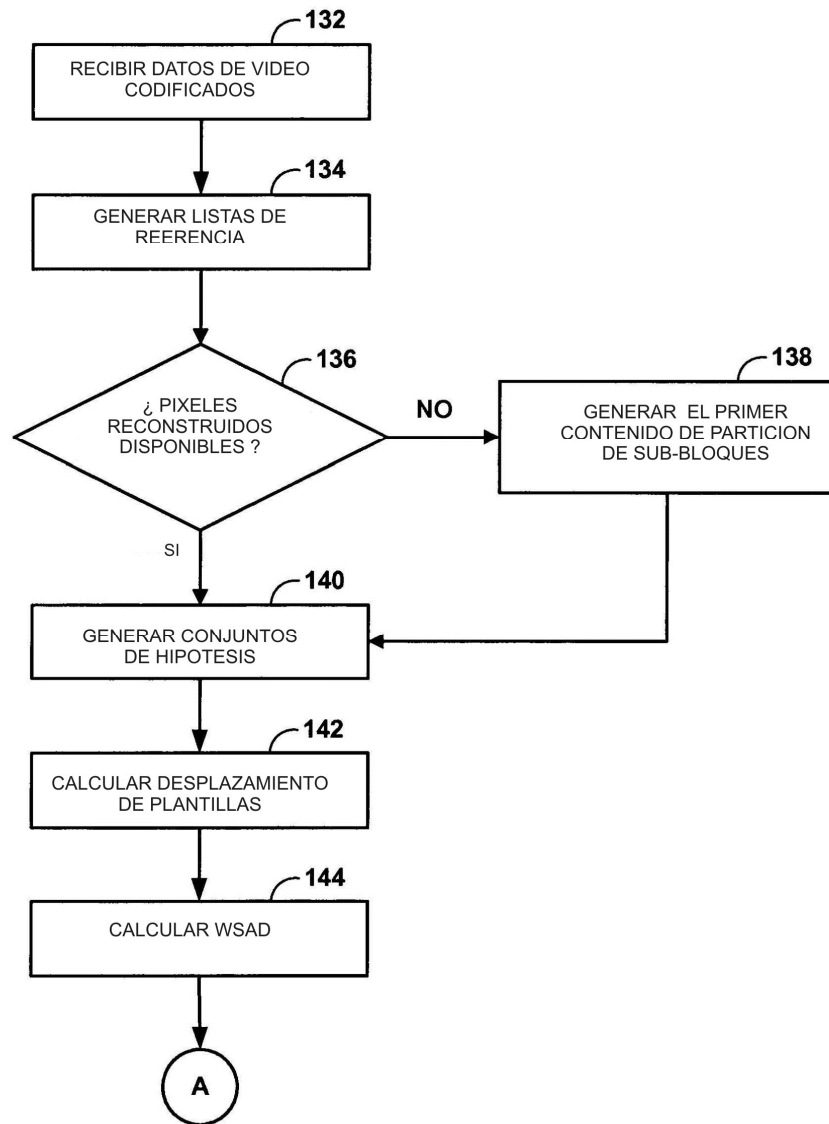


FIG. 6B



A LA FIG. 7B

FIG. 7A

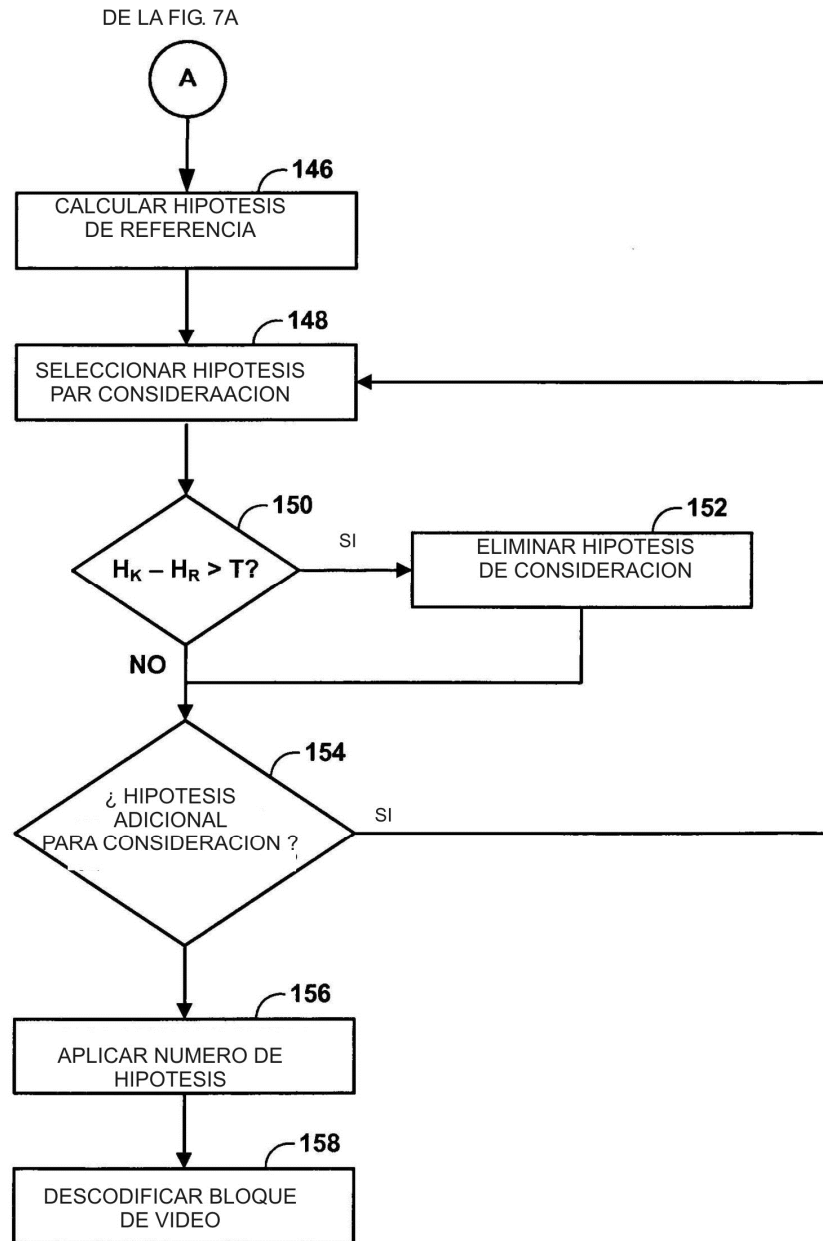


FIG. 7B