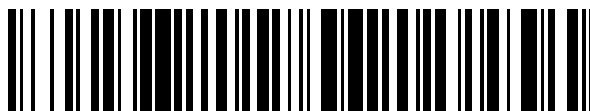


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 527 932**

51 Int. Cl.:

H04N 19/00

(2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.04.2008** **E 08735287 (8)**

97 Fecha y número de publicación de la concesión europea: **01.10.2014** **EP 2279622**

54 Título: **Escalabilidad de profundidad de bits**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
02.02.2015

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**WIEGAND, THOMAS y
WINKEN, MARTIN**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 527 932 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

Escalabilidad de profundidad de bits

5 **[0001]** La presente invención se refiere a la codificación de imágenes y/o vídeo y, en particular, a una codificación de calidad escalable que permite la escalabilidad de profundidad de bits usando flujos de datos de calidad escalable.

10 **[0002]** El Equipo Mixto de Vídeo (JVT, *Joint Video Team*) del Grupo de Expertos de Imágenes en Movimiento (MPEG, *Moving Pictures Experts Group*) de ISO/IEC y el Grupo de Expertos de Codificación de Vídeo (VCEG, *Video Coding Experts Group*) de ITU-T han finalizado recientemente una extensión escalable de la norma de codificación de vídeo de última generación H.264/AVC denominada Codificación de Vídeo Escalable (SVC, *Scalable Video Coding*). SVC soporta la codificación temporal, espacial y escalable en SNR de secuencias de vídeo o cualquier combinación de las mismas.

15 **[0003]** La norma H.264/AVC descrita en el documento "Advanced Video Coding for Generic Audiovisual Services", de ITU-T Rec. & ISO/IEC 14496-10 AVC, versión 3, 2005, especifica un códec de vídeo híbrido en el que señales de predicción de macrobloque se generan en el dominio temporal mediante predicción compensada por movimiento, o en el dominio espacial mediante intra-predicción, donde ambas predicciones van seguidas de una codificación residual. La codificación H.264/AVC sin la extensión de escalabilidad se conoce como codificación H.264/AVC de capa única. Una relación velocidad/distorsión comparable con H.264/AVC de capa única significa que normalmente se consigue la misma calidad de reproducción visual a una velocidad binaria del 10%. A partir de lo anterior, la escalabilidad se considera una funcionalidad para la eliminación de partes del flujo de bits consiguiendo al mismo tiempo una relación velocidad/distorsión en cualquier resolución espacial, temporal o SNR soportada que sea comparable con la codificación H.264/AVC de capa única en esa resolución particular.

20 **[0004]** El diseño básico de la codificación de vídeo escalable (SVC) puede clasificarse como un códec de vídeo estratificado. En cada capa, los conceptos básicos de la predicción compensada por movimiento y la intra-predicción se utilizan como en H.264/AVC. Sin embargo, mecanismos adicionales de predicción entre capas se han integrado con el fin de aprovechar la redundancia entre varias capas espaciales o SNR. La escalabilidad SNR se consigue básicamente mediante cuantificación residual, mientras que para la escalabilidad espacial se utiliza una combinación de predicción compensada por movimiento y una descomposición piramidal sobremuestreada. Se mantiene el enfoque de escalabilidad temporal de H.264/AVC.

30 **[0005]** En general, la estructura del codificador depende del espacio de escalabilidad requerido por una aplicación. A modo de ilustración, la Fig. 8 muestra una estructura de codificador 900 típica con dos capas espaciales 902a, 902b. En cada capa se utiliza una estructura de predicción compensada por movimiento jerárquica e independiente 904a,b con parámetros de movimiento específicos de capa 906a,b. La redundancia entre capas 902a,b consecutivas es aprovechada por los conceptos de predicción entre capas 908 que incluyen mecanismos de predicción para parámetros de movimiento 906a,b, así como datos de textura 910a,b. Una representación base 912a,b de las imágenes de entrada 914a,b de cada capa 902a,b se obtiene mediante codificación de transformada 916a,b similar a la de H.264/AVC, donde las unidades NAL (Capa de Abstracción de Red, *Network Abstraction Layer*) correspondientes contienen información de movimiento y datos de textura; las unidades NAL de la representación base de la capa más baja, es decir, 912a, son compatibles con H.264/AVC de capa única.

40 **[0006]** Los flujos de bits resultantes proporcionados por la codificación de capa base 916a,b y la codificación de textura de refinamiento SNR progresivo 918a,b de las capas respectivas 902a,b, respectivamente, se multiplexan por un multiplexor 920 con el fin de obtener el flujo de bits escalable 922. Este flujo de bits 922 puede escalarse en el tiempo, en el espacio y en calidad SNR.

50 **[0007]** En resumen, según la extensión escalable anterior de la norma de codificación de vídeo H.264/AVC, la escalabilidad temporal se proporciona usando una estructura de predicción jerárquica. Para esta estructura de predicción jerárquica, una de las normas H.264/AVC de capa única puede usarse sin modificaciones. Para la escalabilidad espacial y SNR tienen que añadirse herramientas adicionales a H.264/MPEG4.AVC de capa única como se describe en la extensión SVC de H.264/AVC. Los tres tipos de escalabilidad pueden combinarse para generar un flujo de bits que soporte, en gran medida, una escalabilidad combinada.

60 **[0008]** Surgen problemas cuando una señal fuente de vídeo tiene un intervalo dinámico diferente al requerido por el descodificador o reproductor, respectivamente. En la norma SVC actual mencionada anteriormente, las herramientas de escalabilidad solo se especifican para el caso en que tanto la capa base como la capa de mejora representan una fuente de vídeo dada con la misma profundidad de bits de las disposiciones correspondientes de muestras de luminancia y/o crominancia. Por tanto, considerando diferentes descodificadores y reproductores, respectivamente, que requieren diferentes profundidades de bits, varios flujos de codificación dedicados para cada una de las profundidades de bits tendrían que proporcionarse por separado. Sin embargo, desde el punto de vista de

la relación velocidad/distorsión, esto supone una mayor sobrecarga y una menor eficiencia, respectivamente.

[0009] Se ha propuesto añadir a la norma SVC una escalabilidad en lo que respecta a la profundidad de bits. Por ejemplo, Shan Liu et al. describen en un documento de JVT, concretamente el JVT-X075, la posibilidad de obtener una predicción entre capas a partir de una representación de profundidad de bits inferior de una capa base mediante la utilización de una correlación de tonos inversa según la cual un valor de píxel p' predicho entre capas o correlacionado con tonos de manera inversa se calcula a partir de un valor de píxel de capa base p_b mediante la expresión $p' = p_b \cdot escala + desfase$, donde cabe señalar que la predicción entre capas se lleva a cabo en macrobloques o en tamaños de bloque más pequeños. En el documento JVT-Y067, Shan Liu presenta resultados para este esquema de predicción entre capas. Asimismo, Andrew Segall et al. proponen en el documento JVT-X071 una predicción entre capas para la escalabilidad de profundidad de bits según la cual se usa una operación de ganancia más desfase para la correlación de tonos inversa. Los parámetros de ganancia se indexan y se transmiten en el flujo de bits de capa de mejora bloque a bloque. La señalización de los factores de escala y de los factores de desfase se consigue mediante una combinación de predicción y refinamiento. Además, se describe que una sintaxis de alto nivel soporta granularidades más toscas que la transmisión bloque a bloque. También se hace referencia al documento "Scalable Coding of High Dynamic Range Video", de Andrew Segall en ICIP 2007, I-1 a I-4 y a los documentos JVT-X067 y JVT-W113 de JVT, elaborados asimismo por Andrew Segall.

[0010] Aunque las propuestas mencionadas anteriormente para usar una correlación de tonos inversa con el fin de obtener una predicción a partir de una capa base de profundidad de bits inferior elimina parte de la redundancia entre la información de profundidad de bits inferior y la información de profundidad de bits superior, sería beneficioso obtener una eficacia aún mayor a la hora de proporcionar tal flujo de bits de profundidad de bits escalable, especialmente en lo que respecta a la relación velocidad/distorsión.

[0011] El objeto de la presente invención es proporcionar un esquema de codificación que permita una manera más eficiente de proporcionar una codificación de una imagen o vídeo que sea adecuada para diferentes profundidades de bits.

[0012] Este objeto se consigue mediante un codificador según la reivindicación 1, un descodificador según la reivindicación 11, un procedimiento según la reivindicación 22 ó 23, o un flujo de datos de calidad escalable según la reivindicación 24.

[0013] La presente invención se basa en el hecho de que la eficacia de un flujo de datos de profundidad de bits escalable puede aumentar cuando se obtiene una predicción entre capas correlacionando muestras de la representación de los datos fuente de imagen o vídeo que presentan una primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a una segunda profundidad de bits de muestras de imagen que es mayor que la primera profundidad de bits de muestras de imagen mediante la utilización de una o más funciones de correlación global que son constantes en los datos fuente de imagen o vídeo o que varían en una primera granularidad, y una función de correlación local que modifica localmente la una o más funciones de correlación global y que varía en una segunda granularidad más pequeña que la primera granularidad, formándose el flujo de datos de calidad escalable conforme a la función de correlación local de modo que la función de correlación local puede obtenerse a partir del flujo de datos de calidad escalable. Aunque la provisión de una o más funciones de correlación global, además de una función de correlación local que, a su vez, modifica localmente la una o más funciones de correlación global, aumenta a primera vista la cantidad de información complementaria en el flujo de datos escalable, este aumento está más que justificado por el hecho de que esta subdivisión en una función de correlación global, por un lado, y una función de correlación local, por otro, permite que la función de correlación local y sus parámetros para su parametrización sean pequeños y que, por tanto, puedan codificarse de manera muy eficaz. La función de correlación global puede codificarse en el flujo de datos de calidad escalable y, puesto que es constante en los datos fuente de imagen o vídeo, o varía en una granularidad mayor, la sobrecarga o la flexibilidad para definir esta función de correlación global puede aumentar, de modo que esta función de correlación global puede ajustarse de manera precisa a la estadística media de los datos fuente de imagen o vídeo, reduciendo así adicionalmente la magnitud de la función de correlación local.

[0014] A continuación se describirán realizaciones preferidas de la presente solicitud con referencia a las figuras. En particular,

la Fig. 1 muestra un diagrama de bloques de un codificador de vídeo según una realización de la presente invención;

la Fig. 2 muestra un diagrama de bloques de un descodificador de vídeo según una realización de la presente invención;

la Fig. 3 muestra un diagrama de flujo de una posible implementación de un modo de funcionamiento del módulo de predicción 134 de la Fig. 1 según una realización;

la Fig. 4 muestra una representación esquemática de un vídeo y su subdivisión en secuencias de imágenes, imágenes, pares de macrobloques, macrobloques y bloques de transformación según una realización;

la Fig. 5 muestra una representación esquemática de una parte de una imagen, subdividida en bloques según la fina granularidad subyacente en la función de adaptación/correlación local, e ilustra asimismo un esquema de codificación predictivo para codificar los parámetros de función de adaptación/correlación local según una realización;

la Fig. 6 muestra un diagrama de flujo para ilustrar un proceso de correlación de tonos inversa en el codificador según una realización;

la Fig. 7 muestra un diagrama de flujo de un proceso de correlación de tonos inversa en el decodificador correspondiente al de la Fig. 6, según una realización; y

la Fig. 8 muestra un diagrama de bloques de una estructura de codificador convencional para la codificación de vídeo escalable.

[0015] La Fig. 1 muestra un codificador 100 que comprende un medio de codificación base 102, un medio de predicción 104, un medio de codificación residual 106 y un medio de combinación 108, así como una entrada 110 y una salida 112. El codificador 100 de la Fig. 1 es un codificador de vídeo que recibe una señal de vídeo de alta calidad en la entrada 110 y que proporciona un flujo de bits de calidad escalable en la salida 112. El medio de codificación base 102 codifica los datos de la entrada 110 en un flujo de datos de codificación base que representa el contenido de esta señal de vídeo en la entrada 110 con una profundidad de bits de muestras de imagen reducida y, opcionalmente, una menor resolución espacial en comparación con la señal de entrada en la entrada 110. El medio de predicción 104 está adaptado para, en función del flujo de datos de codificación base proporcionado por el medio de codificación base 102, proporcionar una señal de predicción con una profundidad de bits de muestras de imagen total o aumentada y, opcionalmente, una resolución espacial total o aumentada para la señal de vídeo en la entrada 110. Un restador 114 también comprendido en el codificador 100 forma un residuo de predicción de la señal de predicción proporcionada por el medio 104 con respecto a la señal de entrada de alta calidad en la entrada 110, codificándose la señal residual mediante el medio de codificación residual 106 en un flujo de datos de capa de mejora de calidad. El medio de combinación 108 combina el flujo de datos de codificación base procedente del medio de codificación base 102 y el flujo de datos de capa de mejora de calidad proporcionado por el medio de codificación residual 106 para formar un flujo de datos de calidad escalable 112 en la salida 112. Calidad escalable significa que el flujo de datos en la salida 112 está compuesto, por un lado, por una parte que está autocontenida ya que permite una reconstrucción de la señal de vídeo 110 con la profundidad de bits reducida y, opcionalmente, la resolución espacial reducida sin ninguna información adicional y despreciando el resto del flujo de datos 112, y por una parte adicional que permite, en combinación con la primera parte, una reconstrucción de la señal de vídeo en la entrada 110 en la profundidad de bits original y siendo la resolución espacial original mayor que la profundidad de bits y/o la resolución espacial de la primera parte.

