(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 775 070**

(51) Int. CI.:

H04N 19/114	(2014.01)	***H04N 19/48***	(2014.01)
H04N 19/117	(2014.01)	***H04N 19/51***	(2014.01)
H04N 19/137	(2014.01)	***H04N 19/513***	(2014.01)
H04N 19/147	(2014.01)	***H04N 19/573***	(2014.01)
H04N 19/162	(2014.01)	***H04N 19/577***	(2014.01)
H04N 19/17	(2014.01)	***H04N 19/58***	(2014.01)
H04N 19/18	(2014.01)	***H04N 19/61***	(2014.01)
H04N 19/186	(2014.01)	***H04N 19/82***	(2014.01)
H04N 19/433	(2014.01)		
H04N 19/44	(2014.01)		

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **27.06.2003 E 10005839 (5)**

(97) Fecha y número de publicación de la concesión europea: **04.12.2019 EP 2254339**

(54) Título: **Interpolación mejorada de cuadros de compresión de vídeo**

(30) Prioridad:

28.06.2002 US 187395

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
23.07.2020

(73) Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (100.0%)
100 Potrero Avenue
San Francisco, CA 94103-4813, US

(72) Inventor/es:

DEMOS, GARY, A.

(74) Agente/Representante:

ELZABURU, S.L.P

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

Interpolación mejorada de cuadros de compresión de vídeo

Campo técnico

Esta invención se refiere a un método de compresión de video y, más particularmente, a un método de interpolación mejorada de cuadros de compresión de video en sistemas de codificación de tipo MPEG.

Antecedentes

Compresión de video MPEG

MPEG-2 y MPEG-4 son estándares internacionales de compresión de video que definen sintaxis de video respectivas que proporcionan una forma eficiente de representar secuencias de imágenes en forma de datos codificados más compactos. Una introducción a MPEG-2 se proporciona por Tudor P.N., "MPEG-2 video compression tutorial", Coloquio del IEE sobre MPEG-2 (Resumen Nº 1995/012), Londres, Reino Unido, 24 de enero de 1995, páginas 2/1 a 2/8. El lenguaje de los bits codificados es la "sintaxis". Por ejemplo, unos pocos testigos pueden representar un bloque entero de muestras (por ejemplo, 64 muestras para MPEG-2). Ambos estándares MPEG también describen un proceso de decodificación (reconstrucción) donde los bits codificados se correlacionan desde la representación compacta a una aproximación del formato original de la secuencia de imágenes. Por ejemplo, una marca en el flujo de bits codificado puede señalar si los siguientes bits han de ser precedidos de un algoritmo de predicción antes de ser decodificados con un algoritmo de transformada de coseno discreto (DCT). Los algoritmos que comprenden el proceso de decodificación están regulados por la semántica definida por estos estándares MPEG. Esta sintaxis se puede aplicar para explotar características de video comunes tales como redundancia especial, redundancia temporal, movimiento uniforme, enmascaramiento espacial, etc. Un decodificador MPEG debe ser capaz de analizar sintácticamente y decodificar un flujo de datos entrantes, pero siempre que el flujo de datos cumpla con la sintaxis MPEG correspondiente, se puede usar una amplia variedad de posibles estructuras de datos y técnicas de compresión (aunque técnicamente esto se desvía del estándar dado que la semántica no es conforme). También es posible llevar la semántica necesaria dentro de una sintaxis alternativa.

Estos estándares MPEG usan una variedad de métodos de compresión, incluyendo métodos intracuadro e intercuadro. En la mayoría de las escenas de video, el fondo permanece relativamente estable mientras que la acción tiene lugar en primer plano. El fondo puede moverse, pero gran parte de la escena a menudo es redundante. Estos estándares MPEG comienzan la compresión creando un cuadro de referencia denominado "intra" cuadro o cuadro "I". Los cuadros I se comprimen sin referencia a otros cuadros y, de este modo, contienen un cuadro entero de información de video. Los cuadros I proporcionan puntos de entrada en un flujo de bits de datos para acceso aleatorio, pero solamente se pueden comprimir moderadamente. Típicamente, los datos que representan cuadros I se colocan en el flujo de bits cada 12 a 15 cuadros (aunque también es útil en algunas circunstancias usar una separación mucho más amplia entre cuadros I). A partir de e ntonces, dado que solamente una parte pequeña de los cuadros que caen entre los cuadros I de referencia son diferentes de los cuadros I que delimitan, solamente se capturan, comprimen y almacenan las diferencias de imagen. Se usan dos tipos de cuadros para tales diferencias, cuadros predichos (cuadros P) y cuadros predichos (o interpolados) bidireccionales (cuadros B).

Los cuadros P generalmente se codifican con referencia a un cuadro pasado (o bien un cuadro I o bien un cuadro P anterior), y, en general, se usan como referencia para cuadros P posteriores. Los cuadros P reciben una cantidad bastante alta de compresión. Los cuadros B proporcionan la cantidad más alta de compresión pero requieren tanto un cuadro de referencia pasado como uno futuro con el fin de poder ser codificados. Los cuadros P e I son "cuadros referenciables" debido a que se pueden referenciar por cuadros P o B.

Los macrobloques son regiones de píxeles de imagen. Para MPEG-2, un macrobloque es una agrupación de píxeles de 16x16 de cuatro bloques DCT de 8x8, junto con un vector de movimiento para cuadros P, y uno o dos vectores de movimiento para cuadros B. Los macrobloques dentro de los cuadros P se pueden codificar individualmente usando o bien codificación intracuadro o bien intercuadro (predicha). Los macrobloques dentro de los cuadros B se pueden codificar individualmente usando codificación intracuadro, codificación predicha hacia delante, codificación predicha hacia atrás o tanto codificación predicha hacia delante como hacia atrás (es decir, interpolada bidireccionalmente). Se usa una estructura ligeramente diferente pero similar en codificación de video MPEG-4.

Después de la codificación, un flujo de bits de datos MPEG comprende una secuencia de cuadros I, P y B. Una secuencia puede consistir en casi cualquier patrón de cuadros I, P y B (hay unas pocas restricciones semánticas menores en su colocación). No obstante, es común en la práctica industrial tener un patrón de cuadros fijo (por ejemplo, IBBPBBPBBPBBPBB).

Predicción de vector de movimiento

En MPEG-2 y MPEG-4 (y estándares similares, tales como H.263), el uso de cuadros de tipo B (predichos bidireccionalmente) ha resultado beneficiar la eficiencia de la compresión. Los vectores de movimiento para cada macrobloque de tales cuadros se pueden predecir mediante uno cualquiera de los tres siguientes métodos:

Modo 1: Predicho hacia delante desde el cuadro I o P anterior (es decir, un cuadro predicho no bidireccionalmente).

Modo 2: Predicho hacia atrás desde el cuadro I o P posterior.

Modo 3: Predicho bidireccionalmente tanto desde el cuadro I o P posterior como anterior.

El Modo 1 es idéntico al método de predicción hacia delante usado para cuadros P. El Modo 2 es el mismo concepto, excepto que trabaja hacia atrás desde un cuadro posterior. El Modo 3 es un modo interpolativo que combina información tanto de cuadros anteriores como posteriores.

Además de estos tres modos, MPEG-4 también soporta un segundo modo de predicción de vector de movimiento interpolativo para cuadros B: predicción en modo directo usando el vector de movimiento del cuadro P posterior, más un valor delta (si el vector de movimiento del macrobloque P situado conjuntamente se divide en modo 8x8, dando como resultado cuatro vectores de movimiento para el macrobloque de 16x16, entonces la delta se aplica a todos los cuatro vectores de movimiento independientes en el cuadro B). El vector de movimiento del cuadro P posterior apunta al cuadro P o I anterior. Se usa una proporción para ponderar el vector de movimiento del cuadro P posterior. La proporción es la posición de tiempo relativa del cuadro B actual con respecto a los cuadros P posterior y P (o I) anterior.

La FIG. 1 es una línea de tiempo de cuadros y vectores de movimiento en modo directo MPEG-4 según la técnica anterior. El concepto de modo directo MPEG-4 (modo 4) es que el movimiento de un macrobloque en cada cuadro B intermedio probablemente esté cerca del movimiento que se usó para codificar la misma ubicación en el siguiente cuadro P. Una delta se usa para hacer correcciones menores a un vector de movimiento proporcional derivado del vector de movimiento (MV) 103 correspondiente para el cuadro P posterior. Mostrada en la FIG. 1 está la ponderación proporcional dada a los vectores de movimiento 101, 102 para cada cuadro B intermedio 104a, 104b como una función de la “distancia de tiempo” entre el cuadro P o I anterior 105 y el siguiente cuadro P 106. El vector de movimiento 101, 102 asignado a un cuadro B intermedio 104a, 104b correspondiente es igual al valor de ponderación asignado (1/3 y 2/3, respectivamente) de veces el vector de movimiento 103 para el siguiente cuadro P, más el valor delta.

Con MPEG-2, todos los modos de predicción para cuadros B se prueban en la codificación, y se comparan para encontrar la mejor predicción para cada macrobloque. Si ninguna predicción es buena, entonces el macrobloque se codifica autónomo como un macrobloque “I” (para “intra”). El modo de codificación se selecciona como el mejor modo entre hacia delante (modo 1), hacia atrás (modo 2) y bidireccional (modo 3), o intracodificación. Con MPEG-4, no se permite la opción de intracodificación. En su lugar, el modo directo llega a ser la cuarta opción. De nuevo, se elige el mejor modo de codificación, en base a algunos criterios de mejor coincidencia. En los codificadores de software MPEG-2 y MPEG-4 de referencia, la mejor coincidencia se determina usando una coincidencia de DC (Suma de Diferencia Absoluta o “SAD”).

El número de cuadros B sucesivos en un flujo de bits de datos codificados se determina por el valor del parámetro “M” en MPEG. M menos uno es el número de cuadros B entre cada cuadro P y el siguiente P (o I). De este modo, para M=3, hay dos cuadros B entre cada cuadro P (o I), como se ilustra en la FIG. 1. La limitación principal en restringir el valor de M, y por lo tanto el número de cuadros B secuenciales, es que la cantidad de cambio de movimiento entre cuadros P (o I) llega a ser grande. Números altos de cuadros B significa cantidades más largas de tiempo entre cuadros P (o I). De este modo, la eficiencia y las limitaciones del rango de codificación de los vectores de movimiento crean el límite final en el número de cuadros B intermedios.

También es significativo señalar que los cuadros P llevan “energía de cambio” hacia delante con el flujo de imágenes en movimiento, dado que cada cuadro P decodificado se usa como punto de partida para predecir el siguiente cuadro P posterior. Los cuadros B, no obstante, se descartan después de su uso. De este modo, cualquier bit usado para crear cuadros B se usa solamente para ese cuadro, y no proporciona correcciones que ayuden a la decodificación de cuadros posteriores, a diferencia de los cuadros P.

Un documento de Flierl et al. “A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction”, Actas de la Conferencia de Compresión de Datos, Snowbird, Ut., EE.UU., 30 de marzo al 1 de abril de 1998, Los Alamitos, Ca., EE.UU., Soc. de Comput. del IEEE, EE.UU., 30 de marzo de 1998, páginas 239 a 248 (ISBN-978-0-8186-8406-7) describe predicción con compensación de movimiento para codificación de video en la que un cuadro predicho se puede codificar por referencia a dos o más cuadros referenciables anteriores en la secuencia, tales como en cuadros predichos bidireccionales (cuadros B). El documento generaliza esto para un marco de distorsión de tasa, y propone el uso de varios cuadros en la operación de predicción.

Los siguientes documentos adicionales fueron citados durante el examen:

BJONTEGAARD G. (ED.): “H.26L TEST MODEL LONG TERM NUMBER 5 (TML-5) DRAFT0”, SECTOR DE ESTANDARIZACIÓN DE TELECOMUNICACIONES UIT-T DE LA UIT, GINEBRA, CH, 11ª REUNIÓN, PORTLAND, OR, EE.UU., 22-25 DE AGOSTO DE 2000, 22 de agosto de 2000 (22-08-2000), páginas 1-31,

GRUPO DE ESTUDIO 16 DE LA ITU - GRUPO DE EXPERTOS DE CODIFICACIÓN DE VIDEO - ISO/IEC MPEG E ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 E ITU-T SG16 Q6), 11ª REUNIÓN, PORTLAND, OR, EE.UU., 22-25 DE AGOSTO 2000, Nº q15k44, 16 de agosto de 2000 (16-08-2000), páginas 1-2,

HANNUKSELA: "Generalized B/MH-Picture Averaging", GRUPO DE ESTUDIO 16 DE LA UIT - GRUPO DE EXPERTOS DE CODIFICACIÓN DE VIDEO –I SO/IEC MPEG E ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 E ITU-T SG16 Q6), 3ª REUNIÓN, FAIRFAX, VA, EE.UU., 6-10 DE MAYO DE 2002, Nº JVT-C077, 6 de mayo de 2002 (06-05-2002), páginas 1-8,

GISLE BJÒONTEGAARD ET AL: "H.26L Test Model Long Term Number 4 (TML-4)", 10. REUNIÓN DE VCEG; 16-05-2000 - 19-05-2000; OSAKA, JP; (GRUPO DE EXPERTOS DE CODIFICACIÓN DE VIDEO DE LA ITU-T SG.16), Nº q15j72d0, 16 de junio de 2000 (16-06-2000) ISSN: 0000-0464;

KIKUCHI Y.: "Improved multiframe motion compensation using frame interpolation", GRUPO DE ESTUDIO 16 DE LA UIT - GRUPO DE EXPERTOS DE CODIFICACIÓN DE VIDEO - ISO/IEC MPEG E ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 E ITU-T SG16 Q6), 2ª REUNIÓN, GINEBRA, CH, 29 DE ENERO-1 DE FEBRERO DE 2002, Nº JVT-B075, 29 de enero de 2002 (29-01-2002), páginas 1-8

Compendio

La invención se dirige a métodos y sistemas que se definen en las reivindicaciones 1, 3, 5 y 6. Las realizaciones preferidas se describen en las reivindicaciones dependientes.