[0016] Tras haber descrito en términos generales la estructura y la funcionalidad del codificador 100, a continuación se describe en mayor detalle su estructura interna. En particular, el medio de codificación base 102 comprende un módulo de conversión descendente 116, un restador 118, un módulo de transformación 120 y un módulo de cuantificación 122 conectados en serie, en el orden mencionado, entre la entrada 110 y el medio de combinación 108 y el medio de predicción 104, respectivamente. El módulo de conversión descendente 116 es para reducir la profundidad de bits de las muestras de imagen y, opcionalmente, la resolución espacial de las imágenes de la señal de vídeo en la entrada 110. Dicho de otro modo, el módulo de conversión descendente 116 convierte de manera descendente e irreversible la señal de vídeo de entrada de alta calidad de la entrada 110 en una señal de vídeo de calidad base. Como se describirá posteriormente en mayor detalle, esta conversión descendente puede incluir reducir la profundidad de bits de las muestras de señal, es decir, los valores de píxel, en la señal de vídeo en la entrada 110 usando cualquier esquema de correlación de tonos, tal como redondear los valores de muestra, submuestrear las componentes de crominancia en caso de que la señal de vídeo se proporcione en forma de componentes de luminancia y crominancia, filtrar la señal de entrada de la entrada 110, por ejemplo mediante conversión de RGB a YCbCr, o cualquier combinación de los mismos. Posteriormente se ofrecerán más detalles de los posibles mecanismos de predicción. En particular, es posible que el módulo de conversión descendente 116 use diferentes esquemas de conversión descendente para cada imagen de la entrada de señal de vídeo o de secuencias de imágenes de la entrada 110 o use el mismo esquema para todas las imágenes. Esto también se describe posteriormente en mayor detalle.

[0017] El restador 118, el módulo de transformación 120 y el módulo de cuantificación 122 actúan conjuntamente para codificar la señal de calidad base proporcionada por el módulo de conversión descendente 116 usando, por ejemplo, un esquema de codificación de vídeo no escalable, tal como H.264/AVC. Según el ejemplo de la Fig. 1, el restador 118, el módulo de transformación 120 y el módulo de cuantificación 122 actúan conjuntamente con un filtro de bucle de predicción opcional 124, un módulo de predicción 126, un módulo de transformación inversa

128 y un sumador 130 comprendidos conjuntamente por el medio de codificación base 102 y el medio de predicción 104 para formar la parte de reducción de irrelevancia de un codificador híbrido que codifica la señal de vídeo de calidad base proporcionada por el módulo de conversión descendente 116 mediante una predicción basada en compensación de movimiento y después de la compresión del residuo de predicción. En particular, el restador 118 extrae de una imagen o macrobloque actual de la señal de vídeo de calidad base una imagen predicha o una parte de macrobloque predicha reconstruida a partir de imágenes previamente codificadas de la señal de vídeo de calidad base usando, por ejemplo, compensación de movimiento. El módulo de transformación 120 aplica una transformada al residuo de predicción, tal como una transformada DCT, FFT o de ondículas. La señal residual transformada puede representar una representación espectral y sus coeficientes de transformada se cuantifican de manera irreversible en el módulo de cuantificación 122. La señal residual cuantificada resultante representa la parte residual del flujo de datos de codificación base proporcionado por el medio de codificación base 102.

[0018] Aparte del filtro de bucle de predicción opcional 124, el módulo de predicción 126, el módulo de transformación inversa 128 y el sumador 130, el medio de predicción 104 comprende un filtro opcional para reducir los artefactos de codificación 132 y un módulo de predicción 134. El módulo de transformación inversa 128, el sumador 130, el filtro de bucle de predicción opcional 124 y el módulo de predicción 126 actúan conjuntamente para reconstruir la señal de vídeo con una profundidad de bits reducida y, opcionalmente, una resolución espacial reducida, como se define mediante el módulo de conversión descendente 116. Dicho de otro modo, crean una señal de vídeo de baja profundidad de bits y, opcionalmente, de baja resolución espacial para el filtro opcional 132, que representa una representación de baja calidad de la señal fuente de la entrada 110 que también puede reconstruirse en el lado del descodificador. En particular, el módulo de transformación inversa 128 y el sumador 130 están conectados en serie entre el módulo de cuantificación 122 y el filtro opcional 132, mientras que el filtro de bucle de predicción opcional 124 y el módulo de predicción 126 están conectados en serie, en el orden mencionado, entre una salida del sumador 130 y una entrada adicional del sumador 130. La salida del módulo de predicción 126 también está conectada a una entrada de inversión del restador 118. El filtro opcional 132 está conectado entre la salida del sumador 130 y el módulo de predicción 134 que, a su vez, está conectado entre la salida del filtro opcional 132 y la entrada de inversión del restador 114.

[0019] El módulo de transformación inversa 128 transforma de manera inversa las imágenes residuales de codificación base proporcionadas por el medio de codificación base 102 para conseguir imágenes residuales de baja profundidad de bits y, opcionalmente, de baja resolución espacial. Por consiguiente, el módulo de transformación inversa 128 lleva a cabo una transformación inversa que es una inversión de la transformación y la cuantificación llevadas a cabo por los módulos 120 y 122. Como alternativa, un módulo de descuantificación puede proporcionarse por separado en el lado de entrada del módulo de transformación inversa 128. El sumador 130 añade una predicción a las imágenes residuales reconstruidas, estando basada la predicción en imágenes previamente reconstruidas de la señal de vídeo. En particular, el sumador 130 proporciona una señal de vídeo reconstruida con una profundidad de bits reducida y, opcionalmente, una resolución espacial reducida. Estas imágenes reconstruidas son filtradas por el filtro de bucle 124 para reducir los artefactos, por ejemplo, y son usadas posteriormente por el módulo de predicción 126 para predecir la imagen que va a reconstruirse en este momento mediante, por ejemplo, compensación de movimiento, a partir de imágenes reconstruidas anteriormente. La señal de calidad base así obtenida en la salida del sumador 130 es usada por la conexión serie del filtro opcional 132 y el módulo de predicción 134 para obtener una predicción de la señal de entrada de alta calidad en la entrada 110, usándose esta última predicción para formar la señal de mejora de alta calidad en la salida del medio de codificación residual 106. Esto se describe posteriormente en mayor detalle.

[0020] En particular, la señal de baja calidad obtenida del sumador 130 es filtrada opcionalmente por el filtro opcional 132 para reducir los artefactos de codificación. Los filtros 124 y 132 pueden hacerse funcionar de la misma manera y, por tanto, aunque los filtros 124 y 132 se muestran por separado en la Fig. 1, pueden ser sustituidos por un solo filtro dispuesto entre la salida del sumador 130 y la entrada del módulo de predicción 126 y el módulo de predicción 134, respectivamente. Después, la señal de vídeo de baja calidad es usada por el módulo de predicción 134 para formar una señal de predicción para la señal de vídeo de alta calidad recibida en la entrada de no inversión del sumador 114 que está conectada a la entrada 110. Este proceso de formación de la predicción de alta calidad puede incluir correlacionar las muestras de imagen de señal de calidad base descodificadas mediante la utilización de una función de correlación combinada como se describe posteriormente en mayor detalle, usar el valor respectivo de las muestras de señal de calidad base para indexar una tabla de consulta que contiene los valores de muestra de alta calidad correspondientes, usar el valor de la muestra de señal de calidad base para un proceso de interpolación para obtener el valor de muestra de alta calidad correspondiente, muestrear de manera ascendente las componentes de crominancia, filtrar la señal de calidad base usando, por ejemplo conversión de YCbCr a RGB, o cualquier combinación de lo anterior. Posteriormente se describirán otros ejemplos.

[0021] Por ejemplo, el módulo de predicción 134 puede correlacionar las muestras de la señal de vídeo de calidad base desde un primer intervalo dinámico hasta un segundo intervalo dinámico, mayor que el primer intervalo dinámico, y, opcionalmente, mediante la utilización de un filtro de interpolación espacial, interpolar espacialmente muestras de la señal de vídeo de calidad base para aumentar la resolución espacial para que corresponda a la

resolución espacial de la señal de vídeo en la entrada 110. De manera similar a la descripción anterior del módulo de conversión descendente 116, es posible usar un proceso de predicción diferente para diferentes imágenes de la secuencia de señales de vídeo de calidad base, así como usar el mismo proceso de predicción en todas las imágenes.

[0022] El restador 114 extrae la predicción de alta calidad recibida desde el módulo de predicción 134 de la señal de vídeo de alta calidad recibida desde la entrada 110 para proporcionar una señal residual de predicción de alta calidad, es decir, con la profundidad de bits y, opcionalmente, la resolución espacial originales, al medio de codificación residual 106. En el medio de codificación residual 106, la diferencia entre la señal de entrada de alta calidad original y la predicción obtenida a partir de la señal de calidad base descodificada se codifica usando, por ejemplo, un esquema de codificación por compresión tal como, por ejemplo, el especificado en la norma H.264/AVC. Para ello, el medio de codificación residual 106 de la Fig. 1 comprende, a modo de ejemplo, un módulo de transformación 136, un módulo de cuantificación 138 y un módulo de codificación por entropía 140 conectados en serie entre una salida del restador 114 y el medio de combinación 108 en el orden mencionado. El módulo de transformación 136 transforma la señal residual o las imágenes de la misma, respectivamente, en un dominio de transformación o dominio espectral, respectivamente, donde las componentes espectrales son cuantificadas por el módulo de cuantificación 138 y los valores de transformada cuantificados son codificados por entropía mediante el módulo de codificación por entropía 140. El resultado de la codificación por entropía representa el flujo de datos de capa de mejora de alta calidad proporcionado por el medio de codificación residual 106. Si los módulos 136 a 140 implementan una codificación H.264/AVC, que soporta transformadas con un tamaño de muestras 4x4 u 8x8 para codificar el contenido de luminancia, el tamaño de transformada para transformar la componente de luminancia de la señal residual del restador 114 en el módulo de transformación 136 puede elegirse de manera arbitraria para cada macrobloque y no tiene que ser necesariamente el mismo que el utilizado para codificar la señal de calidad base en el módulo de transformación 120. Para codificar las componentes de crominancia, la norma H.264/AVC no ofrece elección. Cuando se cuantifican los coeficientes de transformada en el módulo de cuantificación 138, puede usarse el mismo esquema de cuantificación que en la norma H.264/AVC, lo que significa que el valor de incremento de cuantificador puede controlarse mediante un parámetro de cuantificación QP que puede tomar valores entre -6* (profundidad de bits de componente de señal de vídeo de alta calidad -8) y 51. El QP usado para codificar el macrobloque de representación de calidad base en el módulo de cuantificación 122 y el QP usado para codificar el macrobloque de mejora de alta calidad en el módulo de cuantificación 138 no tiene por qué ser el mismo.

[0023] El medio de combinación 108 comprende un módulo de codificación por entropía 142 y el multiplexor 144. El módulo de codificación por entropía 142 está conectado entre una salida del módulo de cuantificación 122 y una primera entrada del multiplexor 144, mientras que una segunda entrada del multiplexor 144 está conectada a una salida del módulo de codificación por entropía 140. La salida del multiplexor 144 representa la salida 112 del codificador 100.

[0024] El módulo de codificación por entropía 142 codifica mediante entropía los valores de transformada cuantificados proporcionados por el módulo de cuantificación 122 para formar un flujo de datos de capa de calidad base a partir del flujo de datos de codificación base proporcionado por el módulo de cuantificación 122. Por lo tanto, como se ha mencionado anteriormente, los módulos 118, 120, 122, 124, 126, 128, 130 y 142 pueden diseñarse para que actúen conjuntamente según la norma H.264/AVC y para que representen conjuntamente un codificador híbrido, donde el codificador de entropía 142 lleva a cabo una compresión sin pérdidas del residuo de predicción cuantificado.

[0025] El multiplexor 144 recibe el flujo de datos de capa de calidad base y el flujo de datos de capa de alta calidad y los fusiona para formar el flujo de datos de calidad escalable.

[0026] Como ya se ha descrito anteriormente y como se muestra en la Fig. 3, la manera en la que el módulo de predicción 134 lleva a cabo la predicción desde la señal de calidad base reconstruida hasta el dominio de señales de alta calidad puede incluir una expansión de la profundidad de bits de muestra, también denominada correlación de tonos inversa 150, y, opcionalmente, una operación de muestreo ascendente espacial, es decir, una operación de filtrado de muestreo ascendente 152 en caso de que la señal base y la señal de alta calidad tengan una resolución espacial diferente. El orden en que el módulo de predicción 134 lleva a cabo la correlación de tonos inversa 150 y la operación de muestreo ascendente espacial opcional 152 puede ser fijo y, por consiguiente, ser ya conocido por el codificador y el descodificador, o puede elegirse de manera adaptativa bloque a bloque o imagen a imagen, o con alguna otra granularidad, en cuyo caso el módulo de predicción 134 señaliza información en el orden entre las etapas 150 y 152 usadas por alguna entidad, tal como el codificador 106, para introducirse como información complementaria en el flujo de bits 112 para señalizarse en el lado del descodificador como parte de la información complementaria. La adaptación del orden entre las etapas 150 y 152 se ilustra en la Fig. 3 mediante una flecha discontinua de doble cabeza 154, y la granularidad en la que el orden puede elegirse de manera adaptativa también puede señalizarse e incluso variar en el vídeo.

[0027] Para llevar a cabo la correlación de tonos inversa 150, el módulo de predicción 134 usa dos partes,

concretamente una o más funciones de correlación de tonos inversa global y una adaptación local de las mismas. Generalmente, la una o más funciones de correlación de tonos inversa global tienen en cuenta las características generales y habituales de la secuencia de imágenes del vídeo y, por consiguiente, de la correlación de tonos que se ha aplicado inicialmente a la señal de vídeo de entrada de alta calidad para obtener la señal de vídeo de calidad base en el módulo de conversión descendente 116. En comparación, la adaptación local tendrá en cuenta las desviaciones individuales con respecto al modelo de correlación de tonos inversa global para los bloques individuales de las imágenes del vídeo.

[0028] Para ilustrar esto, la Fig. 4 muestra una parte de un vídeo 160 que consiste, a modo de ejemplo, en cuatro imágenes consecutivas 162a a 162d del vídeo 160. Dicho de otro modo, el vídeo 160 comprende una secuencia de imágenes 162, de las cuales se muestran cuatro a modo de ejemplo en la Fig. 4. El vídeo 160 puede dividirse en secuencias no solapadas de imágenes consecutivas para las que se transmiten parámetros globales o elementos de sintaxis en el flujo de datos 112. Solamente con fines ilustrativos se supone que las cuatro imágenes consecutivas 162a a 162d mostradas en la Fig. 4 formarán tal secuencia 164 de imágenes. Cada imagen, a su vez, se subdivide en una pluralidad de macrobloques 166 como los ilustrados en la esquina inferior izquierda de la imagen 162d. Un macrobloque es un contenedor dentro del cual los coeficientes de transformación y otros elementos de sintaxis de control que pertenecen al tipo de codificación del macrobloque se transmiten dentro del flujo de bits 112. Un par 168 de macrobloques 166 cubre una parte continua de la imagen respectiva 162d. Dependiendo de un modo de par de macrobloques del par de macrobloques 168 respectivo, el macrobloque superior 162 de este par 168 cubre las muestras de la mitad superior del par de macrobloques 168 o las muestras de cada línea impar del par de macrobloques 168, donde el macrobloque inferior se refiere a las otras muestras en el mismo, respectivamente. A su vez, cada macrobloque 160 puede subdividirse en bloques de transformación como los ilustrados en 170, donde estos bloques de transformación forman la base de bloque en la que el módulo de transformación 120 lleva a cabo la transformación y el módulo de transformación inversa 128 lleva a cabo la transformación inversa.