Otros aspectos descritos incluyen métodos que abarcan:

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprende regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho se codifica por referencia a uno o más cuadros referenciables en orden de visualización, donde al menos uno de tales cuadros referenciables no es el cuadro referenciable anterior más cercano en orden de visualización al por lo menos un cuadro predicho.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprende macrobloques, en la que al menos un macrobloque dentro de al menos un cuadro predicho se codifica por interpolación de dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho bidireccional se codifica para incluir más de dos vectores de movimiento, cada vector de movimiento tal que hace referencia a una región de imagen correspondiente en al menos un cuadro referenciable.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho se codifica para incluir al menos dos vectores de movimiento, cada uno de tales vectores de movimiento que hace referencia a una región de imagen correspondiente en un cuadro referenciable, donde cada región de imagen tal de tal al menos un cuadro predicho se codifica por interpolación de dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho bidireccional se codifica como una ponderación desigual de regiones de imagen seleccionadas de dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho bidireccional se codifica por interpolación de dos o más cuadros referenciables, donde al menos uno de los dos o más cuadros referenciables está separado del cuadro predicho bidireccional por al menos un cuadro referenciable intermedio en orden de visualización, y donde tal al menos una región de imagen se codifica como una ponderación desigual de regiones de imagen seleccionadas de tales al menos dos o más cuadros referenciales.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho bidireccional se codifica por interpolación de dos o más cuadros referenciables, donde al menos uno de los dos o más cuadros referenciales está separado del cuadro predicho bidireccional por al menos un cuadro referenciable posterior intermedio en orden de visualización.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales que comprenden regiones de imagen, en la que al menos una región de imagen de al menos un cuadro predicho bidireccional se codifica como una ponderación desigual de regiones de imagen seleccionadas de dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros predichos y predichos bidireccionales, cada uno que comprende valores de píxeles dispuestos en macrobloques, en donde al menos un macrobloque dentro de un cuadro predicho bidireccional se determina usando predicción en modo directo en base a vectores de movimiento de dos o más cuadros predichos.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales, cada uno que comprende valores de píxeles dispuestos en macrobloques, en donde al menos un macrobloque dentro de un cuadro predicho bidireccional se determina usando predicción en modo directo en base a vectores de movimiento de uno o más cuadros predichos en orden de visualización, en donde al menos uno de tales uno o más cuadros predichos es anterior en orden de visualización al cuadro predicho bidireccional.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales, cada uno que comprende valores de píxeles dispuestos en macrobloques, en donde al menos un macrobloque dentro de un cuadro predicho bidireccional se determina usando predicción en modo directo en base a vectores de movimiento de uno o más cuadros predichos, en donde al menos uno de tales uno o más cuadros predichos es posterior en orden de visualización al cuadro predicho bidireccional y separado del cuadro predicho bidireccional por al menos un cuadro referenciable intermedio.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros que comprende una pluralidad de regiones de imagen que tienen un valor de DC, cada región de imagen tal que comprende píxeles cada uno que tiene un valor de píxel de AC, en donde al menos uno del valor de DC y los valores de píxeles de AC de al menos una región de imagen de al menos un cuadro se determina como una interpolación ponderada de los respectivos valores de DC y valores de píxeles de AC correspondientes de al menos otro cuadro.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprenden una pluralidad de regiones de imagen que tienen un valor DC, cada región de imagen tal que comprende píxeles, cada uno que tiene un valor de píxel de AC, en el que al menos uno del valor DC y de los valores de píxeles de AC de al menos una región de imagen de al menos un cuadro predicho se interpola de los respectivos valores de DC y los valores de píxeles de AC correspondientes de dos o más cuadros referenciables.

- Mejorar la calidad de imagen de una secuencia de dos o más cuadros intermedios predichos bidireccionales en un sistema de compresión de imágenes de video, cada cuadro que comprende una pluralidad de regiones de imagen que tienen un valor de DC, cada región de imagen tal que comprende píxeles cada uno que tiene un valor de píxel de AC, incluyendo al menos uno de los siguientes: determinar los valores de píxeles de AC de cada región de imagen de un cuadro intermedio predicho bidireccional como una primera proporción ponderada de valores de píxeles de AC correspondientes en cuadros referenciables que delimitan la secuencia de cuadros intermedios predichos bidireccionalmente; y determinar el valor de DC de cada región de imagen de tal cuadro intermedio predicho bidireccional como una segunda proporción ponderada de los valores de DC correspondientes en cuadros referenciables que delimitan la secuencia de cuadros intermedios predichos bidireccionales.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros que comprende una pluralidad de píxeles que tienen una representación inicial, en la que los píxeles de al menos un cuadro se interpolan de píxeles correspondientes de al menos otros dos cuadros, en donde tales píxeles correspondientes de los al menos otros dos cuadros se interpolan mientras que se transforman a una representación diferente, y los píxeles interpolados resultantes se transforman de nuevo a la representación inicial.

- En un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables y predichos bidireccionales, determinar dinámicamente un patrón de código de tales cuadros que tienen un número variable de cuadros predichos bidireccionales, incluyendo: seleccionar una secuencia inicial que comienza con un cuadro referenciable, que tiene al menos un cuadro predicho bidireccional inmediatamente posterior, y que termina en un cuadro referenciable; añadir un cuadro referenciable al final de la secuencia inicial para crear una secuencia de prueba; evaluar la secuencia de prueba frente a unos criterios de evaluación seleccionados; para cada paso satisfactorio de evaluación de la secuencia de prueba, insertar un cuadro bidireccional antes del cuadro referenciable añadido y repetir el paso de evaluación; y si la evaluación de la secuencia de prueba no es satisfactoria, entonces aceptar la secuencia de prueba anterior como patrón de código actual.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables separados por al menos un cuadro predicho bidireccional, en donde el número de tales cuadros predichos bidireccionales varía en tal secuencia, y en donde al menos una región de imagen de al menos un cuadro predicho bidireccional tal se determina usando una ponderación desigual de valores de píxeles correspondientes a al menos dos cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros codificados por un codificador para decodificar por un decodificador, en donde al menos una región de imagen de al menos un cuadro se basa en interpolaciones ponderadas de dos o más de otros cuadros, tales interpolaciones ponderadas que se basan en al menos un conjunto de ponderaciones disponibles para el codificador y un decodificador, en donde una designación de uno seleccionado de tal al menos un conjunto de ponderaciones se comunica a un decodificador desde el codificador para seleccionar una o más ponderaciones activas actualmente.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros codificados por un codificador para decodificar por un decodificador, en donde al menos una región de imagen de al menos un cuadro se basa en interpolaciones ponderadas de dos o más de otros cuadros, tales interpolaciones ponderadas que se basan en al menos un conjunto de ponderaciones, en donde al menos un conjunto de ponderaciones se descarga a un decodificador y, a partir de entonces, una designación para uno seleccionado de tal al menos un conjunto de ponderaciones se comunica a un decodificador desde el codificador para seleccionar una o más ponderaciones activas actualmente.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables codificados por un codificador para decodificar por un decodificador, en donde los cuadros predichos en la secuencia de cuadros referenciables se transmiten por el codificador al decodificador en un orden de entrega que difiere del orden de visualización de tales cuadros predichos después de la decodificación.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprende píxeles dispuestos en regiones de imagen, en que al menos una región de imagen de al menos un cuadro predicho se codifica por referencia a dos o más cuadros de referencia, en donde cada región de imagen tal se determina usando una ponderación desigual de valores de píxeles correspondientes a tales dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros predichos, predichos bidireccionales e intracuadros, cada uno que comprende regiones de imagen, en donde al menos un filtro seleccionado del conjunto de filtros de nitidez y suavizado se aplica a al menos una región de imagen de un cuadro predicho o predicho bidireccional durante la predicción compensada de vector de movimiento de tal cuadro predicho o predicho bidireccional.

Los detalles de una o más realizaciones de la invención se exponen en los dibujos que se acompañan y la descripción a continuación. Otras características, objetos y ventajas de la invención serán evidentes a partir de la descripción y los dibujos, y de las reivindicaciones.

Descripción de los dibujos

La FIG. 1 es una línea de tiempo de cuadros y vectores de movimiento en modo directo MPEG-4 según la técnica anterior.

La FIG. 2 es una línea de tiempo de cuadros y valores de ponderación de píxeles proporcionales.

La FIG. 3 es una línea de tiempo de cuadros y valores de ponderación de píxeles mezclados, proporcionales e iguales.

La FIG. 4 es un diagrama de flujo que muestra un ejemplo ilustrativo.

La FIG. 5 es un diagrama que muestra un ejemplo de múltiples referencias anteriores por un cuadro P actual a dos cuadros P anteriores, y a un cuadro I anterior.

La FIG. 6A es un diagrama de un patrón de codificación MPEG-2 típico de la técnica anterior, que muestra un número constante de cuadros B entre cuadros I y/o cuadros P que delimitan.

La Figura 6B es un diagrama de un patrón de codificación de video MPEG-4 teóricamente posible de la técnica anterior, que muestra un número variable de cuadros B entre cuadros I y/o cuadros P que delimitan, así como una distancia variable entre cuadros I.

La FIG. 7 es un diagrama de patrones de código.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros referenciables que comprende píxeles dispuestos en regiones de imagen, en que al menos una región de imagen de al menos un cuadro predicho se codifica por referencia a dos o más cuadros de referencia, en donde cada región de imagen tal se determina usando una ponderación desigual de valores de píxeles correspondientes a tales dos o más cuadros referenciables.

- Un sistema de compresión de imágenes de video que tiene una secuencia de cuadros predichos, predichos bidireccionales e intracuadros, cada uno que comprende regiones de imagen, en donde al menos un filtro seleccionado del conjunto de filtros de nitidez y suavizado se aplica a al menos una región de imagen de un cuadro predicho o predicho bidireccional durante la predicción compensada de vector de movimiento de tal cuadro predicho o predicho bidireccional.

Los detalles de una o más realizaciones de la invención se exponen en los dibujos que se acompañan y la descripción a continuación. Otras características, objetos y ventajas de la invención serán evidentes a partir de la descripción y los dibujos, y de las reivindicaciones.

Descripción de los dibujos

La FIG. 1 es una línea de tiempo de cuadros y vectores de movimiento en modo directo MPEG-4 según la técnica anterior.

La FIG. 2 es una línea de tiempo de cuadros y valores de ponderación de píxeles proporcionales según una realización de la invención.

La FIG. 3 es una línea de tiempo de cuadros y valores de ponderación de píxeles mezclados, proporcionales e iguales según una realización de la invención.

La FIG. 4 es un diagrama de flujo que muestra un ejemplo ilustrativo de esta realización de la invención como método que se puede implementar por ordenador.

La FIG. 5 es un diagrama que muestra un ejemplo de múltiples referencias anteriores por un cuadro P actual a dos cuadros P anteriores, y a un cuadro I anterior.

La FIG. 6A es un diagrama de un patrón de codificación MPEG-2 típico de la técnica anterior, que muestra un número constante de cuadros B entre cuadros I y/o cuadros P que delimitan.

La Figura 6B es un diagrama de un patrón de codificación de video MPEG-4 teóricamente posible de la técnica anterior, que muestra un número variable de cuadros B entre cuadros I y/o cuadros P que delimitan, así como una distancia variable entre cuadros I.

La FIG. 7 es un diagrama de patrones de código.

La FIG. 8 es un diagrama de flujo que muestra una realización de un método de interpolación con interpolación de DC que es distinta de interpolación de AC.

La FIG. 9 es un diagrama de flujo que muestra una realización de un método para interpolación de píxeles de luminancia usando una representación alternativa.

La FIG. 10 es un diagrama de flujo que muestra una realización de un método para interpolación de píxeles de croma usando una representación alternativa.

La FIG. 11 es un diagrama que muestra tamaños de regiones de vectores de movimiento únicos para cada uno de dos cuadros P.

La FIG. 12 es un diagrama que muestra una secuencia de cuadros P y B con ponderaciones de interpolación para los cuadros B determinada como una función de la distancia desde un cuadro P posterior a 2 distancias que hace referencia a un cuadro P posterior a 1 distancia.

La FIG. 13 es un diagrama que muestra una secuencia de cuadros P y B con ponderaciones de interpolación para los cuadros B determinada como una función de la distancia desde un cuadro P posterior a 1 distancia que hace referencia a un cuadro P anterior a 2 distancias.

La FIG. 14 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P posterior tiene múltiples vectores de movimiento que hacen referencia a cuadros P anteriores.

La FIG. 15 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P posterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior, y un siguiente cuadro P posterior más cercano tiene múltiples vectores de movimiento que hacen referencia a cuadros P anteriores.

La FIG. 16 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P anterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior.

La FIG. 17 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P anterior más cercano tiene dos vectores de movimiento que hacen referencia a cuadros P anteriores.

La FIG. 18 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P anterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior.

La FIG. 19 es una secuencia de cuadros que muestra el caso de tres cuadros P P1, P2 y P3, donde P3 usa una referencia interpolada con dos vectores de movimiento, uno para cada uno de P1 y P2.

La FIG. 20 es una secuencia de cuadros que muestra el caso de cuatro cuadros P P1, P2, P3 y P4, donde P4 usa una referencia interpolada con tres vectores de movimiento, uno para cada uno de P1, P2 y P3.

La FIG. 21 es un diagrama que muestra una secuencia de cuadros P y B en que diversos cuadros P tienen uno o más vectores de movimiento que hacen referencia a diversos cuadros P anteriores, y

Descripción detallada

Visión general

Es una práctica común usar un valor para M de 3, que proporciona dos cuadros B entre cada cuadro P (o I). No obstante, M=2, y M=4 o más altos, son todos útiles. Es de particular importancia señalar que el valor de M (el número de cuadros B más 1) también soporta una relación natural con la tasa de cuadros. A 24 cuadros por segundo (fps), la velocidad de las películas, la 1/24 de la segunda distancia de tiempo entre cuadros puede dar como resultado cambios sustanciales de imagen cuadro a cuadro. A 60 fps, 72 fps o a tasas de cuadros más altas, no obstante, la distancia de tiempo entre cuadros adyacentes llega a ser reducida correspondientemente. El resultado es que números más altos de cuadros B (es decir, valores más altos de M) llegan a ser útiles y beneficiosos en la eficacia de compresión a medida que se aumenta la tasa de cuadros.

Tanto la compresión de video MPEG-2 como MPEG-4 utilizan un método de interpolación demasiado simplificado.

Por ejemplo, para el modo 3, la predicción bidireccional para cada macrobloque de un cuadro es un promedio igual de los macrobloques de cuadros posteriores y anteriores, a medida que se desplazan por los dos vectores de movimiento correspondientes. Este promedio igual es apropiado para M=2 (es decir, cuadros B intermedios únicos), dado que el cuadro B será equidistante en el tiempo de los cuadros P (o I) anteriores y posteriores. No obstante, para todos los valores más altos de M, solamente los cuadros B centrados simétricamente (es decir, el cuadro intermedio si M=4, 6, 8, etc.) se predecirán de manera óptima usando una ponderación igual. De manera similar, en el modo directo 4 de MPEG-4, incluso aunque los vectores de movimiento se ponderen proporcionalmente, los valores de píxeles predichos para el cuadro B intermedio son una proporción igual de los píxeles correspondientes del cuadro P (o I) anterior y del P posterior.

De este modo, representa una mejora aplicar una ponderación proporcional apropiada, para M>2, a los valores de píxeles predichos para cada cuadro B. La ponderación proporcional para cada píxel en un cuadro B actual corresponde a la posición relativa del cuadro B actual con respecto a los cuadros P (o I) anteriores y posteriores. De este modo, si M=3, el primer cuadro B usaría 2/3 del valor de píxel correspondiente (vector de movimiento ajustado) del cuadro anterior, y 1/3 del valor de píxel correspondiente del cuadro posterior (vector de movimiento ajustado).

La FIG. 2 es una línea de tiempo de cuadros y valores de ponderación de píxeles proporcionales. Los valores de píxeles dentro de cada macrobloque de cada cuadro B intermedio 201a, 201b se ponderan como una función de la "distancia" entre el cuadro P o I anterior A y el siguiente cuadro P o I B, con mayor ponderación que se acuerda a cuadros I o P más cercanos. Es decir, cada valor de píxel de un cuadro B predicho bidireccionalmente es una combinación ponderada de los valores de píxeles correspondientes de los cuadros predichos no bidireccionalmente A y B que delimitan. En este ejemplo, para M=3, la ponderación para el primer cuadro B 201a es igual a 2/3A + 1/3B; la ponderación para el segundo cuadro B 201b es igual a 1/3A + 2/3B. También se muestra la ponderación promedio igual que se asignaría bajo sistemas MPEG convencionales; la ponderación MPEG-1, 2 y 4 para cada cuadro B 201a, 201b sería igual a (A + B)/2.