[0029] Haciendo de nuevo referencia a la función de correlación de tonos inversa global que acaba de mencionarse, el módulo de predicción 134 puede configurarse para usar constantemente una o más de tales funciones de correlación de tonos inversa global para todo el vídeo 160 o, como alternativa, para una subparte del mismo, tal como la secuencia 164 de imágenes consecutivas o una imagen 162 propiamente dicha. Las últimas opciones implican que el módulo de predicción 134 varía la función de correlación de tonos inversa global en una granularidad correspondiente a un tamaño de secuencia de imágenes o un tamaño de imagen. Posteriormente se ofrecerán ejemplos de funciones de correlación de tonos inversa global. Si el módulo de predicción 134 adapta la función de correlación de tonos inversa global a las estadísticas del vídeo 160, el módulo de predicción 134 proporciona información al módulo de codificación por entropía 140 o al multiplexor 144, de modo que el flujo de bits 112 contiene información acerca de la(s) función(es) de correlación de tonos inversa global y de la variación de las mismas en el vídeo 160. En caso de que la una o más funciones de correlación de tonos inversa global usadas por el módulo de predicción 134 se apliquen continuamente a todo el vídeo 160, pueden ser conocidas de antemano por el descodificador o pueden transmitirse como información secundaria en el flujo de bits 112.

[0030] En una granularidad incluso más pequeña, una función de correlación de tonos inversa local usada por el módulo de predicción 134 varía en el vídeo 160. Por ejemplo, tal función de correlación de tonos inversa local varía en una granularidad más pequeña que un tamaño de imagen tal como, por ejemplo, el tamaño de un macrobloque, un par de macrobloques o un tamaño de bloque de transformación.

[0031] Tanto para la(s) función(es) de correlación de tonos inversa global como para la función de correlación de tonos inversa local, la granularidad en la que la función respectiva varía o en la que las funciones se definen en el flujo de bits 112 puede variar en el vídeo 160. A su vez, la variación de la granularidad puede señalizarse en el flujo de bits 112.

[0032] Durante la correlación de tonos inversa 150, el módulo de predicción 134 correlaciona una muestra predeterminada del vídeo 160 desde la profundidad de bits de calidad base hasta la profundidad de bits de alta calidad usando una combinación de una función de correlación de tonos inversa global que se aplica a la imagen respectiva y la función de correlación de tonos inversa local definida en el bloque respectivo al que pertenece la muestra predeterminada.

[0033] Por ejemplo, la combinación puede ser una combinación aritmética y, en particular, una suma. El módulo de predicción puede configurarse para obtener un valor de muestra predicho de alta profundidad de bits S_{alto} a partir del valor de muestra reconstruido correspondiente de baja profundidad de bits S_{bajo} mediante la expresión $S_{\text{alto}} = f_k(S_{\text{bajo}}) + m \cdot S_{\text{bajo}} + n$.

[0034] En esta fórmula, la función f_k representa un operador de correlación de tonos inversa global en el que el índice k selecciona qué operador de correlación de tonos inversa global se selecciona en caso de que se use más de un solo esquema o más de una función de correlación de tonos inversa global. La parte restante de esta fórmula

constituye la adaptación local o la función de correlación de tonos inversa local, siendo n un valor de desfase y siendo m un factor de escalado. Los valores de k , m y n pueden especificarse para cada bloque en el flujo de bits 112. Dicho de otro modo, el flujo de bits 112 permitiría revelar las tríadas $\{k, m, n\}$ para todos los bloques del vídeo 160, donde el tamaño de bloque de estos bloques depende de la granularidad de la adaptación local de la función de correlación de tonos inversa global con esta granularidad que, a su vez, varía posiblemente en el vídeo 160.

[0035] Los siguientes mecanismos de correlación pueden usarse para el proceso de predicción en lo que respecta a la función de correlación de tiempo inversa global $f(x)$. Por ejemplo, una correlación lineal orientada a fragmentos puede usarse si puede especificarse un número arbitrario de puntos de interpolación. Por ejemplo, para una muestra de calidad base con valor x y dos puntos de interpolación dados (x_n, y_n) y (x_{n+1}, y_{n+1}) , la muestra de predicción y correspondiente se obtiene por el módulo 134 según la fórmula siguiente:

$$f(x) = y_n + \frac{x - x_n}{x_{n+1} - x_n} (y_{n+1} - y_n)$$

[0036] Esta interpolación lineal puede llevarse a cabo con poca complejidad computacional usando solamente desplazamiento de bits en lugar de operaciones de división si se limita que $x_{n+1} - x_n$ sea una potencia de dos.

[0037] Otro posible mecanismo de correlación global representa una correlación con una tabla de consulta en la que, mediante los valores de muestra de calidad base, se consulta una tabla de consulta en la que para cada posible valor de muestra de calidad base, en lo que respecta a la función de correlación de tonos inversa global, se especifica el valor de muestra de predicción global correspondiente (x). La tabla de consulta puede proporcionarse al decodificador como información complementaria o puede ser conocida por defecto por el decodificador.

[0038] Además, puede usarse un escalado con un desfase constante en la correlación global. Según esta alternativa, para conseguir la muestra de predicción global de alta calidad correspondiente (x) con una mayor profundidad de bits, el módulo 134 multiplica las muestras de calidad base x por un factor constante 2^{M-N-K} , y después se añade un desfase constante $2^{M-1} - 2^{M-1-K}$, según, por ejemplo, una de las siguientes fórmulas:

$$f(x) = 2^{M-N-K} x + 2^{M-1} - 2^{M-1-K}$$

$$f(x) = \min(2^{M-N-K} x + 2^{M-1} - 2^{M-1-K}, 2^M - 1),$$

respectivamente,

donde M es la profundidad de bits de la señal de alta calidad y N es la profundidad de bits de la señal de calidad base.

[0039] De este modo, el intervalo dinámico de baja calidad $[0; 2^N - 1]$ se correlaciona con el segundo intervalo dinámico $[0; 2^M - 1]$ de una manera según la cual los valores correlacionados de x se distribuyen de manera centralizada con respecto al posible intervalo dinámico $[0; 2^M - 1]$ de mayor calidad en una extensión determinada por K . El valor de K puede ser un valor entero o un valor real, y puede transmitirse como información complementaria al decodificador en, por ejemplo, el flujo de datos de valores escalable, de modo que en el decodificador algunos medios de predicción pueden actuar de la misma manera que el módulo de predicción 134, como se describirá posteriormente. Puede usarse una operación de redondeo para obtener valores enteros $f(x)$.

[0040] Otra posibilidad para el escalado global es escalar con un desfase variable: las muestras de calidad base x se multiplican por un factor constante y después se añade un desfase variable según, por ejemplo, una de las siguientes fórmulas:

$$f(x) = 2^{M-N-K} x + D$$

$$f(x) = \min(2^{M-N-K} x + D, 2^M - 1)$$

[0041] De este modo, el intervalo dinámico de baja calidad se correlaciona de manera global con el segundo intervalo dinámico de una manera según la cual los valores correlacionados de x se distribuyen en una parte del posible intervalo dinámico de las muestras de alta calidad, cuya extensión es determinada por K , y cuyo desfase con respecto al límite inferior es determinado por D . D puede ser un valor entero o real. El resultado $f(x)$ representa un valor de muestra de imagen correlacionado de manera global de la señal de predicción de alta profundidad de bits. Los valores de K y D pueden transmitirse como información complementaria al decodificador en, por ejemplo, el flujo de datos de calidad escalable. De nuevo, puede usarse una operación de redondeo para obtener valores enteros $f(x)$, lo que puede aplicarse asimismo al resto de ejemplos descritos en la presente solicitud para las correlaciones globales de profundidad de bits sin tener que indicarlo continuamente de manera explícita.

[0042] Otra posibilidad para la correlación global es escalar con superposición: las muestras de predicción de alta profundidad de bits correlacionadas de manera global $f(x)$ se obtienen a partir de la muestra de calidad base respectiva x según, por ejemplo, una de las siguientes fórmulas, donde 'redondeo_defecto(a)' redondea al entero inferior más próximo:

$$f(x) = \text{redondeo_defecto}(2^{M-N}x + 2^{M-2N}x)$$

o

$$f(x) = \text{redondeo_defecto}(2^{M-N}x + 2^{M-2N}x), 2^M - 1)$$

[0043] Las posibilidades que acaban de mencionarse pueden combinarse. Por ejemplo, puede usarse un escalado global con superposición y desfase constante: las muestras de predicción de alta profundidad de bits correlacionadas de manera global $f(x)$ se obtienen según, por ejemplo, una de las siguientes fórmulas, donde 'redondeo_defecto(a)' redondea al entero inferior más próximo:

$$f(x) = \text{redondeo_defecto}(2^{M-N-K}x + 2^{M-2N-K}x + 2^{M-1} - 2^{M-1-K})$$

$$f(x) = \min(\text{redondeo_defecto}(2^{M-N-K}x + 2^{M-2N-K}x + 2^{M-1} - 2^{M-1-K}), 2^M - 1)$$

[0044] El valor de K puede especificarse como información complementaria al decodificador.

[0045] Asimismo, puede usarse un escalado global con superposición y desfase variable: las muestras de predicción de alta profundidad de bits correlacionadas de manera global (x) se obtienen según la siguiente fórmula, donde 'redondeo_defecto(a)' redondea al entero inferior más próximo:

$$f(x) = \text{redondeo_defecto}(2^{M-N-K}x + 2^{M-2N-K}x + D)$$

$$f(x) = \min(\text{redondeo_defecto}(2^{M-N-K}x + 2^{M-2N-K}x + D), 2^M - 1)$$

[0046] Los valores de D y K pueden especificarse como información complementaria al decodificador.

[0047] Transfiriendo los ejemplos que acaban de mencionarse para la función de correlación de tonos inversa global a la Fig. 4, los parámetros allí mencionados para definir la función de correlación de tonos inversa global, en concreto (x_1, y_1) , K y D , pueden ser conocidos por el decodificador, pueden transmitirse en el flujo de bits 112 con respecto a todo el vídeo 160 en caso de que la función de correlación de tonos inversa global sea constante en el vídeo 160 o estos parámetros se transmiten en el flujo de bits para diferentes partes del mismo tal como, por ejemplo, para imágenes 162 o secuencias de imágenes 164, dependiendo de la tosca granularidad subyacente a la función de correlación global. En caso de que el módulo de predicción 134 use más de una función de correlación de tonos inversa global, puede considerarse que los parámetros (x_1, y_1) , K y D antes mencionados se proporcionan con un índice k , donde estos parámetros (x_1, y_1) , K_k y D_k definen la k -ésima función de correlación de tonos inversa global f_k .

[0048] Por tanto, sería posible especificar la utilización de diferentes mecanismos de correlación de tonos inversa global para cada bloque de cada imagen señalizando un valor k correspondiente en el flujo de bits 112 así como, de manera alternativa, usar el mismo mecanismo para la secuencia completa del vídeo 160.

[0049] Además, es posible especificar diferentes mecanismos de correlación global para las componentes de luminancia y crominancia de la señal de calidad base para tener en cuenta que las estadísticas, tales como su función de densidad de probabilidad, pueden ser diferentes.

[0050] Como se ha mencionado anteriormente, el módulo de predicción 134 usa una combinación de la función de correlación de tonos inversa global y una función de correlación de tonos inversa local para adaptar de manera local la función global. Dicho de otro modo, se lleva a cabo una adaptación local en cada muestra de predicción de alta profundidad de bits correlacionada de manera global $f(x)$ así obtenida. La adaptación local de la función de correlación de tonos inversa global se lleva a cabo usando una función de correlación de tonos inversa local que se adapta de manera local, por ejemplo, adaptando de manera local algunos parámetros de la misma. En el ejemplo presentado anteriormente, estos parámetros son el factor de escalado m y el valor de desfase n . El factor de escalado m y el valor de desfase n pueden especificarse para cada bloque, donde un bloque puede corresponder a un tamaño de bloque de transformación, lo que es el caso, por ejemplo, de las muestras 4×4 u 8×8 de H.264/AVC, o a un tamaño de macrobloque, lo que es el caso, por ejemplo, de las muestras 16×16 de H.264/AVC. Un bloque puede ser fijo, y por tanto ser conocido de antemano por el codificador y el decodificador, o puede elegirse de manera adaptativa por imagen o por secuencia por el módulo de predicción 134, en cuyo caso tiene que señalizarse al decodificador como parte de información complementaria en el flujo de bits 112. En el caso de H.264/AVC, el conjunto de parámetros de secuencia y/o el conjunto de parámetros de imagen podrían usarse para este fin. Además, en la información complementaria puede especificarse que el factor de escalado m o el valor de desfase n , o ambos valores, sean igual a cero para una secuencia de vídeo completa o para un conjunto bien definido de imágenes de la secuencia de vídeo 160. En caso de que el factor de escalado m o el valor de desfase n , o ambos, se especifiquen para cada bloque para una imagen dada, con el fin de reducir la velocidad binaria requerida para codificar estos valores solo pueden codificarse los valores de diferencia Δm , Δn con respecto a valores predichos m_{pred} , n_{pred} correspondientes, de modo que los valores reales de m , n pueden obtenerse de la siguiente manera: $m = m_{pred} + \Delta m$, $n = n_{pred} + \Delta n$.

[0051] Dicho de otro modo y como se ilustra en la Fig. 5, que muestra una parte de una imagen 162 dividida en bloques 170 que forman la base de la granularidad de la adaptación local de la función de correlación de tonos inversa global, el módulo de predicción 134, el módulo de codificación por entropía 140 y el multiplexor 144 están configurados de modo que los parámetros para adaptar localmente la función de correlación de tonos inversa global, en concreto el factor de escalado m y el valor de desfase n usados para los bloques individuales 170, no se codifican directamente en el flujo de bits 112, sino simplemente como un residuo de predicción con respecto a una predicción obtenida a partir de factores de escalado y valores de desfase de bloques vecinos 170. Por tanto, $\{\Delta m, \Delta n, k\}$ se transmiten para cada bloque 170 en caso de que se use más de una función de correlación global, como se muestra en la Fig. 5. Supóngase, por ejemplo, que el módulo de predicción 134 usó parámetros $m_{i,j}$ y $n_{i,j}$ para la correlación de tonos inversa de las muestras de un determinado bloque i,j , en concreto el bloque central de la Fig. 5. Entonces, el módulo de predicción 134 está configurado para calcular valores de predicción $m_{i,j,pred}$ y $n_{i,j,pred}$ a partir de los factores de escalado y los valores de desfase de bloques vecinos 170 tales como, por ejemplo, $m_{i,j-1}$ y $n_{i,j-1}$. En este caso, el módulo de predicción 134 genera la diferencia entre los parámetros reales $m_{i,j}$ y $n_{i,j}$ y los valores predichos $m_{i,j,pred}$ y $n_{i,j,pred}$ que se insertará en el flujo de bits 112. Estas diferencias están denotadas en la Fig. 5 como $\Delta m_{i,j}$ y $\Delta n_{i,j}$, donde los índices i,j indican el j -ésimo bloque desde la parte superior y el i -ésimo bloque desde el lado izquierdo de la imagen 162, por ejemplo. Como alternativa, la predicción del factor de escalado m y del valor de desfase n puede obtenerse a partir de bloques ya transmitidos y no a partir de bloques vecinos de la misma imagen. Por ejemplo, los valores de predicción pueden obtenerse a partir del bloque 170 de la imagen anterior situada en la misma ubicación espacial o en una correspondiente. En particular, el valor predicho puede ser:

- un valor fijo, que o bien se transmite en la información complementaria o ya es conocido por el codificador y el decodificador,
- el valor de la variable correspondiente y el bloque anterior,
- el valor de mediana de las variables correspondientes en los bloques vecinos,
- el valor promedio de las variables correspondientes en los bloques vecinos,
- un valor lineal interpolado o extrapolado obtenido a partir de los valores de variables correspondientes en los bloques vecinos.

[0052] El mecanismo de predicción que se usa realmente para un bloque particular puede ser conocido tanto por el codificador como por el decodificador o puede depender de valores de m , n en los propios bloques vecinos, si hubiera alguno.

[0053] En la señal codificada de mejora de alta calidad proporcionada por el módulo de codificación por entropía 140 puede transmitirse la siguiente información para cada macrobloque en caso de que los módulos 136, 138 y 140 implementen una codificación según H.264/AVC. Puede incluirse información de patrón de bloque codificado (CBP) que indica cuál de los cuatro bloques 8×8 de transformación de luminancia del macrobloque y cuál de los bloques de transformación de crominancia asociados del macrobloque puede contener coeficientes de transformada distintos de cero. Si no hay ningún coeficiente de transformada distinto de cero, no se transmite información adicional para el macrobloque particular. La información adicional puede referirse al tamaño de transformada usado para codificar la componente de luminancia, es decir, el tamaño de los bloques de transformación en el que el macrobloque que consiste en muestras 16×16 de luminancia se transforma en el módulo

de transformación 136, es decir, en bloques de transformación 4x4 u 8x8. Además, el flujo de datos de capa de mejora de alta calidad puede incluir el parámetro de cuantificación QP usado en el módulo de cuantificación 138 para controlar el valor de incremento del cuantificador. Además, los coeficientes de transformada cuantificados, es decir, los niveles de coeficiente de transformada, pueden incluirse para cada macrobloque en el flujo de datos de capa de mejora de alta calidad proporcionado por el módulo de codificación por entropía 140.

[0054] Además de la información anterior, en el flujo de datos 112 debe incluirse la siguiente información. Por ejemplo, en caso de que se use más de un único esquema de correlación de tonos inversa global para la imagen actual, el valor de índice k correspondiente también tiene que transmitirse para cada bloque de la señal de mejora de alta calidad. En lo que respecta a la adaptación local, las variables Δm y Δn pueden señalizarse para cada bloque de esas imágenes, donde la transmisión de los valores correspondientes se indica en la información complementaria mediante, por ejemplo, el conjunto de parámetros de secuencia y/o el conjunto de parámetros de imagen de H.264/AVC. Para las tres nuevas variables k, Δm y Δn , deben introducirse nuevos elementos de sintaxis con esquemas de binarización correspondientes. Una binarización unaria simple puede usarse para preparar las tres variables para el esquema de codificación aritmética binaria usado en H.264/AVC. Puesto que las tres variables tienen magnitudes normalmente pequeñas, el esquema de binarización unaria simple es muy adecuado. Puesto que Δm y Δn son valores enteros con signo, puede convertirse en valores sin signo como se describe en la tabla siguiente:

Valor con signo de $\Delta m, \Delta n$	Valor sin signo (que se someterá a binarización)
0	0
1	1
2	-1
3	2
4	-2
5	3
6	-3
k	$(-1)^{k+1} \text{Redondeo_exceso}(k+2)$

[0055] Por tanto, resumiendo algunas de las realizaciones anteriores, el medio de predicción 134 puede llevar a cabo las siguientes etapas durante su modo de funcionamiento. En particular, como se muestra en la Fig. 6, el medio de predicción 134 establece, en la etapa 180, una o más funciones de correlación global $f(x)$ o, en caso de más de una función de correlación global, $f_k(x)$, denotando k el valor de índice correspondiente que apunta a la función de correlación global respectiva. Como se ha descrito anteriormente, la una o más funciones de correlación global pueden establecerse o definirse para que sean constantes a lo largo del vídeo 160 o pueden definirse o establecerse en una granularidad tosca tal como, por ejemplo, una granularidad del tamaño de una imagen 162 o de una secuencia de imágenes 164. La una o más funciones de correlación global pueden definirse de la manera descrita anteriormente. En general, las funciones de correlación global pueden ser una función no trivial, es decir, diferente de $f(x) = \text{constante}$, y pueden ser especialmente una función no lineal. En cualquier caso, cuando se use cualquiera de los ejemplos anteriores para una función de correlación global, la etapa 180 da como resultado un parámetro de función de correlación global correspondiente, tal como K, D o (x_n, y_n) , donde $n \in (1, \dots, N)$ se fija para cada sección en la que el vídeo se subdivide según la granularidad tosca.

[0056] Además, el módulo de predicción 134 establece una función de adaptación/correlación local, tal como la indicada anteriormente, en concreto una función lineal según $m \cdot x + n$. Sin embargo, otra función de adaptación/correlación local es también factible, tal como una función constante parametrizada simplemente usando un valor de desfase. El establecimiento 182 se lleva a cabo en una granularidad más fina tal como, por ejemplo, la granularidad de un tamaño más pequeño que una imagen tal como un macrobloque, bloque de transformación o par de macrobloques o incluso un fragmento dentro de una imagen, donde el fragmento es un subconjunto de macrobloques o pares de macrobloques de una imagen. Por tanto, la etapa 182 da como resultado, en el caso de la realización anterior para una función de adaptación/correlación local, un par de valores Δm y Δn fijados o definidos para cada uno de los bloques en los que las imágenes 162 del vídeo se subdividen según la granularidad más fina.

[0057] Opcionalmente, en concreto en caso de que se use más de una función de correlación global en la etapa 180, el medio de predicción 134 establece, en la etapa 184, el índice k para cada bloque de la granularidad fina usada en la etapa 182.

[0058] Aunque sería posible que el módulo de predicción 134 fijara la una o más funciones de correlación global en la etapa 180 dependiendo solamente de la función de correlación usada por el módulo de conversión descendente 116 con el fin de reducir la profundidad de bits de muestra de las muestras de imagen originales tal como, por ejemplo, usando la función de correlación inversa con respecto a las mismas como una función de correlación global en la etapa 180, debe observarse que también es posible que el medio de predicción 134 lleve a cabo todos los ajustes en las etapas 180, 182 y 184 de modo que un determinado criterio de optimización, tal como

la relación velocidad/distorsión del flujo de bits 112 resultante se vuelva extremo, por ejemplo se maximice o se minimice. De este modo, tanto las funciones de correlación global como la adaptación local de las mismas, concretamente la función de adaptación/correlación local, se adaptan a las estadísticas de tono de valor de muestra del vídeo, determinándose el mejor equilibrio entre, por un lado, la sobrecarga de codificar la información complementaria necesaria para la correlación global y, por otro lado, conseguir el mejor ajuste de la función de correlación global con respecto a las estadísticas de tono que necesitan simplemente una pequeña función de adaptación/correlación local.

[0059] Para llevar a cabo la correlación de tonos inversa real en la etapa 186, el módulo de predicción 134 usa, por un lado, para cada muestra de la señal de baja calidad reconstruida, una combinación de la función de correlación global en caso de que solo haya una función de correlación global o una de las funciones de correlación global en caso de que haya más de una y, por otro lado, la función de adaptación/correlación local, como se define en el bloque o en la sección a la que pertenece la muestra actual. En el ejemplo anterior, la combinación ha sido una suma. Sin embargo, también sería posible cualquier combinación aritmética. Por ejemplo, la combinación podría ser una aplicación en serie de ambas funciones.

[0060] Además, con el fin de informar al descodificador acerca de la función de correlación combinada usada para la correlación de tonos inversa, el módulo de predicción 134 hace que al menos información acerca de la función de adaptación/correlación local se proporcione al descodificador a través del flujo de bits 112. Por ejemplo, el módulo de predicción 134 hace que, en la etapa 188, los valores de m y n se codifiquen en el flujo de bits 112 para cada bloque de la granularidad fina. Como se ha descrito anteriormente, la codificación de la etapa 188 puede ser una codificación predictiva, según la cual los residuos de predicción de m y n se codifican en el flujo de bits en lugar de los propios valores reales, donde la predicción de los valores se obtiene a partir de los valores de m y n en los nodos vecinos 170 o de un bloque correspondiente de una imagen anterior. Dicho de otro modo, una predicción local o temporal junto con una codificación residual puede usarse para codificar los parámetros m y n .

[0061] Asimismo, para el caso en que en la etapa 180 se use más de una función de correlación global, el módulo de predicción 134 puede hacer que el índice k se codifique en el flujo de bits para cada bloque en la etapa 190. Además, el módulo de predicción 134 puede hacer que la información acerca de $f(x)$ o $f_k(x)$ se codifique en el flujo de bits en caso de que el descodificar no conozca a priori la misma información. Esto se realiza en la etapa 192. Además, el módulo de predicción 134 puede hacer que la información acerca de la granularidad y el cambio de granularidad en el vídeo para la función de adaptación/correlación local y/o la(s) función(es) de correlación global se codifique en el flujo de bits en la etapa 194.

[0062] Debe observarse que no es necesario llevar a cabo todas las etapas 180 a 194 en el orden mencionado. Incluso un funcionamiento secuencial estricto de estas etapas no es necesario. En cambio, las etapas 180 a 194 se muestran simplemente en orden secuencial con fines ilustrativos y estas etapas se llevarán a cabo preferentemente de manera solapada.

[0063] Aunque no se ha mencionado explícitamente en la descripción anterior, debe observarse que la información complementaria generada en las etapas 190 a 194 puede introducirse en la señal de capa de mejora de alta calidad o en la parte de alta calidad del flujo de bits 112 en lugar de en la parte de calidad base que se obtiene con el módulo de codificación por entropía 142.

[0064] Tras haber descrito una realización para un codificador, con respecto a la Fig. 2, a continuación se describe una realización de un descodificador. El descodificador de la Fig. 2 se indica mediante el signo de referencia 200 y comprende un medio de desmultiplexación 202, un medio de descodificación de base 204, un medio de predicción 206, un medio de descodificación residual 208 y un medio de reconstrucción 210, así como una entrada 212, una primera salida 214 y una segunda salida 216. El descodificador 200 recibe, en su entrada 212, el flujo de datos de calidad escalable proporcionado, por ejemplo, por el codificador 100 de la Fig. 1. Como se ha descrito anteriormente, la calidad escalable puede referirse a la profundidad de bits y, opcionalmente, a la reducción espacial. Dicho de otro modo, el flujo de datos en la entrada 212 puede tener un parte autocontenida que puede usarse de manera aislada para reconstruir la señal de vídeo con una profundidad de bits reducida y, opcionalmente, una resolución espacial reducida, así como una parte adicional que, en combinación con la primera parte, permite reconstruir la señal de vídeo con una mayor profundidad de bits y, opcionalmente, una mayor resolución espacial. La señal de vídeo de reconstrucción de calidad inferior se proporciona en la salida 216, mientras que la señal de vídeo de reconstrucción de calidad superior se proporciona en la salida 214.

[0065] El medio de desmultiplexación 202 divide el flujo de datos de calidad escalable entrante de la entrada 212 en el flujo de datos de codificación base y el flujo de datos de capa de mejora de alta calidad, los cuales se han mencionado con respecto a la Fig. 1. El medio de descodificación base 204 es para descodificar el flujo de datos de codificación base en la representación de calidad base de la señal de vídeo, que puede proporcionarse opcionalmente de manera directa, como en el caso del ejemplo de la Fig. 2, o de manera indirecta a través de un filtro de reducción de artefactos (no mostrado), en la salida 216. En función de la señal de vídeo de representación

de calidad base, el medio de predicción 206 forma una señal de predicción que tiene la profundidad de bits de muestras de imágenes aumentada y/o la resolución de muestreo de crominancia aumentada. El medio de descodificación 208 descodifica el flujo de datos de capa de mejora para obtener el residuo de predicción que presenta la profundidad de bits aumentada y, opcionalmente, la resolución espacial aumentada. El medio de reconstrucción 210 obtiene la señal de vídeo de alta calidad a partir de la predicción y el residuo de predicción y proporciona la misma en la salida 214 a través de un filtro opcional de reducción de artefactos.