Aplicación a rango dinámico extendido y rango de contraste

Si M es mayor que 2, la ponderación proporcional de los valores de píxeles en cuadros B intermedios mejorará la efectividad de la codificación bidireccional (modo 3) y directa (modo 4 MPEG-4) en muchos casos. Los casos de ejemplo incluyen efectos comunes de edición de películas y videos tales como desvanecimientos y disoluciones cruzadas. Estos tipos de efectos de video son casos de codificación problemática tanto para MPEG-2 como para MPEG-4 debido al uso de un algoritmo de coincidencia de DC simple, y el uso común de M=3 (es decir, dos cuadros B intermedios), dando como resultado proporciones iguales para cuadros B. La codificación de tales casos se mejora usando interpolación de cuadros B proporcional.

La interpolación de cuadros B proporcional también tiene aplicación directa a la mejora de eficiencia de codificación para extender el rango dinámico y de contraste. Un caso común en la codificación de imágenes es un cambio en la iluminación. Esto ocurre cuando un objeto se mueve gradualmente hacia (o fuera de) la sombra (bordes de sombra suaves). Si se usa una representación de codificación logarítmica para el brillo (que se incorpora mediante luminancia Y logarítmica, por ejemplo), entonces un cambio en el brillo de la iluminación será un cambio de desplazamiento de DC. Si el brillo de la iluminación cae a la mitad, los valores de píxeles se reducirán todos en una cantidad igual. De este modo, para codificar este cambio, se debería encontrar una coincidencia de AC, y una diferencia de DC codificada aplicada a la región. Tal diferencia de DC que se codifica en un cuadro P también se debería aplicar proporcionalmente en cada cuadro B intermedio. (Véase la Solicitud de Patente de EE.UU. en tramitación Nº 09/905.039, titulada "Method and System for Improving Compressed Image Chroma Information", publicada como US 2003/0112863 A1, para información adicional sobre representaciones de codificación logarítmica).

Además de los cambios en la iluminación, los cambios en el contraste también se benefician de la interpolación de cuadros B proporcional. Por ejemplo, a medida que un avión se mueve hacia un espectador fuera de una nube o neblina, su contraste aumentará gradualmente. Este aumento de contraste se expresará como una amplitud aumentada en los coeficientes de AC de la DCT en los macrobloques codificados de cuadros P correspondientes. De nuevo, los cambios de contraste en los cuadros B intermedios se aproximarán más estrechamente mediante una interpolación proporcional, mejorando de este modo la eficiencia de codificación.

Las mejoras en el rango dinámico y la eficiencia de codificación de contraste usando interpolación de cuadros B proporcional llega a ser cada vez más significativas a medida que las tasas de cuadros llegan a ser más altas y a medida que se aumenta el valor de M.

Aplicación de valores de M altos a capas temporales

El uso de realizaciones de la invención permite un aumento en el valor de M y, por lo tanto, en el número de cuadros B entre cuadros P y/o I que delimitan, mientras que se mantiene u obtiene eficiencia de codificación. Tal uso beneficia a una serie de aplicaciones, que incluyen capas temporales. Por ejemplo, en nuestra Patente de EE.UU. Nº 5.988.863, titulada "Temporal and Resolution Layering for Advanced Television", se señaló que los cuadros B son un mecanismo adecuado para las tasas (cuadros) temporales en capas. La flexibilidad de tales tasas está relacionada con el número de cuadros B consecutivos disponibles. Por ejemplo, los cuadros B únicos (M=2) pueden soportar una capa temporal decodificada de 36 fps dentro de un flujo de 72 fps o una capa temporal decodificada de 30 fps dentro de un flujo de 60 fps. Los cuadros B triples (M=4) pueden soportar tanto capas temporales decodificadas de 36 fps como de 18 fps dentro de un flujo de 72 fps, y capas temporales decodificadas de 30 fps y 15 fps dentro de un flujo de 60 fps. El uso de M=10 dentro de un flujo de 120 fps puede soportar capas temporales decodificadas de 12 fps, 24 fps y 60 fps. M=4 también se puede usar con un flujo de 144 fps para proporcionar capas temporales decodificadas a 72 fps y 36 fps.

Como una mejora para tomar cada cuadro de orden N, se pueden decodificar y mezclar proporcionalmente múltiples cuadros a 120 fps o 72 fps, como se describe en la Solicitud de Patente de EE.UU. en tramitación Nº 09/545.233, titulada "Enhancements to Temporal and Resolution Layering" véase el documento WO 01/17787 A1, para mejorar las características de desenfoque de movimiento de los resultados de 24 fps.

Incluso se pueden sintetizar tasas de cuadros más altas utilizando los métodos descritos en la Solicitud de Patente de EE.UU. en tramitación Nº 09/435.277, titulada "System and Method for Motion Compensation and Frame Rate Conversion" (véase el documento WO 01/35657 A1).

Por ejemplo, se puede utilizar una cámara original de 72 fps, con conversión de tasa de cuadros con compensación de movimiento para crear una tasa de cuadros efectiva de 288 cuadros por segundo. Usando M=12, se pueden derivar tanto tasas de cuadro de 48 fps como de 24 fps, así como otras tasas útiles tales como 144 fps, 96 fps y 32 fps (y, por supuesto, los 72 fps originales). Las conversiones de tasa de cuadros que usan este método no necesitan ser múltiplos integrales. Por ejemplo, una tasa efectiva de 120 fps se puede crear a partir de una fuente de 72 fps, y entonces se puede usar como fuente tanto para tasas de 60 fps como de 24 fps (usando M=10).

De este modo, hay beneficios de capas temporales para optimizar el rendimiento de la interpolación de cuadros B. La interpolación de cuadros B proporcional descrita anteriormente hace que números de cuadros B consecutivos más altos funcionen de manera más eficiente, permitiendo por ello estos beneficios.

Proporciones de interpolación de cuadros B mezcladas

Una razón por la que se ha usado una ponderación promedio igual en sistemas convencionales como el predictor de modo de compensación de movimiento para valores de píxeles de cuadro B es que el cuadro P (o I) antes o después de un cuadro B particular puede ser ruidoso y, por lo tanto, representar una coincidencia imperfecta. La mezcla igual optimizará la reducción de ruido en el bloque de compensación de movimiento interpolado. Hay una diferencia residual que se codifica usando la función DCT cuantificada. Por supuesto, cuanto mejor sea la coincidencia de la proporción compensada de movimiento, menor será la diferencia de bits residuales que se requerirá, y mayor será la calidad de imagen resultante.

En los casos en que hay objetos que se mueven dentro y fuera de la sombra o la neblina, una proporción verdadera donde M>2 proporciona una mejor predicción. No obstante, cuando no están ocurriendo cambios de iluminación y contraste, la ponderación igual puede resultar ser un mejor predictor, dado que los errores de mover un macrobloque hacia delante a lo largo de un vector de movimiento se promediarán con los errores del bloque desplazado hacia atrás, reduciendo de este modo los errores en cada uno a la mitad Aun así, es más probable que los macrobloques de cuadros B más cercanos a un cuadro P (o I) se correlacionen más con ese cuadro que con un cuadro P (o I) más distante.

De este modo, es deseable en algunas circunstancias, tales como contraste regional o cambio de brillo, utilizar una proporción verdadera para la ponderación de píxeles de macrobloque de cuadros B (tanto para luminancia como para color), como se ha descrito anteriormente. En otras circunstancias, puede ser más óptimo utilizar proporciones iguales, como en MPEG-2 y MPEG-4.

De este modo, un aspecto ejemplar utiliza una mezcla de estas dos técnicas de proporción (proporción de promedio igual y distancia de cuadro) para interpolación de píxeles de cuadros B. Por ejemplo, en el caso M=3, 3/4 de las proporciones de 1/3 y 2/3 se pueden mezclar con 1/4 del promedio igual, dando como resultado que las dos proporciones sean 3/8 y 5/8. Esta técnica se puede generalizar usando un “factor de mezcla” F:

Ponderación=F * (Ponderación Proporcional de Distancia de Cuadros) + (1-F) * (Ponderación Promedio Igual)

El rango útil del factor de mezcla F es de 1, que indica interpolación puramente proporcional, a 0, que indica promedio igual puramente (también se puede usar la asignación inversa de valores).

La FIG. 3 es una línea de tiempo de cuadros y valores de ponderación de píxeles mezclados, proporcionales e iguales. Los valores de píxeles de cada macrobloque de cada cuadro B intermedio 301a, 301b se ponderan como una función de la “distancia de tiempo” entre el cuadro P o I anterior A y el siguiente cuadro P o I B, y como una función del promedio igual de A y B. En este ejemplo, para M=3 y un factor de mezcla F=3/4, la ponderación mezclada para el primer cuadro B 301a es igual a 5/8A + 3/8B (es decir, 3/4 de la ponderación proporcional de 2/3A + 1/3B, más 1/4 de la ponderación promedio igual de (A + B)/2). De manera similar, la ponderación para el segundo cuadro B 301b es igual a 3/8A + 5/8B.

El valor del factor de mezcla F se puede establecer en general para una codificación completa, o para cada grupo de imágenes (GOP), un rango de cuadros B, cada cuadro B o cada región dentro de un cuadro B (incluyendo, por ejemplo, tan finamente como para cada macrobloque o, en el caso del modo directo MPEG-4 usando un vector P en modo 8x8, incluso bloques de movimiento 8x8 individuales).

En interés de la economía de bits, y reflejando el hecho de que la proporción de mezcla no es normalmente lo suficientemente importante para ser transportada con cada macrobloque, el uso óptimo de la mezcla se debería relacionar con el tipo de imágenes que se comprimen. Por ejemplo, para imágenes que se desvanecen, se disuelven o donde la iluminación o el contraste general está cambiando gradualmente, un factor de mezcla F cercano o en 1 (es decir, seleccionar la interpolación proporcional) es generalmente el más óptimo. Para imágenes en movimiento sin tales cambios de iluminación o contraste, entonces los valores del factor de mezcla más bajo, tales como 2/3, 1/2 o 1/3, podrían formar la mejor opción, conservando por ello algunos de los beneficios de la interpolación proporcional, así como algunos de los beneficios de la interpolación promedio igual. Todos los valores de factor de mezcla dentro del rango de 0 a 1 serán útiles generalmente, con un valor particular dentro de este rango que resulta óptimo para cualquier cuadro B dado.

Para imágenes de amplio rango dinámico y amplio rango de contraste, el factor de mezcla se puede determinar regionalmente, dependiendo de las características de la región local. En general, no obstante, un amplio rango de luz y contraste se recomienda hacia valores de factor de mezcla que favorecen la interpolación puramente proporcional, en lugar de promedio igual.

Un factor de mezcla óptimo se determina de manera general empíricamente, aunque se puede usar la experiencia con tipos particulares de escenas para crear una tabla de factores de mezcla por tipo de escena. Por ejemplo, se puede usar una determinación de las características de cambio de imagen para seleccionar la proporción de mezcla para un cuadro o región. Alternativamente, los cuadros B se pueden codificar usando una serie de factores de mezcla candidatos (o bien para el cuadro entero, o regionalmente), con cada uno entonces que se evalúa para optimizar la calidad de imagen (determinada, por ejemplo, por la relación señal a ruido, o SNR, más alta) y para el recuento de bits más bajo. Estas evaluaciones candidatas entonces se pueden usar para seleccionar el mejor valor para la proporción de mezcla. También se puede usar una combinación tanto de características de cambio de imagen como de calidad/eficiencia codificada.

Los cuadros B cerca del medio de una secuencia de cuadros B, o que resultan de valores bajos de M, no se ven muy afectados por la interpolación proporcional, dado que las proporciones calculadas ya están cerca del promedio igual. No obstante, para valores más altos de M, las posiciones extremas de cuadros B se pueden ver significativamente afectadas por la elección del factor de mezcla. Señalar que el factor de mezcla puede ser diferente para estas posiciones extremas, utilizando más del promedio, que las posiciones más centrales, que obtienen poco o ningún beneficio de la desviación del promedio, dado que ya tienen altas proporciones de ambos

cuadros P (o I) vecinos. . Por ejemplo, si M=5, el primer y cuarto cuadros B podrían usar un factor de mezcla F que mezcla en más del promedio igual, pero el segundo y tercer cuadros B medios pueden usar las proporciones promedio iguales 2/5 y 3/5 estrictas. Si el factor de mezcla de proporción a promedio varía como una función de la posición de un cuadro B en una secuencia, el valor variable del factor de mezcla se puede transportar en el flujo de bits comprimido o como información lateral al decodificador.

Si se requiere un factor de mezcla general estático (debido a la falta de un método para transportar el valor), entonces el valor de 2/3 normalmente es casi óptimo, y se puede seleccionar como valor estático para la interpolación de cuadros B tanto en el codificador como en el decodificador. Por ejemplo, usando F=2/3 para el factor de mezcla, para M=3 las proporciones de cuadros sucesivos serán 7/18 (7/18 = 2/3 * 1/3 + 1/3 * 1/2) y 11/18 (11/18 = 2/3 * 2/3 + 1/3 * 1/2).

Interpolación linear

Los valores de píxeles de cuadros de video generalmente se almacenan en una representación particular que correlaciona la información de imagen original con valores numéricos. Tal correlación puede dar como resultado una representación lineal o no lineal. Por ejemplo, los valores de luminancia usados en la compresión no son lineales. El uso de diversas formas de representación no lineal incluye logarítmica, exponencial (a diversas potencias) y exponencial con una corrección de negro (comúnmente usada para señales de video).

En rangos dinámicos estrechos, o para interpolaciones de regiones cercanas, la representación no lineal es aceptable, dado que estas interpolaciones cercanas representan interpolaciones lineales por piezas. De este modo, pequeñas variaciones en el brillo se aproximan razonablemente por interpolación lineal. No obstante, para amplias variaciones en el brillo, tales como ocurren en un rango dinámico amplio e imágenes de rango de contraste amplio, el tratamiento de las señales no lineales como lineales será inexacto. Incluso para imágenes de rango de contraste normal, los desvanecimientos lineales y las disoluciones cruzadas se pueden degradar por una interpolación lineal. Algunos desvanecimientos y disoluciones cruzadas utilizan tasas de desvanecimiento y disolución no lineales, añadiendo complejidad adicional.

De este modo, una mejora adicional al uso de mezclas proporcionales, o incluso interpolaciones promedio proporcionales o iguales simples, es realizar tales interpolaciones en valores de píxeles representados en un espacio lineal, o en otros espacios no lineales optimizados que difieren de la representación de luminancia no lineal original.

Esto se puede lograr, por ejemplo, convirtiendo primero las dos señales de luminancia no lineales (de los cuadros P (o I) anteriores y posteriores a una representación lineal, o una representación no lineal diferente. Entonces, se aplica una mezcla proporcional, después de lo cual se aplica la conversión inversa, produciendo el resultado mezclado en la representación de luminancia no lineal original de la imagen. No obstante, la función de proporción se habrá realizado sobre una representación más óptima de las señales de luminancia.

También es útil aplicar de manera beneficiosa esta conversión lineal o no lineal a valores de color (croma), además de luminancia, cuando los colores se desvanecen o llegan a estar más saturados, como ocurre en los cambios de contraste asociados con las variaciones en la neblina y la nubosidad.

Realización de ejemplo

La FIG. 4 es un diagrama de flujo que muestra un ejemplo ilustrativo como método que se puede implementar por ordenador:

Paso 400: En un sistema de compresión de imágenes de video, para el modo directo e interpolativo para calcular cuadros B, determinar un valor de interpolación a aplicar a cada píxel de una secuencia de entrada de dos o más cuadros intermedios predichos bidireccionalmente usando uno de (1) la proporción de distancia de cuadros o (2) una mezcla de ponderación igual y la proporción de distancia de cuadros, derivada de al menos dos cuadros predichos no bidireccionalmente que delimitan tal entrada de secuencia desde una fuente (por ejemplo, un flujo de imagen de video).