[0066] Internamente, el medio de desmultiplexación 202 comprende un desmultiplexor 218 y un módulo de descodificación por entropía 220. Una entrada del desmultiplexor 218 está conectada a la entrada 212 y una primera salida del desmultiplexor 218 está conectada al medio de descodificación residual 208. El módulo de descodificación por entropía 220 está conectado entre otra salida del desmultiplexor 218 y el medio de descodificación base 204. El desmultiplexor 218 divide el flujo de datos de calidad escalable en el flujo de datos de capa base y en el flujo de datos de capa de mejora como si se hubieran introducido por separado en el multiplexor 144, como se ha descrito anteriormente. El módulo de descodificación por entropía 220 lleva a cabo, por ejemplo, una descodificación de Huffman o un algoritmo de descodificación aritmética con el fin de obtener los niveles de coeficiente de transformada, vectores de movimiento, información de tamaño de transformada y otros elementos sintácticos necesarios para obtener de los mismos la representación base de la señal de vídeo. En la salida del módulo de descodificación por entropía 220 se genera el flujo de datos de codificación base.

[0067] El medio de descodificación base 204 comprende un módulo de transformación inversa 222, un sumador 224, un filtro de bucle opcional 226 y un módulo de predicción 228. Los módulos 222 a 228 del medio de descodificación base 204 corresponden, con respecto a la funcionalidad y la interconexión, a los elementos 124 a 130 de la Fig. 1. Para ser más precisos, el módulo de transformación inversa 222 y el sumador 224 están conectados en serie en el orden mencionado entre el medio de desmultiplexación 202, por un lado, y el medio de predicción 206 y la salida de calidad base, respectivamente, por otro, y el filtro de bucle opcional 226 y el módulo de predicción 228 están conectados en serie en el orden mencionado entre la salida del sumador 224 y otra entrada del sumador 224. De este modo, el sumador 224 proporciona la señal de vídeo de representación base con la profundidad de bits reducida y, opcionalmente, la resolución espacial reducida que puede recibirse desde el exterior en la salida 216.

[0068] El medio de predicción 206 comprende un filtro de reducción de artefactos opcional 230 y un módulo de formación de predicciones 232, ambos módulos funcionando de manera sincronizada con respecto a los elementos 132 y 134 de la Fig. 1. Dicho de otro modo, el filtro de reducción de artefactos opcional 230 filtra opcionalmente la señal de vídeo de calidad base para reducir los artefactos de la misma, y el módulo de formación de predicciones 232 recupera imágenes predichas con mayores profundidades de bits y, opcionalmente, una mayor resolución espacial de la manera descrita anteriormente con respecto al módulo de predicción 134. Es decir, el módulo de formación de predicciones 232 puede correlacionar, a través de información complementaria incluida en el flujo de datos de calidad escalable, las muestras de imagen entrantes con respecto a un mayor intervalo dinámico y, opcionalmente, aplicar un filtro de interpolación espacial al contenido de las imágenes para aumentar la resolución espacial.

[0069] El medio de descodificación residual 208 comprende un módulo de descodificación por entropía 234 y un módulo de transformación inversa 236 que están conectados en serie entre el desmultiplexor 218 y el medio de reconstrucción 210 en el orden que acaba de mencionarse. El módulo de descodificación por entropía 234 y el módulo de transformación inversa 236 actúan conjuntamente para invertir la codificación llevada a cabo por los módulos 136, 138 y 140 de la Fig. 1. En particular, el módulo de descodificación por entropía 234 lleva a cabo, por ejemplo, una descodificación de Huffman o algoritmos de descodificación aritméticos para obtener elementos sintácticos que comprenden, entre otros, niveles de coeficiente de transformada, los cuales se transforman de manera inversa, mediante el módulo de transformación inversa 236, para obtener una señal de residuo de predicción o una secuencia de imágenes residuales. Además, el módulo de descodificación por entropía 234 revela la información complementaria generada en las etapas 190 a 194 en el lado del codificador, de modo que el módulo de formación de predicciones 232 puede emular el procedimiento de correlación inversa llevado a cabo en el lado del codificador mediante el módulo de predicción 134, como se ha mencionado anteriormente.

[0070] De manera similar a la Fig. 6, que se refiere al codificador, la Fig. 7 muestra el modo de funcionamiento del módulo de formación de predicciones 232 y, parcialmente, del módulo de descodificación por entropía 234 en mayor detalle. Como se muestra en la misma, el proceso de obtener la predicción a partir de la señal de capa base reconstruida en el lado del descodificador empieza con una acción conjunta del desmultiplexor 218 y del descodificador por entropía 234 para obtener en la etapa 280 la información de granularidad codificada en la etapa 194, obtener en la etapa 282 la información acerca de la(s) función(es) de correlación global codificada en la etapa 192, obtener en la etapa 284 el valor de índice k para cada bloque de granularidad fina codificado en la etapa 190, y obtener en la etapa 286 los parámetros m y n de función de adaptación/correlación local para cada bloque de granularidad fina codificados en la etapa 188. Como se ilustra mediante las líneas discontinuas, las etapas 280 a 284 son opcionales y su aplicación depende de la realización que esté utilizándose.

[0071] En la etapa 288, el módulo de formación de predicciones 232 lleva a cabo la correlación de tonos inversa en función de la información obtenida en las etapas 280 a 286, emulando así exactamente la correlación de tonos inversa que se ha llevado a cabo en el lado del codificador en la etapa 186. De manera similar a la descripción con respecto a la etapa 188, la obtención de los parámetros m , n puede comprender una descodificación predictiva, donde un valor de residuo de predicción se obtiene a partir de la parte de alta calidad del flujo de datos que entra en el desmultiplexor 218 usando, por ejemplo, descodificación por entropía como la llevada a cabo por el descodificador de entropía 234, y obtener los valores reales de m y n sumando estos valores de residuo de predicción a un valor de predicción obtenido mediante predicción local y/o temporal.

[0072] El medio de reconstrucción 210 comprende un sumador 238 cuyas entradas están conectadas a la salida del módulo de formación de predicciones 232 y a la salida del módulo de transformación inversa 236, respectivamente. El sumador 238 suma el residuo de predicción y la señal de predicción para obtener la señal de vídeo de alta calidad que presenta la profundidad de bits aumentada y, opcionalmente, una resolución espacial aumentada, la cual se transfiere a través de un filtro de reducción de artefactos opcional 240 a la salida 214.

[0073] Por tanto, como puede deducirse a partir de la Fig. 2, un descodificador de calidad base puede reconstruir una señal de vídeo de calidad base a partir del flujo de datos de calidad escalable en la entrada 212 y, para ello, puede no incluir los elementos 230, 232, 238, 234, 236 y 240. Por otro lado, un descodificador de alta calidad puede no incluir la salida 216.

[0074] Dicho de otro modo, en el proceso de descodificación, la descodificación de la representación de calidad base es sencilla. Para la descodificación de la señal de alta calidad, en primer lugar tiene que descodificarse la señal de calidad base, lo que es llevado a cabo por los módulos 218 a 228. Después, el proceso de predicción descrito anteriormente con respecto al módulo 232 y al módulo opcional 230, se utiliza usando la representación base descodificada. Los coeficientes de transformación cuantificados de la señal de mejora de alta calidad son escalados y transformados de manera inversa por el módulo de transformación inversa 236, por ejemplo como se especifica en la norma H.264/AVC, con el fin de obtener las muestras residuales o de señales de diferencia, que se añaden a la predicción obtenida a partir de las muestras de representación base descodificadas mediante el módulo de predicción 232. Como etapa final en el proceso de descodificación de la señal de vídeo de alta calidad que va a proporcionarse en la salida 214, puede utilizarse un filtro opcional con el fin de eliminar o reducir artefactos de codificación visualmente molestos. Debe observarse que el bucle de predicción compensada por movimiento que incluye los módulos 226 y 228 está totalmente autocontenido usando solamente la representación de calidad base. Por lo tanto, la complejidad de la descodificación es moderada y no es necesario un filtro de interpolación que opere en la alta profundidad de bits, y, opcionalmente, en datos de imagen de alta resolución espacial, en el proceso de predicción compensada por movimiento del módulo de predicción 228.

[0075] Con respecto a las realizaciones anteriores, debe mencionarse que los filtros de reducción de artefactos 132 y 230 son opcionales y pueden omitirse. Lo mismo se aplica a los filtros de bucle 124 y 226, respectivamente, y al filtro 240. Por ejemplo, con respecto a la Fig. 1, debe observarse que los filtros 124 y 132 pueden sustituirse por un solo filtro común como, por ejemplo, un filtro de desbloqueo cuya salida está conectada tanto a la entrada del módulo de predicción de movimiento 126 como a la entrada del módulo de correlación de tonos inversa 134. Asimismo, los filtros 226 y 230 pueden ser sustituidos por un solo filtro común, tal como un filtro de desbloqueo, cuya salida está conectada a la salida 216, la entrada del módulo de formación de predicciones 232 y la entrada del módulo de predicción 228. Además, en aras de una mayor exhaustividad, debe observarse que los módulos de predicción 228 y 126, respectivamente, no predicen necesariamente de manera temporal muestras en los macrobloques de una imagen actual. En cambio, también puede usarse una predicción espacial o una intra-predicción que usa muestras de la misma imagen. En particular, el tipo de predicción puede elegirse para cada macrobloque, por ejemplo, o en alguna otra granularidad. Además, la presente invención no está limitada a la codificación de vídeo. En cambio, la descripción anterior también puede aplicarse a la codificación de imágenes fijas. Por consiguiente, el bucle de predicción compensada por movimiento que incluye los elementos 118, 128, 130, 126 y 124 y los elementos 224, 228 y 226, respectivamente, también puede suprimirse. Asimismo, la codificación por entropía mencionada no tiene que llevarse a cabo necesariamente.

[0076] Para ser más precisos, en las realizaciones anteriores, la codificación de capa base 118-130, 142 se basó en predicción compensada por movimiento basada en una reconstrucción de imágenes ya codificadas con pérdidas. En este caso, la reconstrucción del proceso de codificación base también puede considerarse como parte del proceso de formación de predicciones de alta calidad, como el realizado en la descripción anterior. Sin embargo, en caso de una codificación sin pérdidas de la representación base, una reconstrucción no sería necesaria y la señal convertida de manera descendente podría reenviarse directamente a los medios 132, 134, respectivamente. En caso de una predicción no basada en compensación de movimiento en una codificación de capa base con pérdidas, la reconstrucción para reconstruir la señal de calidad base en el lado del codificador estaría especialmente dedicada a la formación de predicciones de alta calidad en 104. Dicho de otro modo, la anterior asociación de los elementos 116-134 y 142 a los medios 102, 104 y 108, respectivamente, podría llevarse a cabo de otra manera. En particular, puede considerarse que el módulo de codificación por entropía 142 forma parte del medio de codificación base 102,

donde el medio de predicción comprende simplemente los módulos 132 y 134, y el medio de combinación 108 comprende simplemente el multiplexor 144. Este enfoque está relacionado con la asociación módulo/medio usada en la Fig. 2, ya que el medio de predicción 206 no comprende la predicción base de compensación por movimiento. Sin embargo, adicionalmente, puede considerarse que el medio de desmultiplexación 202 no incluye el módulo de entropía 220, de modo que el medio de descodificación base comprende además el módulo de descodificación por entropía 220. Sin embargo, ambos enfoques generan el mismo resultado ya que la predicción en 104 se lleva a cabo según una representación del material fuente con la profundidad de bits reducida y, opcionalmente, la resolución espacial reducida que se codifica sin pérdidas y que puede obtenerse sin pérdidas a partir del flujo de bits de calidad escalable y el flujo de datos de capa base, respectivamente. Según el enfoque de la Fig. 1, la predicción 134 se basa en una reconstrucción del flujo de datos de codificación base, mientras que en el caso del enfoque alternativo, la reconstrucción comenzaría a partir de una versión codificada intermedia o una versión medio codificada de la señal de calidad base que carece de la codificación sin pérdidas según el módulo 142 para codificarse completamente en el flujo de datos de capa base. A este respecto, debe observarse además que la conversión descendente en el módulo 116 no tiene por qué ser llevada a cabo por el codificador 100. En cambio, el codificador 100 puede tener dos entradas, una para recibir la señal de alta calidad y otra para recibir la versión convertida de manera descendente, desde el exterior.

[0077] En las realizaciones descritas anteriormente, la calidad escalable se refirió simplemente a la profundidad de bits y, opcionalmente, a la resolución espacial. Sin embargo, las realizaciones anteriores pueden extenderse fácilmente para incluir escalabilidad temporal, escalabilidad de formato cromático y escalabilidad de calidad granular fina.

[0078] Por consiguiente, las realizaciones anteriores de la presente invención crean un concepto de codificación escalable de contenido de vídeo o imágenes con diferentes granularidades en lo que respecta a la profundidad de bits de muestra y, opcionalmente, la resolución espacial usando una correlación de tonos inversa localmente adaptativa. Según realizaciones de la presente invención, los procesos de predicción tanto temporal como espacial especificados en la extensión de codificación de vídeo escalable H.264/AVC se amplían de modo que incluyan correlaciones desde una fidelidad de profundidad de bits de muestra inferior a otra superior, así como, opcionalmente, desde una resolución espacial inferior a otra superior. La extensión descrita anteriormente de SVC con vistas a la escalabilidad en lo que respecta a la profundidad de bits de muestra y, opcionalmente, la resolución espacial permite al codificador almacenar una representación de calidad base de una secuencia de vídeo, que puede descodificarse mediante cualquier descodificador de vídeo heredado junto con una señal de mejora para una mayor profundidad de bits y, opcionalmente, una mayor resolución espacial, que es ignorada por los descodificadores de vídeo heredados. Por ejemplo, la representación de calidad base podría incluir una versión de 8 bits de la secuencia de vídeo en resolución CIF, concretamente 352x288 muestras, mientras que la señal de mejora de alta calidad incluye un "refinamiento" hacia una versión de 10 bits en resolución 4CIF, es decir, 704x476 muestras de la misma secuencia. En una configuración diferente, también es posible usar la misma resolución espacial para las representaciones de calidad base y de mejora, de modo que la señal de mejora de alta calidad solo contiene un refinamiento de la profundidad de bits de muestra, por ejemplo de 8 a 10 bits.

[0079] Dicho de otro modo, las realizaciones mencionadas anteriormente permiten formar un codificador de vídeo, es decir, un codificador o un descodificador, para codificar, es decir, codificar o descodificar, una representación estratificada de una señal de vídeo que comprende un procedimiento de codificación de vídeo normalizado para codificar una capa de calidad base, un procedimiento de predicción para llevar a cabo una predicción de la señal de capa de mejora de alta calidad usando la señal de calidad base reconstruida, y un procedimiento de codificación residual para codificar el residuo de predicción de la señal de capa de mejora de alta calidad. A este respecto, la predicción puede llevarse a cabo usando una función de correlación desde el intervalo dinámico asociado a la capa de calidad base hasta el intervalo dinámico asociado a la capa de mejora de alta calidad. Además, la función de correlación puede generarse como la suma de una función de correlación global, que sigue los esquemas de correlación de tonos inversa descritos anteriormente, y una adaptación local. A su vez, la adaptación local puede llevarse a cabo escalando los valores de muestra x de la capa de calidad base y añadiendo un valor de desfase según $m \cdot x \cdot n$. En cualquier caso, la codificación residual puede llevarse a cabo según la norma H.264/AVC.

[0080] Dependiendo de la implementación real, el esquema de codificación inventivo puede implementarse en hardware o en software. Por lo tanto, la presente invención también se refiere a un programa informático que puede almacenarse en un medio legible por ordenador, tal como un CD, un disco o cualquier otro portador de datos. Por lo tanto, la presente invención es también un programa informático que tiene un código de programa que, cuando se ejecuta en un ordenador, lleva a cabo el procedimiento inventivo descrito en relación con las figuras anteriores. En particular, las implementaciones de los medios y módulos de la Fig. 1 y 2 pueden comprender subrutinas que se ejecutan en una CPU, partes de circuito de un ASIC o similares, por ejemplo.

[0081] Por tanto, entre otras cosas, las realizaciones anteriores describen un codificador para codificar datos fuente de imagen o vídeo (160) en un flujo de datos de calidad escalable (112), que comprende un medio de

codificación base (102) para codificar los datos fuente de imagen o vídeo (160) en un flujo de datos de codificación base que representa una representación de los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen; un medio de correlación (104) para correlacionar muestras de la representación de los datos fuente de imagen o vídeo (160) con la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a una segunda profundidad de bits de muestras de imagen mayor que la primera profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en los datos fuente de imagen o vídeo (160) o que varían en una primera granularidad, y una función de correlación local que modifica localmente la una o más funciones de correlación global en una segunda granularidad más fina que la primera granularidad para obtener una predicción de los datos fuente de imagen o vídeo que tienen la segunda profundidad de bits de muestras de imagen; un medio de codificación residual (106) para codificar un residuo de predicción de la predicción en un flujo de datos de capa de mejora de profundidad de bits; y medios de combinación (108) para formar el flujo de datos de calidad escalable en función del flujo de datos de codificación base, la función de correlación local y el flujo de datos de capa de mejora de profundidad de bits, de modo que la función de correlación local puede obtenerse a partir del flujo de datos de calidad escalable.

[0082] El medio de correlación puede comprender medios (124, 126, 128, 130, 132) para reconstruir una imagen o un vídeo de reconstrucción de baja profundidad de bits como la representación de los datos fuente de imagen o vídeo, donde la primera profundidad de bits de muestras de imagen se basa en el flujo de datos de codificación base, presentando la imagen o el vídeo de reconstrucción de baja profundidad de bits la primera profundidad de bits de muestras de imagen.

[0083] El medio de correlación (104) puede estar adaptado para correlacionar las muestras de la representación de los datos fuente de imagen o vídeo (160) con la primera profundidad de bits de muestra de imagen mediante el uso de una función de correlación combinada que es una combinación aritmética de una de la una o más funciones de correlación global y la función de correlación local. El medio de correlación (104) puede usar más de una función de correlación global, y el medio de combinación (108) está adaptado para formar el flujo de datos de calidad escalable (112) de modo que la una de las más de una función de correlación global puede obtenerse a partir del flujo de datos de calidad escalable. La combinación aritmética puede comprender una operación de suma.

[0084] El medio de combinación (108) y el medio de correlación (104) pueden adaptarse de modo que la segunda granularidad subdivide los datos fuente de imagen o vídeo (160) en una pluralidad de bloques de imagen (170) y el medio de correlación puede adaptarse de modo que la función de correlación local sea $m \cdot s + n$, donde m y n varían en la segunda granularidad y donde el medio de combinación está adaptado de modo que m y n se definen en el flujo de datos de calidad escalable para cada bloque de imagen (170) de los datos fuente de imagen o vídeo (160), de modo que m y n pueden ser diferentes en la pluralidad de bloques de imagen.

[0085] El medio de correlación (134) puede adaptarse de modo que la segunda granularidad varía en los datos fuente de imagen o vídeo (160), y el medio de combinación (108) puede adaptarse de modo que la segunda granularidad también pueda obtenerse a partir del flujo de datos de calidad escalable.

[0086] El medio de combinación (108) y el medio de correlación (134) pueden adaptarse de modo que la segunda granularidad divida los datos fuente de imagen o vídeo (160) en una pluralidad de bloques de imagen (170), y el medio de combinación (108) puede adaptarse de modo que un residuo de función de correlación local (Δm , Δn) se incorpore en el flujo de datos de calidad escalable (112) para cada bloque de imagen (170), y la función de correlación local de un bloque de imagen predeterminado de los datos fuente de imagen o vídeo (160) puede obtenerse a partir del flujo de datos de calidad escalable mediante el uso de una predicción espacial y/o temporal a partir de uno o más bloques de imagen vecinos o un bloque de imagen correspondiente de una imagen de los datos fuente de imagen o vídeo anterior a una imagen a la que pertenece el bloque de imagen predeterminado, y el residuo de función de correlación local del bloque de imagen predeterminado.

[0087] El medio de combinación (108) puede adaptarse de modo que al menos una de las una o más funciones de correlación global también pueda obtenerse a partir del flujo de datos de calidad escalable.

[0088] El medio de codificación base (102) puede comprender un medio (116) para correlacionar muestras que representan la imagen con la segunda profundidad de bits de muestras de imagen desde el segundo intervalo dinámico hasta el primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen para obtener una imagen de calidad reducida; y medios (118, 120, 122, 124, 126, 128, 130) para codificar la imagen de calidad reducida para obtener el flujo de datos de codificación base.

REIVINDICACIONES

1. Codificador para codificar datos fuente de imagen o vídeo (160) en un flujo de datos de calidad escalable (112), que comprende:

un medio de codificación base (102) para codificar los datos fuente de imagen o vídeo (160) en un flujo de datos de codificación base que representa una representación de los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen;
 un medio de correlación (104) para correlacionar muestras de la representación de los datos fuente de imagen o vídeo (160) que presentan la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a una segunda profundidad de bits de muestras de imagen mayor que la primera profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en los datos fuente de imagen o vídeo (160) o que varían en una primera granularidad, y una función de correlación local que modifica localmente la una o más funciones de correlación global en una segunda granularidad más fina que la primera granularidad para obtener una predicción de los datos fuente de imagen o vídeo que presentan la segunda profundidad de bits de muestras de imagen;
 un medio de codificación residual (106) para codificar un residuo de predicción de la predicción en un flujo de datos de capa de mejora de profundidad de bits; y
 un medio de combinación (108) para formar el flujo de datos de calidad escalable en función del flujo de datos de codificación base, la función de correlación local y el flujo de datos de capa de mejora de profundidad de bits, de modo que la función de correlación local puede obtenerse a partir del flujo de datos de calidad escalable.

2. Decodificador para decodificar un flujo de datos de calidad escalable en el que están codificados datos fuente de imagen o vídeo, comprendiendo el flujo de datos de calidad escalable un flujo de datos de capa base que representa los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen, representando un flujo de datos de capa de mejora de profundidad de bits un residuo de predicción con una segunda profundidad de bits de muestras de imagen que es mayor que la primera profundidad de bits de muestras de imagen, y una función de correlación local definida en una segunda granularidad, comprendiendo el decodificador:

un medio (204) para decodificar el flujo de datos de capa base en datos de imagen o vídeo reconstruidos con profundidad de bits inferior;
 un medio (208) para decodificar el flujo de datos de mejora de profundidad de bits en el residuo de predicción;
 un medio (206) para correlacionar muestras de los datos de imagen o vídeo reconstruidos con profundidad de bits inferior que presentan la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a la segunda profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en el vídeo o que varían en una primera granularidad, y una función de correlación local que modifica localmente la una o más funciones de correlación global en la segunda granularidad que es más pequeña que la primera granularidad, para obtener una predicción de los datos fuente de imagen o vídeo que presentan la segunda profundidad de bits de muestras de imagen; y
 un medio (210) para reconstruir la imagen con la segunda profundidad de bits de muestras de imagen en función de la predicción y el residuo de predicción.

3. Decodificador según la reivindicación 2, en el que el medio de correlación (206) está adaptado para correlacionar las muestras de datos de imagen o vídeo reconstruidos con profundidad de bits inferior que presentan la primera profundidad de bits de muestras de imagen mediante la utilización de una función de correlación combinada que es una combinación aritmética de una de las una o más funciones de correlación global y la función de correlación local.

4. Decodificador según la reivindicación 3, en el que el medio de correlación (206) usa más de una función de correlación global y el medio (208) para decodificar el flujo de datos de mejora de profundidad de bits está adaptado para identificar la una de las más de una función de correlación global a partir del flujo de datos de mejora de profundidad de bits.

5. Decodificador según la reivindicación 3 ó 4, en el que la combinación aritmética comprende una operación de suma.

6. Decodificador según cualquiera de las reivindicaciones 2 a 5, en el que el medio (208) para

descodificar el flujo de datos de mejora de profundidad de bits y el medio de correlación (104) están adaptados de modo que la segunda granularidad subdivide los datos fuente de imagen o vídeo (160) en una pluralidad de bloques de imagen (170) y el medio de correlación (206) está adaptado de modo que la función de correlación local es $m \cdot s + n$, donde m y n varían en la segunda granularidad y el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits está adaptado para obtener m y n a partir del flujo de datos de mejora de profundidad de bits para cada bloque de imagen (170) de los datos fuente de imagen o vídeo (160) de modo que m y n pueden ser diferentes entre la pluralidad de bloques de imagen.

7. Descodificador según cualquiera de las reivindicaciones 2 a 6, en el que el medio de correlación (206) está adaptado de modo que la segunda granularidad varía en los datos fuente de imagen o vídeo (160), y el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits está adaptado para obtener la segunda granularidad a partir del flujo de datos de mejora de profundidad de bits.

8. Descodificador según cualquiera de las reivindicaciones 2 a 7, en el que el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits y el medio de correlación (206) están adaptados de modo que la segunda granularidad divide los datos fuente de imagen o vídeo (160) en una pluralidad de bloques de imagen (170), y el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits está adaptado para obtener un residuo de función de correlación local ($\Delta m, \Delta n$) a partir del flujo de datos de mejora de profundidad de bits para cada bloque de imagen (170), y obtener la función de correlación local de un bloque de imagen predeterminado de los datos fuente de imagen o vídeo (160) mediante la utilización de una predicción espacial y/o temporal a partir de uno o más bloques de imagen vecinos o un bloque de imagen correspondiente de una imagen de los datos fuente de imagen o vídeo anterior a una imagen a la que pertenece el bloque de imagen predeterminado, y el residuo de función de correlación local del bloque de imagen predeterminado.

9. Descodificador según cualquiera de las reivindicaciones 2 a 8, en el que el medio de correlación (206) está adaptado de modo que al menos una de las una o más funciones de correlación global no es lineal.

10. Descodificador según cualquiera de las reivindicaciones 2 a 9, en el que el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits está adaptado para obtener al menos una de las una o más funciones de correlación global a partir del flujo de datos de mejora de profundidad de bits.

11. Descodificador según cualquiera de las reivindicaciones 2 a 10, en el que el medio de correlación (206) está adaptado de modo que al menos una de las funciones de correlación global se define como

$2^{M-N-K} x + 2^{M-1} - 2^{N-1-K}$, donde x es una muestra de la representación de los datos fuente de imagen o vídeo con la primera profundidad de bits de muestras de imagen, N es la primera profundidad de bits de muestras de imagen, M es la segunda profundidad de bits de muestras de imagen y K es un parámetro de correlación, $2^{M-N-K} x + D$, donde N es la primera profundidad de bits de muestras de imagen, M es la segunda profundidad de bits de muestras de imagen, K y D son parámetros de correlación, $\text{redondeo_defecto}(2^{M-N-K} x + 2^{M-2N-K} x + D)$, donde $\text{redondeo_defecto}(a)$ redondea al entero inferior más próximo, N es la primera profundidad de bits de muestras de imagen, M es la segunda profundidad de bits de muestras de imagen, y K y D son parámetros de correlación, una función lineal orientada a fragmentos para correlacionar las muestras desde el primer intervalo dinámico hasta el segundo intervalo dinámico con información de puntos de interpolación que define la correlación lineal orientada a fragmentos, o una tabla de consulta para indexarse mediante la utilización de muestras del primer intervalo dinámico y para proporcionar a continuación muestras del segundo intervalo dinámico.

12. Descodificador según la reivindicación 11, en el que el medio (208) para descodificar el flujo de datos de mejora de profundidad de bits está adaptado para obtener el (los) parámetro(s) de correlación, información de puntos de interpolación o la tabla de consulta a partir del flujo de datos de mejora de profundidad de bits.

13. Procedimiento para codificar datos fuente de imagen o vídeo (160) en un flujo de datos de calidad escalable (112), que comprende:

codificar los datos fuente de imagen o vídeo (160) en un flujo de datos de codificación base que representa una representación de los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen;
correlacionar muestras de la representación de los datos fuente de imagen o vídeo (160) que presentan la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a una segunda profundidad de bits de muestras de imagen mayor que la primera profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en los datos fuente de imagen o vídeo (160) o que varían en una

primera granularidad, y una función de correlación local que modifica localmente las una o más funciones de correlación global en una segunda granularidad más fina que la primera granularidad para obtener una predicción de los datos fuente de imagen o vídeo que presentan la segunda profundidad de bits de muestras de imagen;

codificar un residuo de predicción de la predicción en un flujo de datos de capa de mejora de profundidad de bits; y

formar el flujo de datos de calidad escalable en función del flujo de datos de codificación base, la función de correlación local y el flujo de datos de capa de mejora de profundidad de bits, de modo que la función de correlación local puede obtenerse a partir del flujo de datos de calidad escalable.