Paso 401: Optimizar el valor de interpolación con respecto a una unidad de imagen (por ejemplo, un grupo de imágenes (GOP), una secuencia de cuadros, una escena, un cuadro, una región dentro de un cuadro, un macrobloque, un bloque DCT , o agrupación útil o selección de píxeles similares). El valor de interpolación se puede establecer estáticamente para toda la sesión de codificación, o dinámicamente para cada unidad de imagen.

Paso 402: Optimizar además el valor de interpolación con respecto al tipo de escena o la simplicidad de codificación. Por ejemplo, se puede establecer un valor de interpolación: estáticamente (tal como proporcional 2/3 y promedio igual 1/3); proporcionalmente para cuadros cerca del promedio igual, pero mezclados con el promedio igual cerca de los cuadros P (o I) adyacentes; dinámicamente en base a las características de escena generales, tales como desvanecimientos y disoluciones cruzadas; dinámicamente (y localmente) en base a las características de región de imagen local, tales como contraste local y rango dinámico local; o dinámicamente (y localmente) en base al rendimiento de codificación (tal como SNR codificada más alta) y bits codificados mínimos generados.

Paso 403: Transportar las cantidades de proporción apropiadas al decodificador, si no se determinan estáticamente.

Paso 404: Opcionalmente, convertir la información de luminancia (y, opcionalmente, croma) para cada cuadro en una representación lineal o no lineal alternativa, y transportar esta representación alternativa al decodificador, si no se determina estáticamente.

Paso 405: Determinar los valores de píxeles proporcionales usando el valor de interpolación determinado.

Paso 406: Si es necesario (debido al Paso 404), reconvertir a la representación original.

Referencia de cuadro P extendido

Como se ha señalado anteriormente, en los métodos de compresión MPEG-1, 2 y 4 de la técnica anterior, los cuadros P hacen referencia al cuadro P o I anterior, y los cuadros B hacen referencia a los cuadros P y/o I anteriores y posteriores más cercanos. La misma técnica se usa en los estándares de compresión DCT con compensación de movimiento H.261 y H.263, que abarcan técnicas de compresión de tasa de bits baja.

En el estándar H.263++ y H.26L en desarrollo, la referencia de cuadro B se extendió para apuntar a cuadros P o I que no estaban delimitando directamente un cuadro actual. Es decir, los macrobloques dentro de cuadros B podrían apuntar a un cuadro P o I antes del cuadro P anterior, o a un cuadro P o I después del cuadro P posterior. Con uno o más bits por macrobloque, la omisión del cuadro P anterior o posterior se puede señalar simplemente. Conceptualmente, el uso de cuadros P anteriores para referencia en cuadros B solamente requiere almacenamiento. Para el uso de codificación de baja tasa de bits de H.263++ o H.26L, esta es una cantidad pequeña de memoria adicional. Para referencia de cuadro P posterior, el orden de codificación de cuadros P se debe modificar con respecto a la codificación de cuadros B, de manera que los cuadros P futuros (o posiblemente cuadros I) se deben decodificar antes que los cuadros B intermedios. De este modo, el orden de codificación también es un problema para referencias de cuadros P posteriores.

Las distinciones principales entre los tipos de cuadros P y B son: (1) los cuadros B se pueden referenciar bidireccionalmente (hasta dos vectores de movimiento por macrobloque); (2) los cuadros B se descartan después de su uso (lo que también significa que se pueden omitir durante la decodificación para proporcionar capas temporales); y (3) los cuadros P se usan como “peldaños”, uno al siguiente, dado que cada cuadro P se debe decodificar para su uso como referencia para cada cuadro P posterior.

Como ejemplo los cuadros P (en oposición a los cuadros B) se decodifican con referencia a uno o más cuadros P o I anteriores (excluyendo el caso de cada cuadro P que hace referencia solamente al cuadro P o I anterior más cercano). De este modo, por ejemplo, se pueden usar dos o más vectores de movimiento por macrobloque para un cuadro P actual, todos apuntando hacia atrás en el tiempo (es decir, a uno o más cuadros decodificados anteriormente). Tales cuadros P aún mantienen un carácter de “peldaño”. La FIG. 5 es un diagrama que muestra un ejemplo de múltiples referencias anteriores por un cuadro P actual 500 a dos cuadros P anteriores 502, 504, y a un cuadro I anterior 506.

Además, es posible aplicar los conceptos de interpolación de macrobloques, como se ha descrito anteriormente, en tales referencias de cuadro P. De este modo, además de señalar referencias únicas a más de un cuadro P o I anterior, también es posible mezclar proporciones de múltiples cuadros P o I anteriores, usando un vector de movimiento para cada referencia de cuadro tal. Por ejemplo, la técnica descrita anteriormente de usar un modo de interpolación de cuadros B que tiene dos referencias de cuadro se puede aplicar para permitir que cualquier macrobloque en un cuadro P haga referencia a dos cuadros P anteriores o un cuadro P anterior y un cuadro I anterior, usando dos vectores de movimiento. Esta técnica interpola entre dos vectores de movimiento, pero no es bidireccional en el tiempo (como es el caso con la interpolación de cuadros B), dado que ambos vectores de movimiento apuntan hacia atrás en el tiempo. Los costes de memoria han disminuido hasta el punto en que mantener múltiples cuadros P o I anteriores en la memoria para tal referencia concurrente es bastante práctico.

Al aplicar tal interpolación de cuadros P, es constructivo seleccionar y señalar a un decodificador diversas proporciones útiles de los dos o más cuadros P anteriores (y, opcionalmente, un cuadro I anterior). En particular, una mezcla igual de cuadros es una de las proporciones de mezcla útiles. Por ejemplo, con dos cuadros P anteriores como referencias, se puede mezclar una cantidad igual de 1/2 de cada cuadro P. Para tres cuadros P anteriores, se podría usar una mezcla igual de 1/3.

Otra combinación útil de dos cuadros P es 2/3 del cuadro anterior más reciente, y 1/3 del cuadro anterior menos reciente. Para tres cuadros P anteriores, otra mezcla útil es 1/2 del cuadro anterior más reciente, 1/3 del siguiente cuadro anterior más reciente y 1/6 del cuadro anterior menos reciente.

En cualquier caso, se puede utilizar un conjunto simple de mezclas útiles de múltiples cuadros P anteriores (y, opcionalmente, un cuadro I) y señalarlos simplemente desde un codificador a un decodificador. Las proporciones de mezcla específicas utilizadas se pueden seleccionar tan a menudo como sea útil para optimizar la eficiencia de codificación para una unidad de imagen. Se puede seleccionar una serie de proporciones de mezcla usando un

número pequeño de bits, que se pueden transportar al decodificador siempre que sea adecuado para una unidad de imagen deseada.

Como ejemplo, también puede ser útil conmutar-seleccionar referencias de cuadro P único a partir del cuadro P (o I) anterior más reciente a un cuadro P (o I) anterior más "distante". De esta forma, los cuadros P utilizarían un único vector de movimiento por macrobloque (u, opcionalmente, por bloque de 8x8 en codificación de estilo MPEG-4), pero utilizarían uno o más bits para indicar que la referencia se refiere a un único cuadro anterior específico. Los macrobloques de cuadros P en este modo no serían interpolativos, sino que, en su lugar, harían referencia a un cuadro anterior seleccionado, siendo seleccionado de dos, tres o más opciones de cuadros P (o I) anteriores posibles para referencia. Por ejemplo, un código de 2 bits podría designar uno de hasta cuatro cuadros anteriores como el cuadro de referencia único de elección. Este código de 2 bits se podría cambiar en cualquier unidad de imagen conveniente.

Número adaptativo de cuadros B

Es típico en codificación MPEG usar un patrón fijo de tipos de cuadros I, P y B. El número de cuadros B entre cuadros P es típicamente una constante. Por ejemplo, es típico en codificación MPEG-2 usar dos cuadros B entre cuadros P (o I). La FIG. 6A es un diagrama de un patrón de codificación MPEG-2 típico de la técnica anterior, que muestra un número constante de cuadros B (es decir, dos) entre los cuadros I 600 y/o los cuadros P 602 que delimitan.

El estándar de codificación de video MPEG-4 permite conceptualmente un número variable de cuadros B entre los cuadros I y/o los cuadros P que delimitan, y una cantidad variable de distancia entre cuadros I. La FIG. 6B es un diagrama de un patrón de codificación de video MPEG-4 teóricamente posible de la técnica anterior, que muestra un número variable de cuadros B entre los cuadros I 600 y/o los cuadros P 602 que delimitan, así como una distancia variable entre cuadros I 600.

Esta estructura de codificación flexible se puede utilizar teóricamente para mejorar la eficiencia de codificación haciendo coincidir los tipos de codificación de cuadros B y P más efectivos con los cuadros de imágenes en movimiento. Aunque esta flexibilidad se ha permitido específicamente, se ha explorado muy poco y no se conoce ningún mecanismo para determinar realmente la colocación de los cuadros B y P en tal estructura flexible.

Los conceptos descritos en la presente memoria pueden aplicar a esta estructura de codificación flexible así como a los patrones de codificación fijos simples de uso común. Los cuadros B, de este modo, se pueden interpolar usando los métodos descritos anteriormente, mientras que los cuadros P pueden hacer referencia a más de un cuadro P o I anterior y se pueden interpolar según la presente descripción.

En particular, los macrobloques dentro de los cuadros B pueden utilizar mezclas proporcionales apropiadas para una estructura de codificación flexible tan eficazmente como con una estructura fija. Las mezclas proporcionales también se pueden utilizar cuando los cuadros B hacen referencia a cuadros P o I que están más lejos que los cuadros P o I más cercanos que delimitan.

De manera similar, los cuadros P pueden hacer referencia a más de un cuadro P o I anterior en esta estructura de codificación flexible tan eficazmente como con una estructura de patrón fija. Además, las proporciones de mezcla se pueden aplicar a macrobloques en tales cuadros P cuando hacen referencia a más de un cuadro P anterior (más, opcionalmente, un cuadro I).

(A) Determinación de la colocación en patrones de codificación flexibles

El siguiente método permite que un codificador optimice la eficiencia tanto del patrón de codificación de cuadros como de las proporciones de mezcla utilizadas. Para un rango seleccionado de cuadros, se puede intentar un número de patrones de codificación candidatos, para determinar un patrón óptimo o casi óptimo (en relación a criterios específicos). La FIG. 7 es un diagrama de patrones de código que se pueden examinar. Se selecciona arbitrariamente una secuencia inicial 700, que termina en un cuadro P o I, y se usa como base para añadir cuadros P y/o B adicionales, que entonces se evalúan (como se describe a continuación). En una realización, se añade un cuadro P a la secuencia inicial 700 para crear una primera secuencia de prueba 702 para su evaluación. Si la evaluación es satisfactoria, se inserta un cuadro B intermedio para crear una segunda secuencia de prueba 704. Para cada evaluación satisfactoria, se insertan cuadros B adicionales para crear secuencias de prueba cada vez más largas 706-712, hasta que los criterios de evaluación lleguen a ser insatisfactorios. En ese punto, se acepta la secuencia de codificación anterior. Este proceso se repite entonces, usando el cuadro P final para la secuencia de codificación anteriormente aceptada como el punto de partida para añadir un nuevo cuadro P y entonces insertar nuevos cuadros B.

Un patrón de codificación óptimo o casi optimo se puede seleccionar en base a diversos criterios de evaluación, que implican a menudo compromisos de diversas características de codificación, tales como calidad de imagen codificada frente al número de bits de codificación requeridos. Los criterios de evaluación comunes incluyen el menor número de bits usados (en una prueba de parámetros de cuantificación fija), o la mejor relación señal a ruido (en una prueba de tasa de bits fija), o una combinación de ambos.

También es común minimizar una suma de diferencia absoluta (SAD), que forma una medida de coincidencia de DC. Como se describe en la Patente de EE.UU. en tramitación Nº 09/904.192, titulada “Motion Estimation for Video Compression Systems”, (publicada como US 2003/0112873 A1), un criterio de coincidencia de AC también es una medida útil de la calidad de una coincidencia candidata en particular (la solicitud de patente también describe otras optimizaciones útiles). De este modo, los criterios de coincidencia de AC y DC, acumulados sobre las mejores coincidencias de todos los macrobloques, se pueden examinar para determinar la calidad de coincidencia general de cada patrón de codificación candidato. Esta técnica de coincidencia de AC/DC puede aumentar o sustituir las pruebas de relación señal a ruido (SNR) y de menos bits usados cuando se usa junto con una estimación del número de bits codificados para cada tipo de patrón de cuadro. Es típico codificar macrobloques dentro de cuadros B con un valor de parámetro de cuantificación (QP) más alto que para los cuadros P, afectando tanto a la calidad (medida a menudo como relación señal a ruido) como al número de bits usados dentro de los diversos patrones de codificación candidatos.

(B) Optimización de proporción de mezcla en patrones de codificación flexibles

Opcionalmente, para cada patrón candidato determinado según el método anterior, las proporciones de mezcla se pueden probar para su idoneidad (por ejemplo, proporciones de mezcla óptimas o casi óptimas) en relación con uno o más criterios. Esto se puede hacer, por ejemplo, probando la mejor calidad (SNR más baja) y/o eficiencia (menos bits usados). El uso de una o más referencias anteriores para cada macrobloque en cuadros P también se puede determinar de la misma forma, probando cada patrón de referencia candidato y proporción de mezcla, para determinar un conjunto de una o más referencias adecuadas.

Una vez que se ha seleccionado el patrón de codificación para este siguiente paso (Paso 700 en la Figura 7), entonces los pasos posteriores (Pasos 702-712) se pueden probar para diversos patrones de codificación candidatos. De esta forma, se puede determinar una codificación más eficiente de una secuencia de imágenes en movimiento. De este modo, la eficiencia se puede optimizar/mejorar como se describe en la subsección (A) anterior; se puede aplicar optimización de mezcla en cada paso de codificación probado.

Interpolación de DC frente a AC

En muchos casos de codificación de imágenes, tales como cuando se usa una representación logarítmica de cuadros de imagen, la interpolación de valores de píxeles de cuadro descrita anteriormente codificará de manera óptima los cambios en la iluminación. No obstante, en representaciones alternativas de “curva gamma” de vídeo, lineal y otras, a menudo resultará útil aplicar diferentes factores de mezcla de interpolación a los valores de DC que a los valores de AC de los píxeles. La FIG. 8 es un diagrama de flujo que muestra una realización de un método de interpolación con la interpolación de DC que es distinta de la interpolación de AC. Para una región de imagen seleccionada (normalmente un bloque DCT o macrobloque) de un primer y segundo cuadros de entrada 802, 802’, el valor de píxeles promedio para cada región tal se resta 804, 804’, separando por ello el valor de DC (es decir, el valor promedio de toda la región seleccionada) 806, 806’ de los valores de AC (es decir, los valores de píxeles señalados restantes) 808, 808’ en las regiones seleccionadas. Los respectivos valores de DC 806, 806’ entonces se pueden multiplicar por ponderaciones de interpolación 810, 810’ diferentes de las ponderaciones de interpolación 814, 814’ usadas para multiplicar los valores de píxeles de AC (señalados) 808, 808’. El valor de DC recién interpolado 812 y los valores de AC recién interpolados 816 entonces se pueden combinar 818, dando como resultado una nueva predicción 820 para la región seleccionada.

Como con los otros valores de interpolación descritos, las ponderaciones apropiadas se pueden señalar a un decodificador por unidad de imagen. Un número pequeño de bits puede seleccionarse entre una serie de valores de interpolación, así como seleccionar la interpolación independiente de los aspectos de AC frente a DC de los valores de píxeles.

Interpolación lineal y no lineal

La interpolación es un promedio ponderado lineal. Dado que la operación de interpolación es lineal, y dado que los valores de píxeles en cada cuadro de imagen a menudo se representan de una forma no lineal (tal como representaciones gamma de video o logarítmicas), llega a ser posible una optimización adicional del proceso de interpolación. Por ejemplo, la interpolación de píxeles para una secuencia de cuadros particular, así como la interpolación de valores de DC por separado de valores de AC, algunas veces será óptima o casi óptima con una representación lineal de píxeles. No obstante, para otras secuencias de cuadros, tal interpolación será óptima o casi óptima, si los píxeles se representan como valores logarítmicos o en otras representaciones de píxeles. Además, las representaciones óptimas o casi óptimas para interpolar los componentes de señal U y V (croma) pueden diferir de las representaciones óptimas o casi óptimas para la componente de señal Y (luminancia). Por lo tanto, es útil convertir una representación de píxeles a una representación alternativa como parte del procedimiento de interpolación.

La FIG. 9 es un diagrama de flujo que muestra una realización de un método para interpolación de píxeles de luminancia usando una representación alternativa. Comenzando con una región o bloque de píxeles de luminancia (Y) en una representación inicial (por ejemplo, gamma de vídeo o logarítmica) (Paso 900), los datos de píxeles se

transforman a una representación alternativa (por ejemplo, lineal, logarítmica, gamma de video) diferente de la representación inicial (Paso 902). La región o bloque de píxeles transformados se interpola entonces como se ha descrito anteriormente (Paso 904), y se transforma de nuevo a la representación inicial (Paso 906). El resultado son valores de luminancia de píxeles interpolados (Paso 908).

La FIG. 10 es un diagrama de flujo que muestra una realización de un método para interpolación de píxeles de croma usando una representación alternativa. Comenzando con una región o bloque de píxeles de croma (U, V) en una representación inicial (por ejemplo, gamma de vídeo o logarítmica) (Paso 1000), los datos de píxeles se transforman a una representación alternativa (por ejemplo, lineal, logarítmica, gamma de video) diferente de la representación inicial (Paso 1002). La región o bloque de píxeles transformados entonces se interpola como se ha descrito anteriormente (Paso 1006), y se transforma de nuevo a la representación inicial (Paso 1006). El resultado son valores de croma de píxeles interpolados (Paso 1008).

Las transformaciones entre representaciones se pueden realizar según las enseñanzas de la Solicitud de Patente de EE.UU. Nº 09/905.039, titulada “Method and System for Improving Compressed Image Chroma Information”, publicada como US 2003/0112863 A1. Señalar que la transformación de representación alternativa y su inversa a menudo se pueden realizar usando una tabla de búsqueda simple.

Como una variación, el espacio de representación alternativa (lineal o no lineal) para la interpolación de AC puede diferir del espacio de representación alternativa para la interpolación de DC.

Como con las ponderaciones de interpolación, la selección de qué representación de interpolación alternativa se ha de usar para cada una de las representaciones de píxeles de luminancia (Y) y croma (U y V) se puede señalar al decodificador usando un número pequeño de bits para cada unidad de imagen seleccionada.

Número de vectores de movimiento por macrobloque

En MPEG-2, se permite un vector de movimiento por macrobloque de 16x16 en cuadros P. En cuadros B, MPEG-2 permite un máximo de 2 vectores de movimiento por macrobloque de 16x16, correspondiente al modo interpolativo bidireccional. En codificación de video MPEG-4, se permiten hasta 4 vectores de movimiento por macrobloque de 16x16 en cuadros P, correspondientes a un vector de movimiento por bloque DCT de 8x8. En cuadros B MPEG-4, se permiten un máximo de dos vectores de movimiento para cada macrobloque de 16x16, cuando se usa el modo interpolativo. Una delta de vector de movimiento único en modo directo MPEG-4 puede dar como resultado cuatro vectores de movimiento “implícitos” independientes, si el macrobloque de cuadros P posteriores correspondiente se estableció en modo 8x8 teniendo cuatro vectores. Esto se logra añadiendo la delta del vector de movimiento transportada en un macrobloque de cuadros B 16x16 a cada uno de los cuatro vectores de movimiento independientes correspondientes del siguiente macrobloque de cuadros P, después de escalar la distancia en el tiempo (el cuadro B está más cerca en el tiempo que la referencia de cuadro P o I anterior del cuadro P).

Existe la opción de aumentar el número de vectores de movimiento por región de imagen, tal como un macrobloque. Por ejemplo, algunas veces resultará beneficioso tener más de dos vectores de movimiento por macrobloque de cuadros B. Éstos se pueden aplicar haciendo referencia a cuadros P o I adicionales y teniendo tres o más términos de interpolación en la suma ponderada. También se pueden aplicar vectores de movimiento adicionales para permitir vectores independientes para los bloques DCT de 8x8 del macrobloque de cuadros B. También, se pueden usar cuatro deltas independientes para extender el concepto de modo directo aplicando una delta separada a cada uno de los cuatro vectores de movimiento de regiones de 8x8 del cuadro P posterior.

Además, los cuadros P se pueden adaptar usando técnicas de implementación de cuadro B para hacer referencia a más de un cuadro anterior en un modo interpolativo, usando la técnica de dos términos de interpolación de cuadros B descrita anteriormente. Esta técnica se puede extender fácilmente a más de dos cuadros P o I anteriores, con una interpolación resultante que tiene tres o más términos en la suma ponderada.

Como con otros aspectos del método (por ejemplo, representación de píxeles y métodos de interpolación de DC frente a AC), se pueden comunicar sumas ponderadas particulares a un decodificador usando un número pequeño de bits por unidad de imagen.

Al aplicar este aspecto del método, la correspondencia entre los bloques DCT de 8x8 píxeles y el campo de vector de movimiento no necesita ser tan estricta como con MPEG-2 y MPEG-4. Por ejemplo, puede ser útil usar tamaños de regiones alternativos distintos de 16x16, 16x8 (usado solamente con entrelazado en MPEG-4) y 8x8 para vectores de movimiento. Tales alternativas podrían incluir cualquier número de tamaños de regiones útiles, tales como 4x8, 8x12, 8x16, 6x12, 2x8, 4x8, 24x8, 32x32, 24x24, 24x16, 8x24, 32x8, 32x4, etc. Usando un número pequeño de tales tamaños útiles, unos pocos bits pueden señalar a un decodificador la correspondencia entre los tamaños de regiones de vectores de movimiento y los tamaños de bloques DCT. En sistemas donde se usa un bloque DCT de 8x8 convencional, un conjunto simple de correspondencias con el campo de vector de movimiento es útil para simplificar el procesamiento durante la compensación de movimiento. En sistemas donde el tamaño del bloque DCT es diferente de 8x8, entonces se puede lograr una mayor flexibilidad al especificar el campo de vector de movimiento, como se describe en la Solicitud de Patente de EE.UU. en tramitación Nº 09/545.233, titulada “Enhanced Temporal and Resolution Layering in Advanced Television”, véase el documento WO 01/177871 A1.

Señalar que los límites de la región de vector de movimiento no necesitan corresponder con los límites de la región DCT. De hecho, a menudo es útil definir regiones de vector de movimiento de tal forma que un borde de la región de vector de movimiento caiga dentro de un bloque DCT (y no en su borde).

También se aplica el concepto de extender la flexibilidad del campo de vector de movimiento a la interpolación. Siempre que se especifique la correspondencia entre cada píxel y uno o más vectores de movimiento a uno o más cuadros de referencia, el método de interpolación descrito anteriormente se puede aplicar a la flexibilidad total de vectores de movimiento útiles usando toda la generalidad de este método. Incluso el tamaño de las regiones correspondientes a cada vector de movimiento puede diferir para cada referencia de cuadro anterior cuando se usan cuadros P, y cada referencia de cuadro anterior y futuro cuando se usan cuadros B. Si los tamaños de regiones para los vectores de movimiento difieren cuando se aplica el método de interpolación mejorado, entonces la interpolación refleja la región común de superposición. La región común de superposición para referencias de vector de movimiento se puede utilizar como la región sobre la cual se determina el término de DC cuando se interpolan por separado los valores de píxeles de DC y de AC.

La FIG. 11 es un diagrama que muestra tamaños de regiones de vectores de movimiento únicos 1100, 1102 para cada uno de los dos cuadros P 1104, 1106. Antes de calcular los valores de interpolación, se determina la unión 1108 de los tamaños de regiones de vectores de movimiento. La unión 1108 define todas las regiones que se consideran que tienen un vector de movimiento asignado.

De este modo, por ejemplo, al interpolar regiones DCT de 4x4 de un cuadro B 1112 hacia atrás con el cuadro P anterior 1104, una región de 4x4 1110 dentro de la unión 1108 usaría el vector de movimiento correspondiente a la región de 8x16 1114 en el cuadro P anterior. Si se predice hacia delante, la región 1110 dentro de la unión 1108 usaría el vector de movimiento correspondiente a la región de 4x16 1115 en el siguiente cuadro P. De manera similar, la interpolación de la región 116 dentro de la unión 1108 hacia atrás usaría el vector de movimiento correspondiente a la región de 8x16 1114, mientras que la predicción de la misma región hacia delante usaría el vector de movimiento correspondiente a la región de 12x16 1117.

En un método, se usan dos pasos para lograr la interpolación de vectores de movimiento generalizados (es decir, de tamaño no uniforme). El primer paso es determinar las regiones comunes de vector de movimiento, como se describe con respecto a la FIG. 11. Esto establece la correspondencia entre píxeles y vectores de movimiento (es decir, el número de vectores de movimiento por tamaño de región de píxeles especificado) para cada referencia de cuadro anterior o posterior. El segundo paso es utilizar el método de interpolación apropiado y los factores de interpolación activos para cada región de píxeles. Es una tarea del codificador asegurar que se especifiquen regiones de vector de movimiento y métodos de interpolación óptimos o casi óptimos, y que todos los píxeles tengan sus vectores y métodos de interpolación completamente especificados. Esto puede ser muy simple en el caso de un patrón fijo de vectores de movimiento (tales como un vector de movimiento para cada bloque de 32x8, especificado para un cuadro entero), con un único método de interpolación especificado (tal como una mezcla de proporción fija a cada distancia de cuadro referenciado, especificada para el cuadro entero). Este método también puede llegar a ser bastante complejo si se hacen cambios regionales en los tamaños de regiones de vectores de movimiento, y donde los tamaños de región difieren dependiendo de a qué cuadro anterior o posterior se hace referencia (por ejemplo, bloques de 8x8 para el cuadro anterior más cercano y bloques de 32x8 para el siguiente cuadro anterior más cercano). Además, el método de interpolación se puede especificar regionalmente dentro del cuadro.

Cuando se codifica, es el trabajo del codificador determinar el uso óptimo o casi óptimo de los bits para seleccionar entre formas y tamaños de regiones de vectores de movimiento, y para seleccionar el método de interpolación óptimo o casi optimo. También se requiere una determinación para especificar el número y la distancia de los cuadros referenciados. Estas especificaciones se pueden determinar mediante pruebas exhaustivas de una serie de tamaños de regiones de vectores de movimiento candidatos, cuadros candidatos para referencia y métodos de interpolación para cada región de vector de movimiento tal, hasta que se encuentra una codificación óptima o casi óptima. La optimización (en relación con un criterio seleccionado) se puede determinar encontrando la menor SNR después de codificar un bloque o el menor número de bits para un parámetro de cuantificación (QP) fijo después de codificar el bloque, o mediante la aplicación de otra medida adecuada.

Extensión de modo directo

El modo directo convencional, usado en macrobloques de cuadros B en MPEG-4, puede ser eficiente en codificación de vectores de movimiento, proporcionando los beneficios del modo de bloque de 8x8 con una delta común simple. El modo directo pondera cada vector de movimiento correspondiente del cuadro P posterior, que hace referencia al cuadro P anterior, en la ubicación del macrobloque correspondiente en base a la distancia en el tiempo. Por ejemplo, si M=3 (es decir, dos cuadros B intermedios), con interpolación lineal simple, el primer cuadro B usaría -2/3 de veces el vector de movimiento del cuadro P posterior para determinar un desplazamiento de píxeles con respecto a tal cuadro P, y 1/3 de veces el vector de movimiento del cuadro P posterior para determinar un desplazamiento de píxeles con respecto al cuadro P anterior. De manera similar, el segundo cuadro B usaría -1/3 de veces el mismo vector de movimiento de cuadro P para determinar un desplazamiento de píxeles con respecto a tal cuadro P, y 2/3 de veces el vector de movimiento de cuadro P posterior para determinar un desplazamiento de píxeles con respecto al cuadro P anterior. En modo directo, se añade una delta pequeña a cada vector de movimiento correspondiente.

Según la invención, este concepto se puede extender a referencias de cuadro B que apuntan a uno o más cuadros P a n distancias, que a su vez hacen referencia a uno o más cuadros P o cuadros I anteriores o posteriores, teniendo en cuenta la distancia de cuadro para determinar una fracción de escala de cuadro.

La FIG. 12 es un diagrama que muestra una secuencia de cuadros P y B con ponderaciones de interpolación para los cuadros B determinada como una función de la distancia desde un cuadro P posterior a 2 distancias que hace referencia a un cuadro P posterior a 1 distancia. En el ejemplo ilustrado, M=3, que indica dos cuadros B 1200, 1202 consecutivos entre los cuadros P 1204, 1206 que delimitan. En este ejemplo, cada macrobloque de ubicación conjunta en el siguiente cuadro P posterior más cercano 1208 (es decir, n=2) podría apuntar al cuadro P intermedio (es decir, el más cercano) 1204, y los dos primeros cuadros B 1200, 1202 pueden hacer referencia al siguiente cuadro P posterior más cercano 1208 en lugar de al cuadro P posterior más cercano 1204, como en MPEG convencional. De este modo, para el primer cuadro B 1200, la fracción de escala de cuadro 5/3 de veces el vector de movimiento mv del siguiente cuadro P posterior más cercano 1208 se usaría como un desplazamiento de píxeles con respecto al cuadro P 1208, y el segundo cuadro B 1202 usaría un desplazamiento de 4/3 de veces ese mismo vector de movimiento.

Si un cuadro P posterior más cercano referenciado por un cuadro B apunta al siguiente cuadro P anterior más cercano, entonces, de nuevo, la distancia de cuadro simple se puede usar para obtener la fracción de escala de cuadro adecuada a aplicar a los vectores de movimiento. La FIG. 13 es un diagrama que muestra una secuencia de cuadros P y B con ponderaciones de interpolación para los cuadros B determinada como una función de la distancia desde un cuadro P posterior a 1 distancia que hace referencia a un cuadro P anterior a 2 distancias. En el ejemplo ilustrado, M=3, y los cuadros B 1300, 1302 hacen referencia al cuadro P posterior más cercano 1304, que a su vez hace referencia al cuadro P a 2 distancias 1306. De este modo, para el primer cuadro B 1300, la fracción de desplazamiento de píxeles es la fracción de escala de cuadros 2/6 multiplicada por el vector de movimiento mv del cuadro P posterior más cercano 1304, y el segundo cuadro B 1302 tendría un desplazamiento de píxeles de la fracción de escala de cuadros 1/6 multiplicada por ese mismo vector de movimiento, dado que el vector de movimiento del cuadro P posterior más cercano 1304 apunta al cuadro P anterior a 2 distancias 1306, que está distante a 6 cuadros.