14. Procedimiento para descodificar un flujo de datos de calidad escalable en el que están codificados datos fuente de imagen o vídeo, comprendiendo el flujo de datos de calidad escalable un flujo de datos de capa base que representa los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen, representando un flujo de datos de capa de mejora de profundidad de bits un residuo de predicción con una segunda profundidad de bits de muestras de imagen que es mayor que la primera profundidad de bits de muestras de imagen, y una función de correlación local definida en una segunda granularidad, comprendiendo el procedimiento:

descodificar el flujo de datos de capa base en datos de imagen o vídeo reconstruidos con profundidad de bits inferior;

descodificar el flujo de datos de mejora de profundidad de bits en el residuo de predicción;

correlacionar muestras de los datos de imagen o vídeo reconstruidos con profundidad de bits inferior que presentan la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a la segunda profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en el vídeo o que varían en una primera granularidad, y una función de correlación local que modifica localmente las una o más funciones de correlación global en la segunda granularidad que es más pequeña que la primera granularidad, para obtener una predicción de los datos fuente de imagen o vídeo que presentan la segunda

profundidad de bits de muestras de imagen; y
reconstruir la imagen con la segunda profundidad de bits de muestras de imagen en función de la predicción y el residuo de predicción.

15. Flujo de datos de calidad escalable en el que están codificados datos fuente de imagen o vídeo, comprendiendo el flujo de datos de calidad escalable:

un flujo de datos de capa base que representa los datos fuente de imagen o vídeo con una primera profundidad de bits de muestras de imagen,

un flujo de datos de capa de mejora de profundidad de bits que representa un residuo de predicción con una segunda profundidad de bits de muestras de imagen que es mayor que la primera profundidad de bits de muestras de imagen, y

una función de correlación local definida en una segunda granularidad,

en el que una reconstrucción de la imagen con la segunda profundidad de bits de muestras de imagen puede obtenerse a partir del residuo de predicción y una predicción obtenida correlacionando muestras de los datos de imagen o vídeo reconstruidos con profundidad de bits inferior que presentan la primera profundidad de bits de muestras de imagen desde un primer intervalo dinámico correspondiente a la primera profundidad de bits de muestras de imagen hasta un segundo intervalo dinámico mayor que el primer intervalo dinámico y correspondiente a la segunda profundidad de bits de muestras de imagen, mediante la utilización de una o más funciones de correlación global que son constantes en el vídeo o que varían en una primera granularidad, y una función de correlación local que modifica localmente las una o más funciones de correlación global en la segunda granularidad que es más pequeña que la primera granularidad.

16. Programa informático que presenta un código de programa para llevar a cabo, cuando se ejecuta en un ordenador, un procedimiento según la reivindicación 13 ó 14.