En general, en el caso de un cuadro B que hace referencia a un único cuadro P en modo directo, el método de distancia de cuadro establece el numerador de una fracción de escala de cuadro igual a la distancia de cuadro de ese cuadro B a su cuadro P referenciado, u “objetivo”, y establece el denominador igual a la distancia de cuadro del cuadro P objetivo a otro cuadro P referenciado por el cuadro P objetivo. El signo de la fracción de escala de cuadro es negativo para mediciones hechas de un cuadro B a un cuadro P posterior, y positivo para mediciones hechas de un cuadro B a un cuadro P anterior. Este método simple de aplicar una distancia de cuadro o la fracción de escala de cuadro a un vector de movimiento de cuadro P puede lograr una codificación efectiva en modo directo.

Además, es posible permitir que el modo directo se aplique a múltiples referencias de vectores de movimiento interpolados de un cuadro P. Por ejemplo, si un cuadro P se interpoló desde los cuadros P anteriores más cercanos y siguientes más cercanos, una referencia en modo directo permite una mezcla interpolada para cada macrobloque de cuadros B en modo directo de referencia múltiple. En general, los dos o más vectores de movimiento de un cuadro P pueden tener aplicada una fracción de escala de cuadro adecuada. Los dos o más vectores de movimiento modificados de distancia de cuadro entonces se pueden usar con las ponderaciones de interpolación correspondientes para cada cuadro B que hace referencia o que es el destino de ese cuadro P, como se describe a continuación, para generar compensación de movimiento de macrobloque de cuadros B interpolado.

La FIG. 14 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P posterior tiene múltiples vectores de movimiento que hacen referencia a cuadros P anteriores. En este ejemplo, un cuadro B 1400 hace referencia a un cuadro P posterior P3. Este cuadro P3 a su vez tiene dos vectores de movimiento, mv1 y mv2, que hacen referencia a los cuadros P anteriores P2, P1, correspondientes. En este ejemplo, cada macrobloque del cuadro B 1400 se puede interpolar en modo directo usando cualquiera de dos términos de ponderación o una combinación de tales términos de ponderación.

Cada macrobloque para el cuadro B 1400 se construiría como una mezcla de:

- píxeles correspondientes del cuadro P2 desplazados por la fracción de escala de cuadro 1/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional i) más píxeles correspondientes del cuadro P3 desplazado por la fracción de escala de cuadro -2/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional j); y

- píxeles correspondientes del cuadro P1 desplazados por la fracción de escala de cuadro 2/3 (4/6) de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional k) más píxeles correspondientes del cuadro P3 desplazados por la fracción de escala de cuadro -1/3 (-2/6) de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional l).

Como con todos los modos directos, se puede usar una delta de vector de movimiento con cada uno de mv1 y mv2.

Los macrobloques predichos en modo directo en cuadros B también pueden hacer referencia a múltiples cuadros P posteriores, usando la misma metodología de interpolación y aplicación de fracción de escala de cuadro de vector de movimiento que con múltiples cuadros P anteriores. La FIG. 15 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P posterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior, y siguiente un cuadro P más cercano tiene múltiples vectores de movimiento que hacen referencia a cuadros P anteriores. En este ejemplo, un cuadro B 1500 hace referencia a dos cuadros P posteriores P2, P3. El cuadro P3 tiene dos vectores de movimiento, mv1 y mv2, que hacen referencia a los cuadros P anteriores P2, P1. El cuadro P2 tiene un vector de movimiento, mv3, que hace referencia al cuadro P anterior P1. En este ejemplo, cada macrobloque del cuadro B 1500 se interpola en modo directo usando tres términos de ponderación. En este caso, las fracciones de escala de cuadro del vector de movimiento pueden ser mayores que 1 o menores que -1.

Las ponderaciones para esta forma de interpolación de macrobloques de cuadros B en modo directo pueden utilizar la generalidad completa de interpolación como se describe en la presente memoria. En particular, cada ponderación, o combinaciones de las ponderaciones, se pueden probar para un mejor rendimiento (por ejemplo, calidad frente a número de bits) para diversas unidades de imagen. La fracción de interpolación establecida para este modo directo mejorado se puede especificar a un decodificador con un número pequeño de bits por unidad de imagen.

Cada macrobloque para el cuadro B 1500 se construiría como una mezcla de:

- píxeles correspondientes del cuadro P3 desplazados por la fracción de escala de cuadro -5/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional i), más píxeles correspondientes del cuadro P2 desplazados por la fracción de escala de cuadro -2/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional j);

- píxeles correspondientes del cuadro P3 desplazados por la fracción de escala de cuadro -5/6 de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional k) más píxeles correspondientes del cuadro P1 desplazados por la fracción de escala de cuadro 1/6 de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional l); y

- píxeles correspondientes del cuadro P2 desplazados por la fracción de escala de cuadro -2/3 de mv3 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional m) más píxeles correspondientes del cuadro P1 desplazados por la fracción de escala dl cuadro 1/3 de mv3 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional n).

Como con todos los modos directos, se puede utilizar una delta de vector de movimiento con cada uno de mv1, mv2 y mv3.

Señalar que a menudo ocurre un modo de codificación directa particularmente beneficioso cuando el siguiente cuadro P posterior más cercano hace referencia a los cuadros P más cercanos que delimitan un cuadro B candidato.

La codificación de modo directo de cuadros B en MPEG-4 siempre usa los vectores de movimiento de cuadro P posterior como referencia. También es posible que un cuadro B haga referencia a los vectores de movimiento de los macrobloques situados conjuntamente del cuadro P anterior, lo que algunas veces resultará una elección beneficiosa de la referencia de codificación en modo directo. En este caso, las fracciones de escala de cuadro del vector de movimiento serán mayores que uno, cuando el siguiente cuadro P anterior más cercano se referencie por el vector de movimiento del cuadro P anterior más cercano. La FIG. 16 es un diagrama que muestra una secuencia de cuadros P y B en los que un cuadro P anterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior. En este ejemplo, un cuadro B 1600 hace referencia al cuadro P anterior a 1 distancia P2. El vector de movimiento mv del cuadro P2 hace referencia al siguiente cuadro P anterior P1 (a 2 distancias en relación con el cuadro B 1600). Se muestran las fracciones de escala de cuadro apropiadas.

Si el cuadro P anterior más cercano se interpola a partir de múltiples vectores y cuadros, entonces métodos similares a los descritos junto con la FIG. 14 se aplican para obtener las fracciones de escala de cuadro del vector de movimiento y las ponderaciones de interpolación. La FIG. 17 es un diagrama que muestra una secuencia de cuadros P y B en los que un cuadro P anterior más cercano tiene dos vectores de movimiento que hacen referencia a cuadros P anteriores. En este ejemplo, un cuadro B 1700 hace referencia al cuadro P anterior P3. Un vector de movimiento mv1 del cuadro P3 anterior hace referencia al siguiente cuadro P anterior P2, mientras que el segundo vector de movimiento mv2 hace referencia al cuadro P anterior a 2 distancias P1. Se muestran las fracciones de escala de cuadro apropiadas.

Cada macrobloque para el cuadro B 1700 se construiría como una mezcla de:

- píxeles correspondientes del cuadro P3 desplazados por la fracción de escala de cuadro 1/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional i) más píxeles correspondientes del cuadro P2 desplazados por la fracción de escala de cuadro 4/3 de mv1 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional j); y

- píxeles correspondientes del cuadro P3 desplazados por la fracción de escala de cuadro 1/6 de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional k) más píxeles correspondientes del cuadro P1 desplazados por la fracción de escala de cuadro 7/6 de mv2 (donde los píxeles entonces se pueden multiplicar por alguna ponderación proporcional l).

Cuando el vector de movimiento de un cuadro P anterior (en relación con un cuadro B) apunta al siguiente cuadro P anterior más cercano, no es necesario utilizar solamente el siguiente cuadro anterior más cercano como la referencia de interpolación, como en la FIG. 16. El cuadro P anterior más cercano puede resultar ser una mejor opción para compensación de movimiento. En este caso, el vector de movimiento del cuadro P anterior más cercano se acorta a la fracción de distancia de cuadro desde un cuadro B a ese cuadro P. La FIG. 18 es un diagrama que muestra una secuencia de cuadros P y B en que un cuadro P anterior más cercano tiene un vector de movimiento que hace referencia a un cuadro P anterior. En este ejemplo, para M=3, un primer cuadro B 1800 usaría fracciones de distancia de cuadro de 1/3 y -2/3 de veces el vector de movimiento mv del cuadro P anterior más cercano P2. El segundo cuadro B 1802 usaría fracciones de distancia de cuadro 2/3 y -1/3 (no se muestran). Tal selección se señalaría al decodificador para distinguir este caso del caso mostrado en la FIG. 16.

Como con todos los otros modos de codificación, el uso del modo directo implica preferiblemente probar el modo candidato frente a otros modos disponibles de interpolación y de codificación de un único vector y cuadros de referencia. Para la prueba en modo directo, el cuadro P posterior más cercano (y, opcionalmente, el siguiente cuadro P posterior más cercano o incluso los cuadros P posteriores más distantes, y/o uno o más cuadros P anteriores) se pueden probar como candidatos, y se puede usar un número pequeño de bits (generalmente uno o dos) para especificar la distancia o distancias de cuadro de referencia P en modo directo a ser usadas por un decodificador.

Valores de interpolación extendidos

Se especifica en MPEG-1, 2 y 4, así como en los estándares H.261 y H.263, que los cuadros B usan una ponderación igual de valores de píxeles de los cuadros referenciados hacia delante y referenciados hacia atrás, como desplazados por los vectores de movimiento. Esta invención incluye la aplicación de diversas ponderaciones desiguales útiles que pueden mejorar significativamente la eficiencia de codificación de cuadro B, así como la extensión de tales ponderaciones desiguales a más de dos referencias, incluyendo dos o más referencias hacia atrás o hacia delante en el tiempo.

La invención también incluye métodos para que más de un cuadro sea referenciado e interpolado para cuadros P. Además, cuando dos o más referencias apuntan hacia delante en el tiempo, o cuando dos o más referencias apuntan hacia atrás en el tiempo, algunas veces será útil usar ponderaciones negativas, así como ponderaciones superiores a 1,0.

Por ejemplo, la FIG. 19 es una secuencia de cuadros que muestra el caso de tres cuadros P P1, P2 y P3, donde P3 usa una referencia interpolada con dos vectores de movimiento, uno para cada uno de P1 y P2. Si, por ejemplo, está ocurriendo un cambio continuo sobre el lapso de cuadros entre P1 y P3, entonces P2-P1 (es decir, los valores de píxeles del cuadro P2, desplazados por el vector de movimiento para P2, menos los valores de píxeles del cuadro P1, desplazados por el vector de movimiento para P1) igualarán a P3-P2. De manera similar, P3-P1 será el doble de la magnitud de P2-P1 y P3-P2. En tal caso, los valores de píxeles para el cuadro P3 se pueden predecir de manera diferencial a partir de P1 y P2 a través de la fórmula:

$$P3 = P1 + 2 \times (P2 - P1) = (2 \times P2) - P1$$

En este caso, las ponderaciones interpolativas para P3 son 2,0 para P2 y -1,0 para P1.

Como otro ejemplo, la FIG. 20 es una secuencia de cuadros que muestra el caso de cuatro cuadros P P1, P2, P3 y P4, donde P4 usa una referencia interpolada con tres vectores de movimiento, uno para cada uno de P1, P2 y P3. De este modo, dado que P4 se predice a partir de P3, P2 y P1, se aplicarían tres vectores de movimiento y ponderaciones interpolativas. Si, en este caso, estuviera ocurriendo un cambio continuo sobre este lapso de cuadros, entonces P2- P1sería igual tanto a P3-P2 como a P4-P3, y P4-P1 sería igual tanto a 3 x (P2- P1) como a 3 x (P3-P2).

De este modo, en este caso de ejemplo, una predicción de P4 en base a P2 y P1 sería:

$$P4 = P1 + 3 \times (P2 - P1) = (3 \times P2) - (2 \times P1)$$ (ponderaciones 3,0 y – 2,0)

La predicción de P4 en base a P3 y P1 sería:

$$P4 = P1 + 3/2 \times (P3 - P1) = (3/2 \times P3) - (1/2 \times P1)$$ (ponderaciones 1,5 y -0,5)

La predicción de P4 en base a P3 y P2 sería:

$$P4 = P2 + 2 \times (P3 - P2) = (2 \times P3) - P2$$ (ponderaciones 2,0 y -1,0)

No obstante, también podría ser probable que el cambio más cercano a P4, que implica a P3 y P2, sea un predictor más fiable de P4 que las predicciones que implican a P1. De este modo, dando 1/4 de ponderación a cada uno de los dos términos anteriores que implican a P1, y 1/2 de ponderación al término que implica solamente a P3 y P2, daría como resultado:

1/2 (2 P3 – P2) + 1/4 (3/2 P3 – 1/2 P1) + 1/4 (3 P2 – 2 P1) =

1 3/8 P3 + 1/4 P2 – 5/8 P1 (ponderaciones 1,375, 0,25 y -0,625)

Por consiguiente, algunas veces será útil usar ponderaciones tanto por encima de 1,0 como por debajo de cero. En otras ocasiones, si hay una variación de tipo ruido de un cuadro al siguiente, un promedio ponderado positivo que tenga coeficientes leves entre 0,0 y 1,0 podría producir el mejor predictor del macrobloque de P4 (u otra región de píxeles). Por ejemplo, una ponderación igual de 1/3 de cada uno de P1, P2 y P3 en la FIG. 20 podría formar el mejor predictor de P4 en algunos casos.

Señalar que el vector de movimiento de la mejor coincidencia se aplica para determinar la región de P1, P2, P3, etc., que se utiliza por los cálculos en este ejemplo. Esta coincidencia podría ser mejor una coincidencia de AC en algunos casos, permitiendo que un término de DC variable se prediga a través de los coeficientes de AC. Alternativamente, si se usa una coincidencia de DC (tal como la Suma de Diferencia Absoluta), entonces a menudo se pueden predecir cambios en los coeficientes de AC. En otros casos, diversas formas de coincidencia de vectores de movimiento formarán una mejor predicción con diversas mezclas de ponderación. En general, el mejor predictor para un caso particular se determina empíricamente usando los métodos descritos en la presente memoria.

Estas técnicas también son aplicables a cuadros B que tienen dos o más vectores de movimiento que apuntan o bien hacia atrás o bien hacia delante en el tiempo. Cuando se apunta hacia delante en el tiempo, el patrón de coeficiente descrito anteriormente para los cuadros P se invierte para predecir con precisión hacia atrás al cuadro P actual. Es posible tener dos o más vectores de movimiento tanto en la dirección hacia delante como hacia atrás, prediciendo por ello en ambas direcciones concurrentemente. Se puede optimizar una mezcla ponderada adecuada de estas diversas predicciones seleccionando la ponderación de mezcla que mejor predice el macrobloque (u otra región de píxeles) de un cuadro B actual.

La FIG. 21 es un diagrama que muestra una secuencia de cuadros P y B en que diversos cuadros P tienen uno o más vectores de movimiento que hacen referencia a diversos cuadros P anteriores, y que muestran diferentes ponderaciones a-e asignadas a las respectivas referencias hacia delante y hacia atrás por un cuadro B particular. En este ejemplo, un cuadro B 2100 hace referencia a tres cuadros P anteriores y a dos cuadros P posteriores.