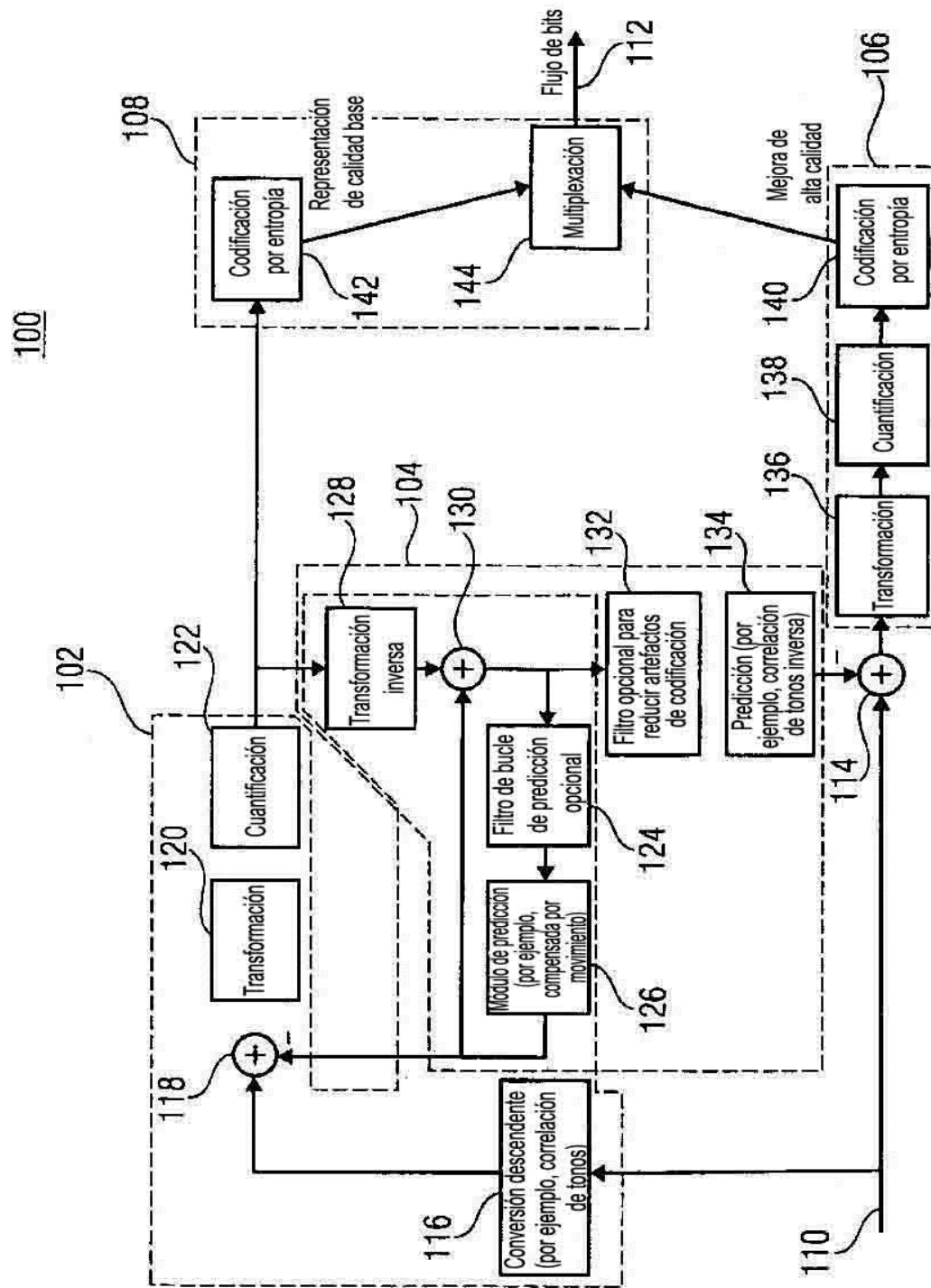


FIG 1

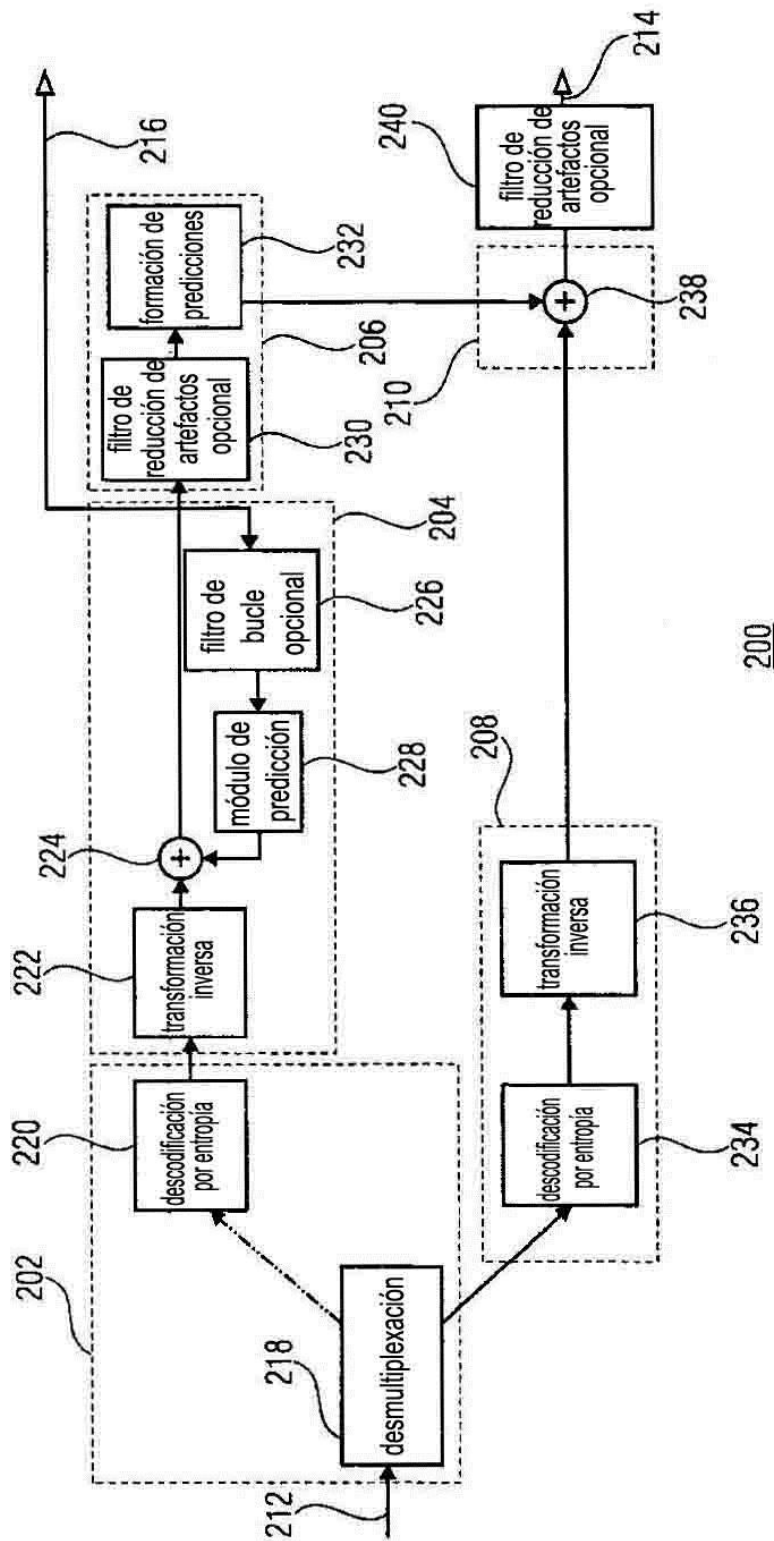


FIG 2

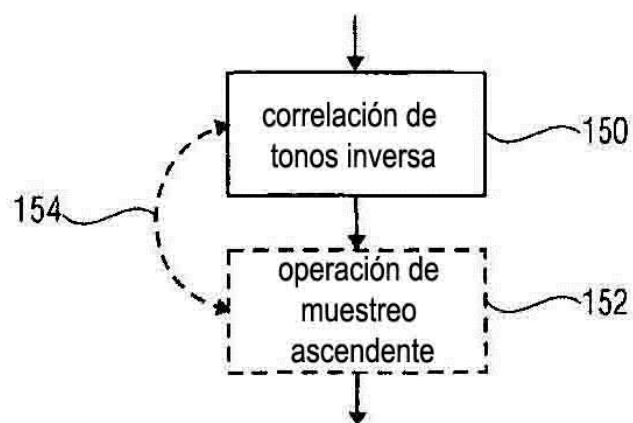


FIG 3

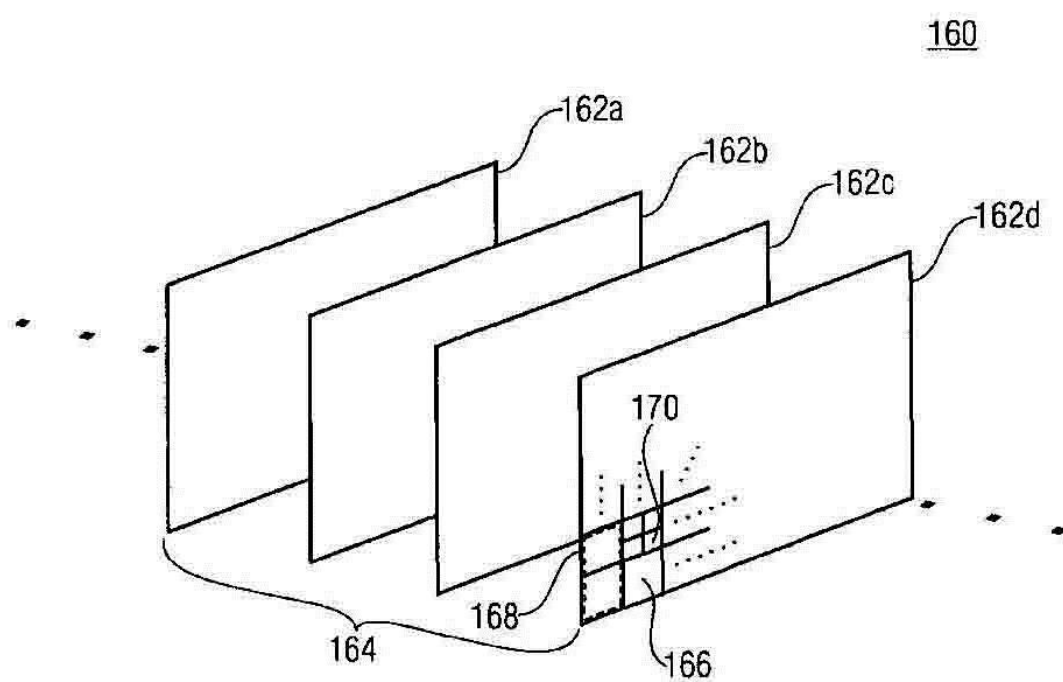


FIG 4

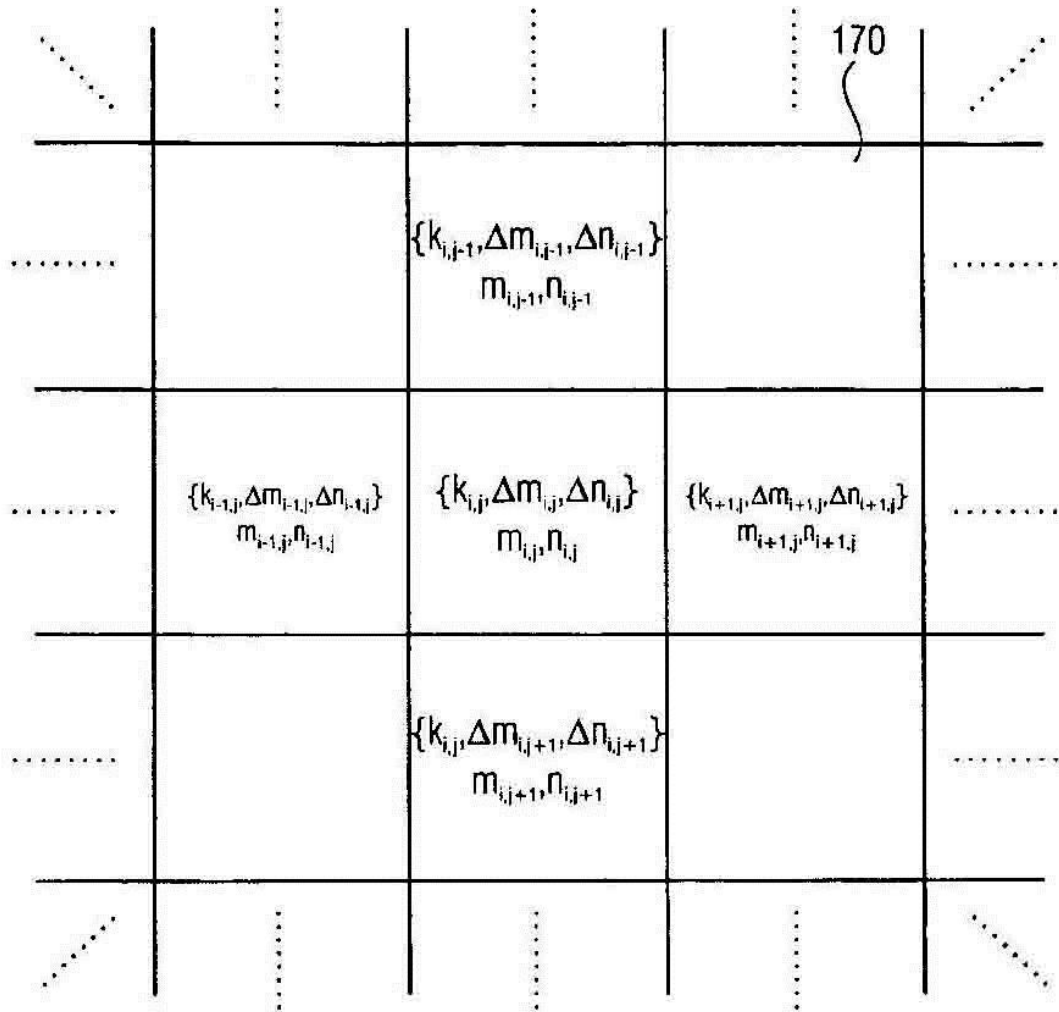


FIG 5

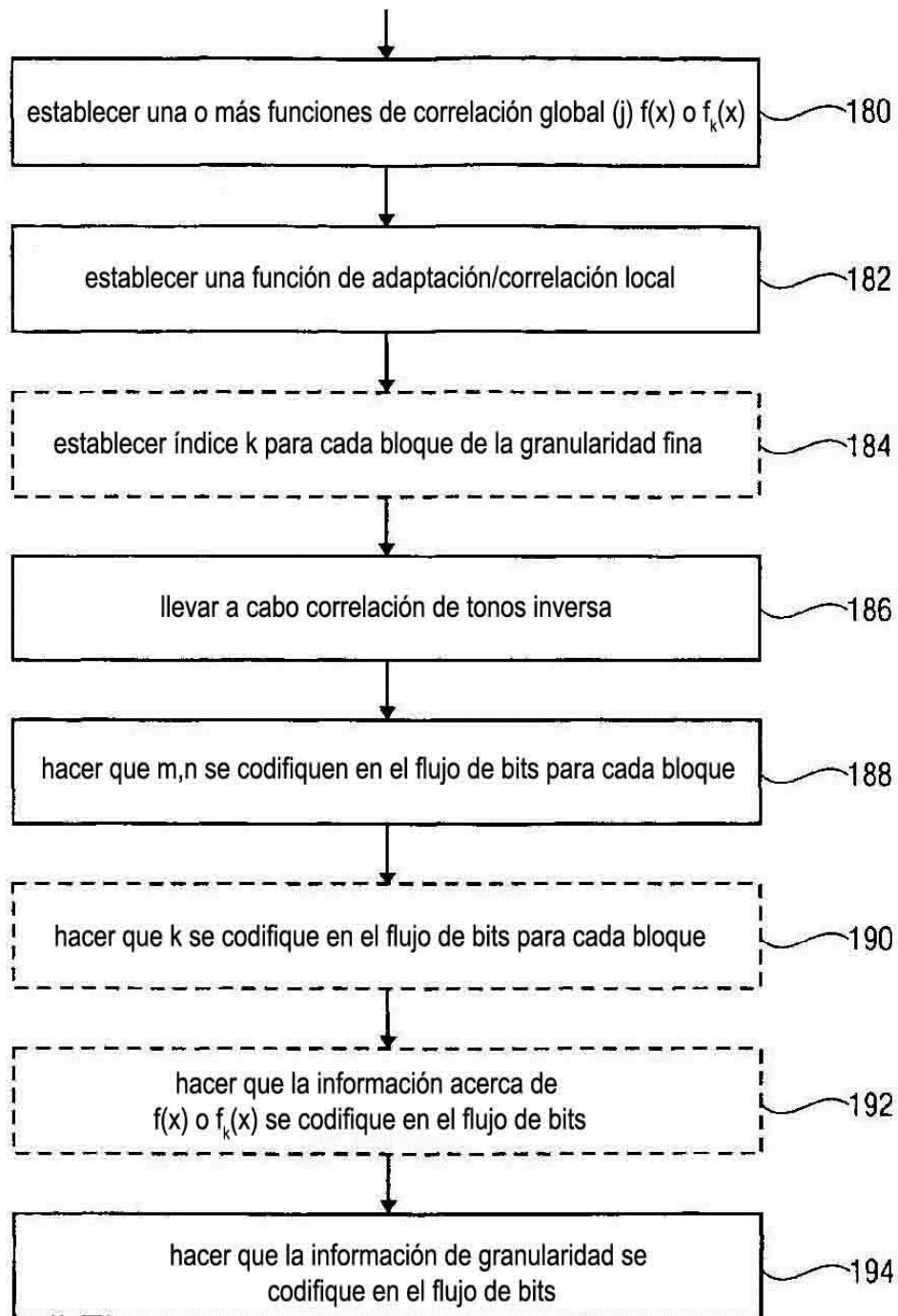


FIG 6

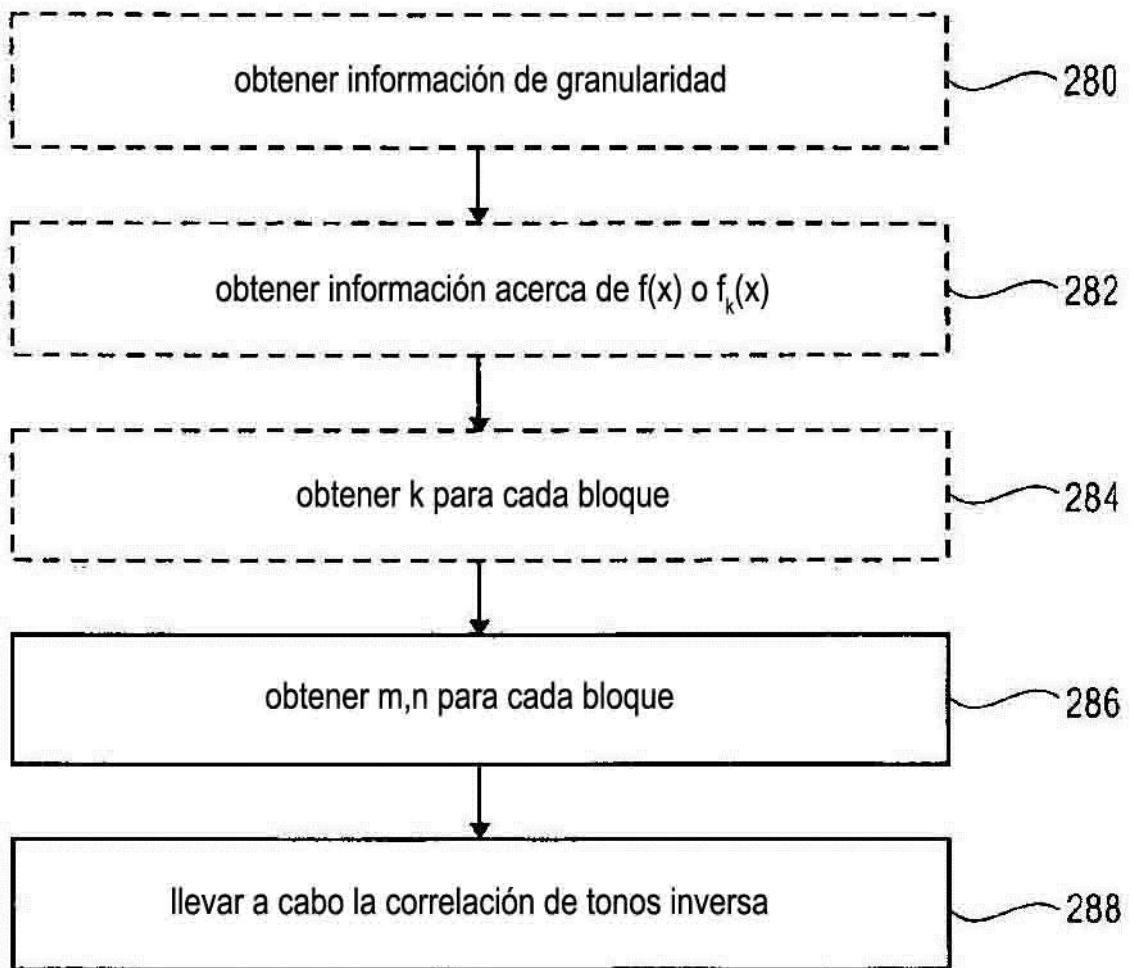


FIG 7

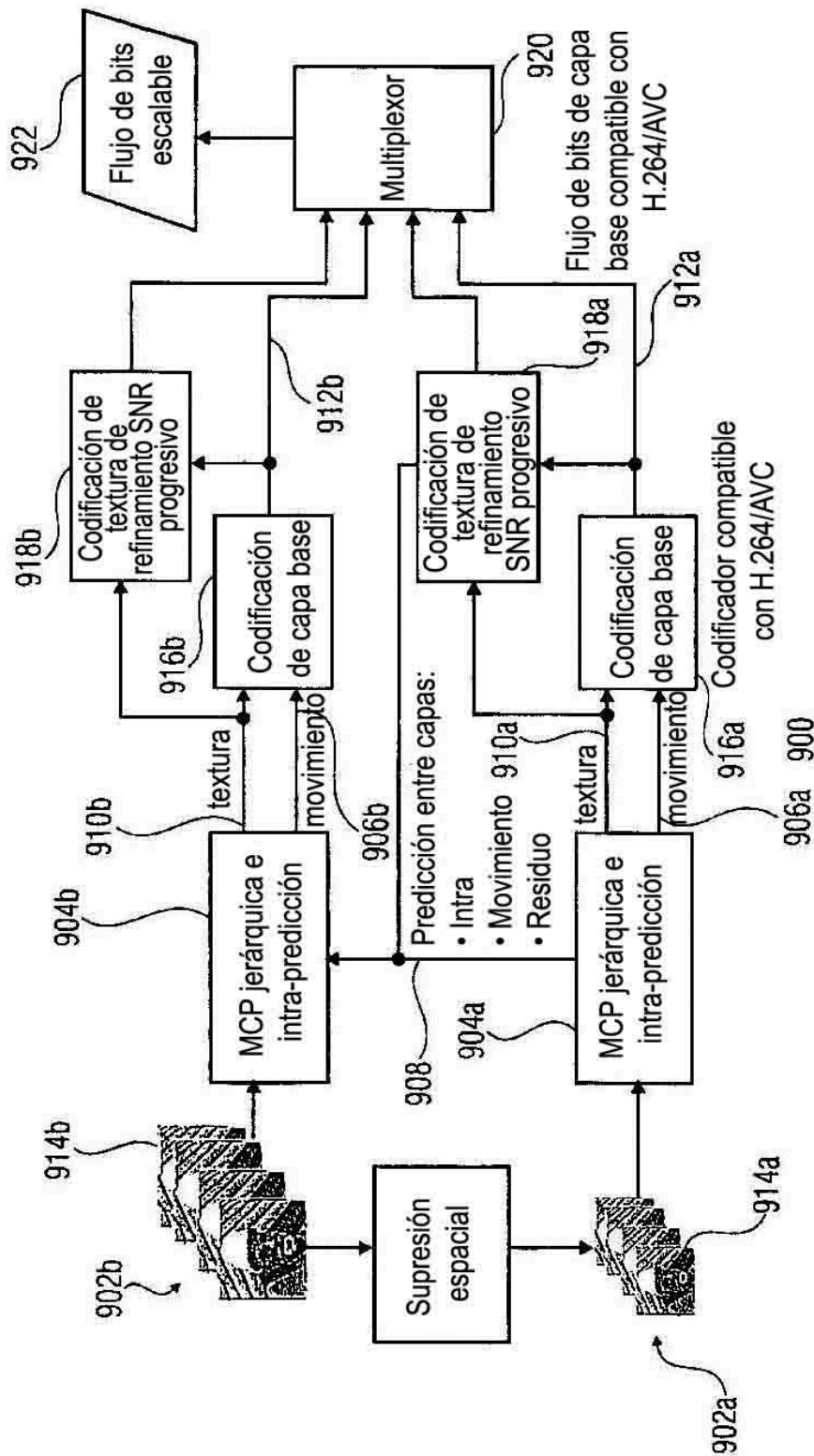


FIG 8