En el ejemplo ilustrado en la FIG. 21, el cuadro P5 se debe decodificar para que este ejemplo funcione. Algunas veces es útil ordenar cuadros en un flujo de bits en el orden necesario para la decodificación (“orden de entrega”), que no es necesariamente el orden de visualización (“orden de visualización”). Por ejemplo, en una secuencia de cuadros que muestra un movimiento cíclico (por ejemplo, rotación de un objeto), un cuadro P particular puede ser más similar a un cuadro P distante que al cuadro P posterior más cercano. La FIG. 22 es un diagrama que muestra una secuencia de cuadros P y B en la que el orden de entrega del flujo de bits de los cuadros P difiere del orden de visualización. En este ejemplo, el cuadro P3 es más similar al cuadro P5 que al cuadro P4. Por lo tanto, es útil entregar y decodificar P5 antes que P4, pero mostrar P4 antes que P5. Preferiblemente, cada cuadro P debería señalar al decodificador cuándo se puede descartar tal cuadro P (por ejemplo, una expiración de n cuadros en orden de flujo de bits o después del cuadro X en el orden de visualización).

Si las ponderaciones se seleccionan de un conjunto pequeño de opciones, entonces un número pequeño de bits puede señalar al decodificador qué ponderación se ha de usar. Como con todas las otras ponderaciones descritas en la presente memoria, esto se puede señalar a un decodificador una vez por unidad de imagen, o en cualquier otro punto del proceso de decodificación donde sea útil un cambio en las ponderaciones.

También es posible descargar nuevos conjuntos de ponderación. De esta forma, un número pequeño de conjuntos de ponderación pueden estar activos en un momento dado. Esto permite que un número pequeño de bits señale a un decodificador cuál de los conjuntos de ponderación activos se ha de usar en cualquier punto dado en el proceso de decodificación Para determinar conjuntos de ponderación adecuados, se puede probar un número grande de ponderaciones durante la codificación. Si se encuentra que un subconjunto pequeño proporciona alta eficiencia, entonces ese subconjunto se puede señalar a un decodificador para su uso. Un elemento particular del subconjunto se puede señalar, de este modo, al decodificador con sólo unos pocos bits. Por ejemplo, 10 bits pueden seleccionar 1 de 1024 elementos de subconjunto. Además, cuando se debería cambiar un subconjunto pequeño particular para mantener la eficiencia, se puede señalar un nuevo subconjunto al decodificador. De este modo, un codificador puede optimizar dinámicamente el número de bits requeridos para seleccionar entre los elementos del conjunto de ponderación frente al número de bits necesarios para actualizar los conjuntos de ponderación. Además, se puede usar un número pequeño de códigos cortos para señalar ponderaciones útiles comunes, tales como 1/2, 1/3, 1/4, etc. De esta forma, un número pequeño de bits se puede usar para señalar el conjunto de ponderaciones, tal como para una predicción de vector hacia delante K en un cuadro P (donde K = 1, 2, 3, ...), o una predicción de vector hacia delante K y vector hacia atrás L en un cuadro B (donde K y L se seleccionan de 0, 1, 2, 3, ...), o una predicción

de vector hacia delante K y de vector hacia atrás L en un cuadro P (donde K y L se seleccionan de 0, 1, 2, 3, ...), como una función del valor M actual (es decir, la posición relativa del cuadro B con respecto a los cuadros P (o I) vecinos).

La FIG. 23 es un diagrama que muestra una secuencia de cuadros P y B con ponderaciones asignadas. Un cuadro B 2300 tiene ponderaciones a-e, los valores de las cuales se asignan a partir de una tabla de conjuntos de ponderaciones de cuadro B 2302. Un cuadro P 2304 tiene ponderaciones m, n, los valores de las cuales se asignan a partir de una tabla de conjuntos de ponderaciones de cuadro P 2306. Algunas ponderaciones pueden ser estáticas (es decir, descargadas permanentemente al decodificador), y señaladas por un codificador. Otras ponderaciones se pueden descargar dinámicamente y entonces señalar.

Esta misma técnica se puede usar para actualizar dinámicamente conjuntos de ponderación para seleccionar interpolación de DC frente a interpolación de AC. Además, se pueden señalar valores de código que seleccionan interpolación normal (lineal) (de valores de píxeles normalmente representados en una representación no lineal) frente a interpolación lineal de valores convertidos (en una representación alternativa lineal o no lineal). De manera similar, tales valores de código pueden señalar qué interpolación tal aplicar a los valores de AC o DC o si dividir las partes de AC y DC de la predicción.

El subconjunto activo también se puede usar para minimizar el número de bits necesarios para seleccionar entre los conjuntos de coeficientes de ponderación actualmente en uso. Por ejemplo, si 1024 conjuntos de ponderación descargados se mantuvieron en un decodificador, tal vez 16 pudieran necesitar estar activos durante una parte particular de un cuadro. De este modo, seleccionando qué subconjunto de 16 (de entre 1024) conjuntos de ponderación han de estar activos, solamente necesitan ser usados 4 bits para seleccionar qué conjunto de ponderación de estos 16 está activo. Los subconjuntos también se pueden señalar usando códigos cortos para los subconjuntos más comunes, permitiendo de este modo que un número pequeño de bits se seleccione entre los subconjuntos usados comúnmente.

Suavidad y nitidez

Como con la separación simple de un componente de DC de las señales de AC a través de la resta del valor promedio, también son posibles otras operaciones de filtrado durante la predicción compensada de vector de movimiento. Por ejemplo, se pueden aplicar diversos filtros de paso alto, paso banda y paso bajo a una región de píxeles (tal como un macrobloque) para extraer diversas bandas de frecuencia. Estas bandas de frecuencia entonces se pueden modificar cuando se realiza compensación de movimiento. Por ejemplo, a menudo podría ser útil en una imagen en movimiento ruidosa filtrar las frecuencias más altas con el fin de suavizar (hacer menos nítido o desenfocar ligeramente) la imagen. Los píxeles de imagen más suaves, combinados con una matriz de inclinación más pronunciada para la cuantificación (una matriz de inclinación más pronunciada ignora el ruido de más alta frecuencia en el bloque actual), normalmente formarán un método de codificación más eficiente. Ya es posible señalar un cambio en la matriz de inclinación de cuantificación para cada unidad de imagen. También es posible descargar matrices de inclinación personalizadas para luminancia y croma. Señalar que la efectividad de la compensación de movimiento se puede mejorar si la matriz de inclinación se cambia o no. No obstante, a menudo será más efectivo cambiar tanto la matriz de inclinación como los parámetros de filtro que se aplican durante la compensación de movimiento.

Es una práctica común usar una resolución reducida para la codificación de croma junto con una matriz de inclinación específica de croma. No obstante, la resolución de la codificación de croma es estática en este caso (tal como la media resolución de codificación 4:2:0 vertical y horizontalmente, o la media resolución de codificación 4:2:2 solamente horizontalmente). La efectividad de codificación se puede aumentar aplicando un proceso de filtro dinámico durante la compensación de movimiento tanto a croma como a luminancia (independientemente o en tándem), seleccionados por unidad de imagen.

La Solicitud de Patente de EE.UU. Nº 09/545.233, titulada “Enhanced Temporal and Resolution Layering in Advanced Television” (referenciada anteriormente, véase el documento WO 01/17787 A1), describe el uso de filtros de desplazamiento mejorados que tienen lóbulos negativos (una función de sincronismo truncada). Estos filtros tienen la ventaja de que conservan la nitidez cuando se realiza la parte de píxel fraccional del desplazamiento del vector de movimiento. Tanto en el punto de desplazamiento de píxeles enteros como en los puntos fraccionales, algunos macrobloques (u otras regiones de imagen útiles) se desplazan de manera más óptima usando filtros que reducen o aumentan su nitidez. Por ejemplo, para un “enfoque de bastidor” (donde algunos objetos en el cuadro se desenfocan con el tiempo, y otras partes del cuadro se enfocan), la transición es uno de cambio tanto en la nitidez como en la suavidad. De este modo, un filtro de compensación de movimiento que puede tanto aumentar la nitidez en ciertas regiones en una imagen al tiempo que se reduce la nitidez en otras regiones puede mejorar la eficiencia de codificación. En particular, si una región de una imagen se desenfoca, puede ser beneficioso disminuir la nitidez, lo que suavizará la imagen (creando por ello potencialmente una mejor coincidencia) y disminuirá el grano y/o el ruido (mejorando por ello posiblemente la eficiencia de codificación). Si una región de la imagen se enfoca, puede ser beneficioso conservar la nitidez máxima, o incluso aumentar la nitidez usando valores de filtro de lóbulo negativo más grandes.

El filtrado de croma también puede beneficiarse del aumento y la disminución de la nitidez durante la codificación. Por ejemplo, muchos de los beneficios de eficiencia de codificación de la codificación 4:2:0 (croma de media resolución horizontal y verticalmente) se pueden lograr usando filtros de compensación de movimiento más suaves para croma mientras que se conserva la resolución completa en los canales U y/o V. Solamente cuando el detalle de color en los canales U y V es alto, será necesario seleccionar los filtros de desplazamiento más nítidos; los filtros más suaves serán más beneficiosos donde haya alto ruido de color o grano.

Además de los cambios en el enfoque, también es común tener un cambio en la dirección y la cantidad de desenfoque de movimiento de un cuadro al siguiente. A la tasa de cuadros de película en movimiento de 24 fps, incluso una simple escena de diálogo puede tener cambios significativos en el desenfoque de movimiento de un cuadro al siguiente. Por ejemplo, un labio superior podría desenfocarse en un cuadro y hacerse nítido en el siguiente, debido totalmente al movimiento del labio durante el tiempo de apertura del obturador de la cámara. Para tal desenfoque de movimiento, será beneficioso no solamente tener filtros de nitidez y suavizado (desenfoque) durante la compensación de movimiento, sino también tener un aspecto direccional para el enfoque y el suavizado. Por ejemplo, si se puede determinar una dirección de movimiento, se puede usar un suavizado o una nitidez a lo largo de esa dirección para corresponder al movimiento o la detención de un rasgo de la imagen. Los vectores de movimiento usados para la compensación de movimiento pueden proporcionar a ellos mismos alguna información útil acerca de la cantidad de movimiento y el cambio en la cantidad de movimiento (es decir, desenfoque de movimiento), para un cuadro particular (o región dentro de un cuadro) con respecto a cualquiera de los cuadros circundantes (o regiones correspondientes). En particular, un vector de movimiento es la mejor coincidencia de movimiento entre cuadros P, mientras que el desenfoque de movimiento resulta del movimiento durante el tiempo de apertura del obturador dentro de un cuadro.

La FIG. 24 es un gráfico de posición de un objeto dentro de un cuadro frente al tiempo. El obturador de una cámara está abierto solamente durante parte de un tiempo de cuadro. Cualquier movimiento del objeto mientras que el obturador está abierto da como resultado un desenfoque. La cantidad de desenfoque de movimiento se indica por la cantidad de cambio de posición durante el tiempo de apertura del obturador. De este modo, la pendiente de la curva de posición 2400 mientras que el obturador está abierto es una medición del desenfoque de movimiento.

La cantidad de desenfoque de movimiento y la dirección de movimiento también se pueden determinar a partir de una combinación de métricas de nitidez, vectores de movimiento circundantes (donde coinciden regiones de imagen), detección de borrosidad de rasgos y designación asistida por humanos de regiones de cuadro. Se puede seleccionar un filtro en base a la cantidad determinada de desenfoque de movimiento y de dirección de movimiento. Por ejemplo, se puede determinar empíricamente una correlación de diversos filtros frente al desenfoque de movimiento y la dirección determinados.

Cuando se combina con los otros rasgos del método, tales filtros aplicados de manera inteligente pueden mejorar significativamente la eficiencia de codificación de compresión. Se puede seleccionar un número pequeño de tales filtros con un número pequeño de bits señalados al decodificador. De nuevo, esto se puede hacer una vez por unidad de imagen o en otros puntos útiles en el proceso de decodificación. Como con los conjuntos de ponderación, se puede usar un conjunto de filtros cargados dinámicamente, así como un mecanismo de subconjuntos activos, para minimizar el número de bits necesarios a seleccionar entre el conjunto de parámetros de filtro más beneficiosos.

Implementación

El método de compresión de video se puede implementar en hardware o software, o una combinación de ambos (por ejemplo, matrices lógicas programables). A menos que se especifique de otro modo, los algoritmos incluidos como parte de la invención no están inherentemente relacionados con ningún ordenador u otro aparato particular. En particular, se pueden usar diversas máquinas de propósito general con programas escritos según las enseñanzas de la presente memoria, o puede ser más conveniente construir aparatos más especializados (por ejemplo, circuitos integrados) para realizar funciones particulares. De este modo, el método se puede implementar en uno o más programas de ordenador que se ejecutan en uno o más sistemas de ordenador programables, cada uno que comprende al menos un procesador, al menos un sistema de almacenamiento de datos (incluyendo memoria volátil y no volátil y/o elementos de almacenamiento), en al menos un dispositivo o puerto de entrada, y al menos un dispositivo o puerto de salida. El código de programa se aplica a los datos de entrada para realizar las funciones descritas en la presente memoria y generar información de salida. La información de salida se aplica a uno o más dispositivos de salida, de forma conocida.

Cada programa tal se puede implementar en cualquier lenguaje de ordenador deseado (incluyendo lenguajes de programación de máquina, ensamblaje o procedimental de alto nivel, lógico u orientado a objetos) para comunicar con un sistema de ordenador. En cualquier caso, el lenguaje puede ser un lenguaje compilado o interpretado.

Cada programa de ordenador tal se almacena o descarga preferiblemente a un medio o dispositivo de almacenamiento (por ejemplo, memoria o medios de estado sólido, o medios magnéticos u ópticos) legible por un ordenador programable de propósito general o especial, para configurar y operar el ordenador cuando el medio o dispositivo de almacenamiento se lee por el sistema de ordenador para realizar los procedimientos descritos en la

presente memoria. También se puede considerar que el sistema se implemente como un medio de almacenamiento legible por ordenador, configurado con un programa de ordenador, donde el medio de almacenamiento así configurado hace que un sistema de ordenador opere de una manera específica y predefinida para realizar las funciones descritas en la presente memoria.

Se han descrito una serie de realizaciones de la invención. Sin embargo, se entenderá que se pueden hacer diversas modificaciones sin apartarse del alcance de la invención como se define por las reivindicaciones. Por ejemplo, algunos de los pasos descritos anteriormente pueden ser independientes del orden, y de este modo se puede realizar en un orden diferente del descrito.

REIVINDICACIONES

1. Un método de compresión de imágenes de video que comprende:

proporcionar una secuencia de cuadros referenciables (I, P) y predichos bidireccionales (B) cada uno que comprende valores de píxeles dispuestos en macrobloques; y

determinar al menos un macrobloque dentro de un cuadro predicho bidireccional (B) usando predicción en modo directo en base a un vector de movimiento entre dos cuadros referenciables (I, P) siendo escalado por una fracción de escala de cuadro de más de uno, en donde los dos cuadros referenciables son anteriores en orden de visualización al cuadro predicho bidireccional.

2. Un método según la reivindicación 1, en donde los dos cuadros referenciables están en orden de visualización.

3. Un sistema de compresión de imágenes de video adaptado para proporcionar una secuencia de cuadros referenciales (I, P) y predichos bidireccionales (B) cada uno que comprende valores de píxeles dispuestos en macrobloques, en donde al menos un macrobloque dentro de un cuadro predicho bidireccional se determina usando predicción en modo directo en base a un vector de movimiento entre dos cuadros referenciables que se escalan por una fracción de escala de cuadro de más de uno, en donde los dos cuadros referenciables son anteriores en orden de visualización al cuadro predicho bidireccional.

4. Un sistema según la reivindicación 3, en donde los dos cuadros referenciables están en orden de visualización.

5. Un método de descompresión de imágenes de video que comprende: recibir una secuencia de cuadros referenciables (I, P) y predichos bidireccionales (B) cada uno que comprende valores de píxeles dispuestos en macrobloques; y

determinar al menos un macrobloque dentro de un cuadro predicho bidireccional (B) usando predicción en modo directo en base a un vector de movimiento entre dos cuadros referenciables (I, P) siendo escalado por una fracción de escala de cuadro de más de uno, en donde los dos cuadros referenciables son anteriores en orden de visualización al cuadro predicho bidireccional.

6. Un sistema de descompresión de video adaptado para

recibir una secuencia de cuadros referenciables (I, P) y predichos bidireccionales (B) cada uno que comprende valores de píxeles dispuestos en macrobloques; y

determinar al menos un macrobloque dentro de un cuadro predicho bidireccional (B) usando predicción en modo directo en base a un vector de movimiento entre dos cuadros referenciables (I, P) siendo escalado por una fracción de escala de cuadro de más de uno, en donde los dos cuadros referenciables son anteriores en orden de visualización al cuadro predicho bidireccional.

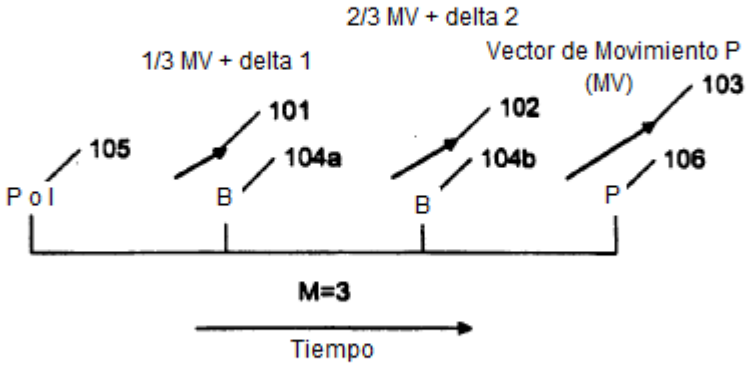

FIG. 1
Ponderación de Vector de Movimiento Proporcional
(Técnica Anterior)

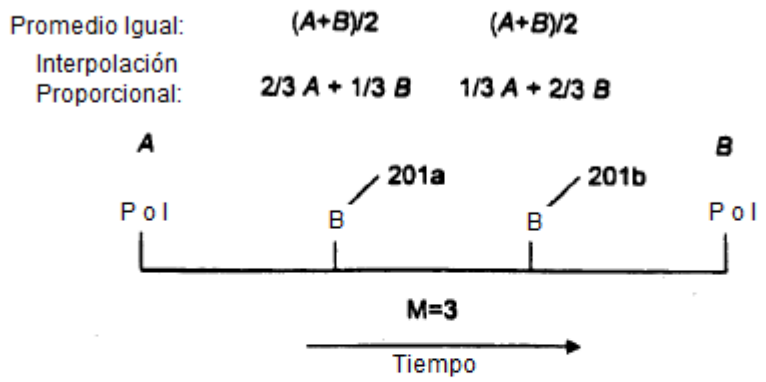

FIG. 2
Ponderación Proporcional de Valores de Píxeles

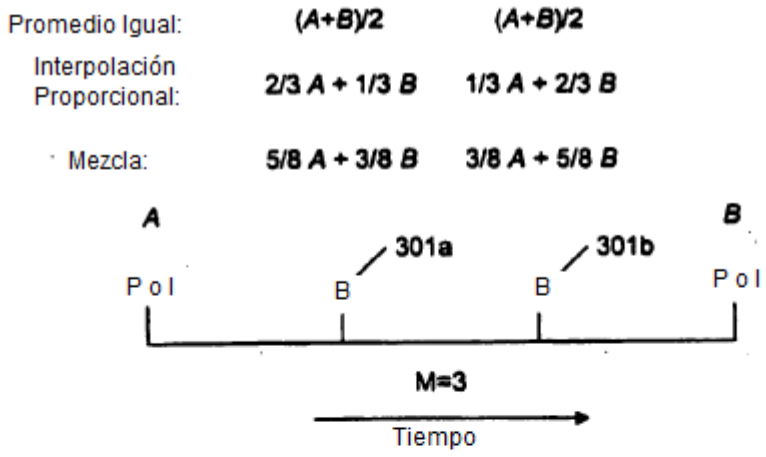

FIG. 3
Ponderación de Valores de Píxeles Mezclados

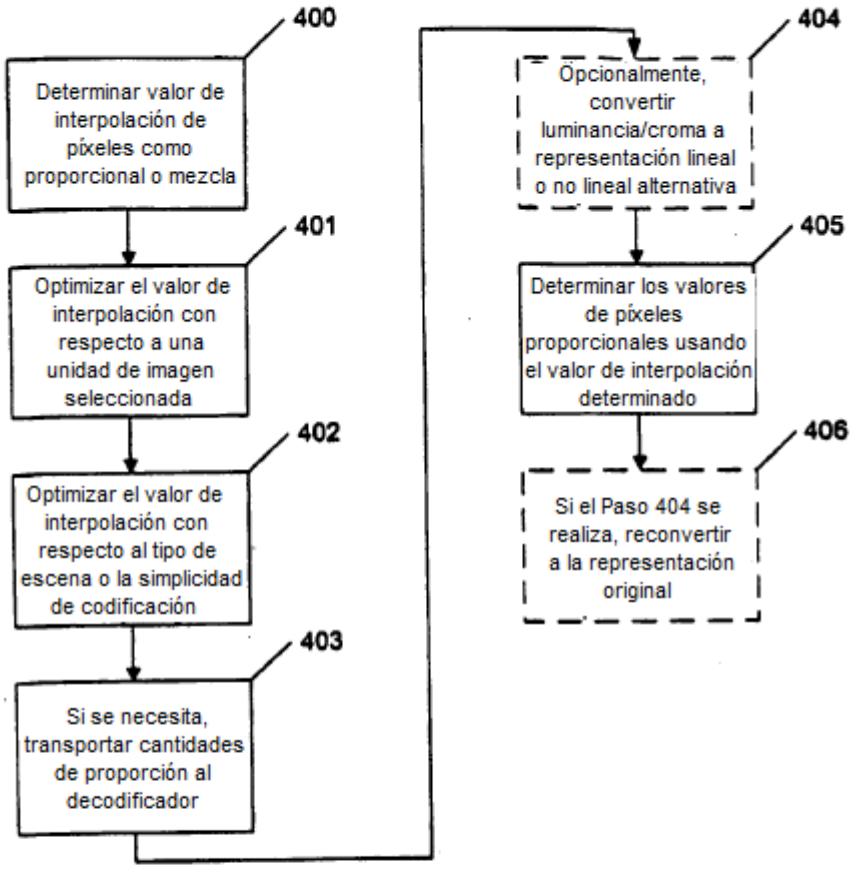
400
Determinar valor de interpolación de píxeles como proporcional o mezcla
401
Optimizar el valor de interpolación con respecto a una unidad de imagen seleccionada
402
Optimizar el valor de interpolación con respecto al tipo de escena o la simplicidad de codificación
403
Si se necesita, transportar cantidades de proporción al decodificador
404
Opcionalmente, convertir luminancia/croma a representación lineal o no lineal alternativa
405
Determinar los valores de píxeles proporcionales usando el valor de interpolación determinado
406
Si el Paso 404 se realiza, reconvertir a la representación original

FIG. 4

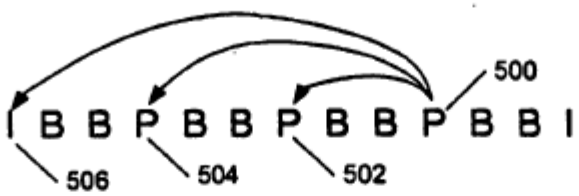

FIG. 5

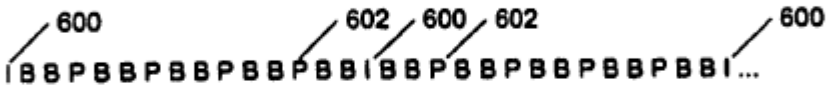

FIG. 6A
Técnica Anterior

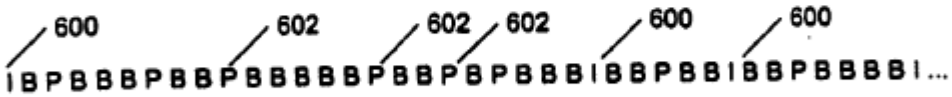

FIG. 6B
Técnica Anterior

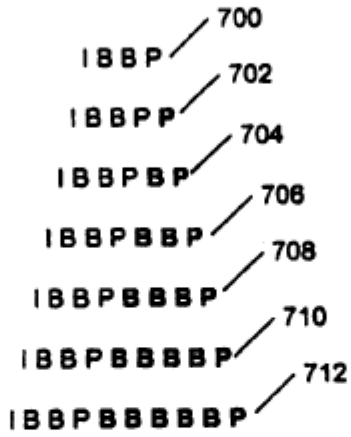
700
I B B P
702
I B B P P
704
I B B P B P
706
I B B P B B P
708
I B B P B B B P
710
I B B P B B B B P
712
I B B P B B B B B P

FIG. 7

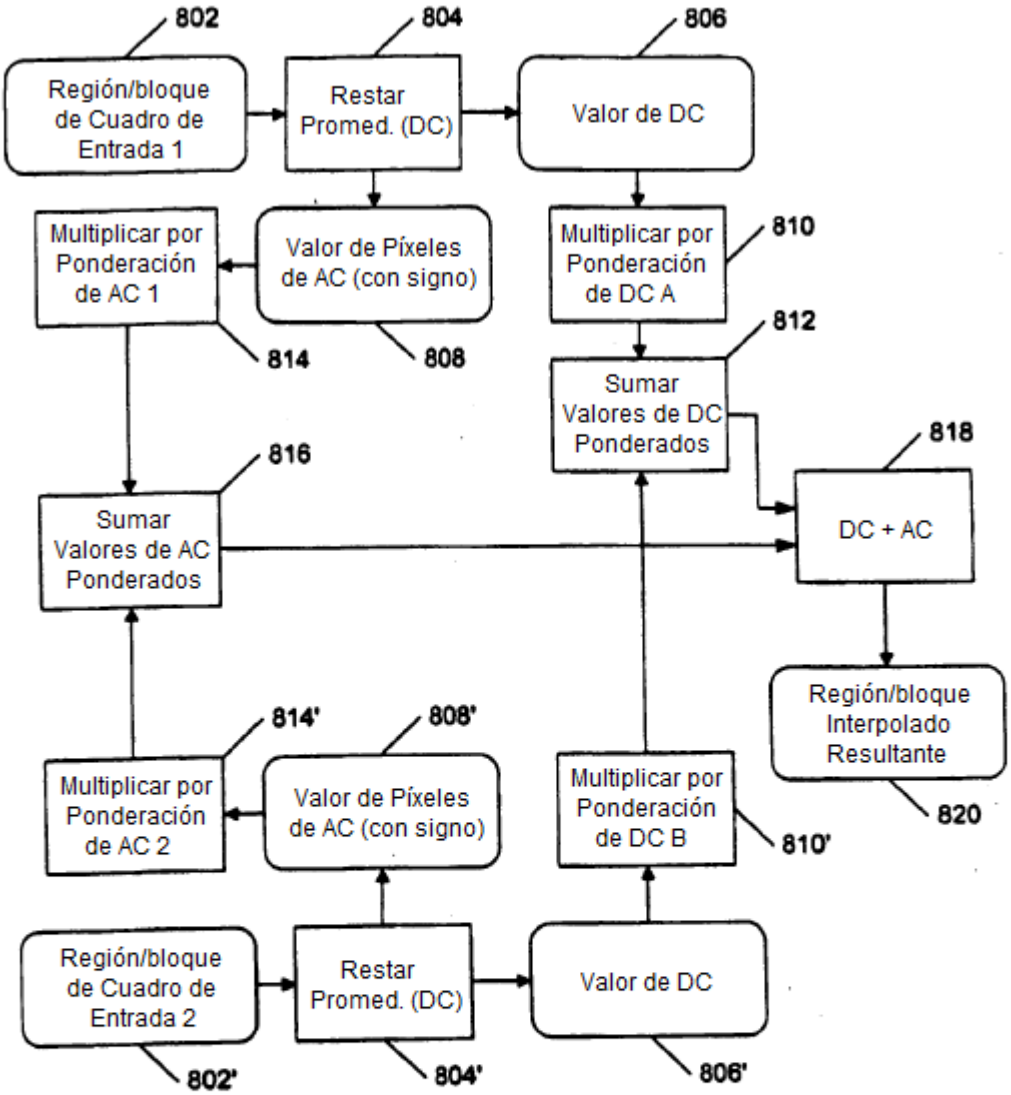
802
Región/bloque de Cuadro de Entrada 1
804
Restar Promed. (DC)
806
Valor de DC
Multiplicar por Ponderación de AC 1
Valor de Píxeles de AC (con signo)
Multiplicar por Ponderación de DC A
810
814
808
812
Sumar Valores de DC Ponderados
818
816
Sumar Valores de AC Ponderados
DC + AC
Región/bloque Interpolado Resultante
820
814'
808'
Multiplicar por Ponderación de AC 2
Valor de Píxeles de AC (con signo)
Multiplicar por Ponderación de DC B
810'
Región/bloque de Cuadro de Entrada 2
Restar Promed. (DC)
Valor de DC
802'
804'
806'

FIG. 8

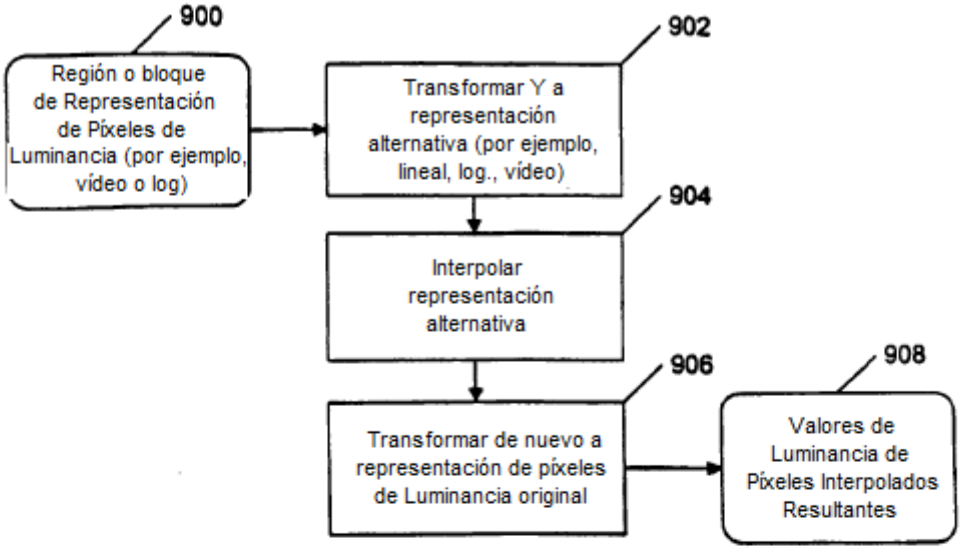
900
Región o bloque de Representación de Píxeles de Luminancia (por ejemplo, vídeo o log)
902
Transformar Y a representación alternativa (por ejemplo, lineal, log., vídeo)
904
Interpolar representación alternativa
906
Transformar de nuevo a representación de píxeles de Luminancia original
908
Valores de Luminancia de Píxeles Interpolados Resultantes

FIG. 9

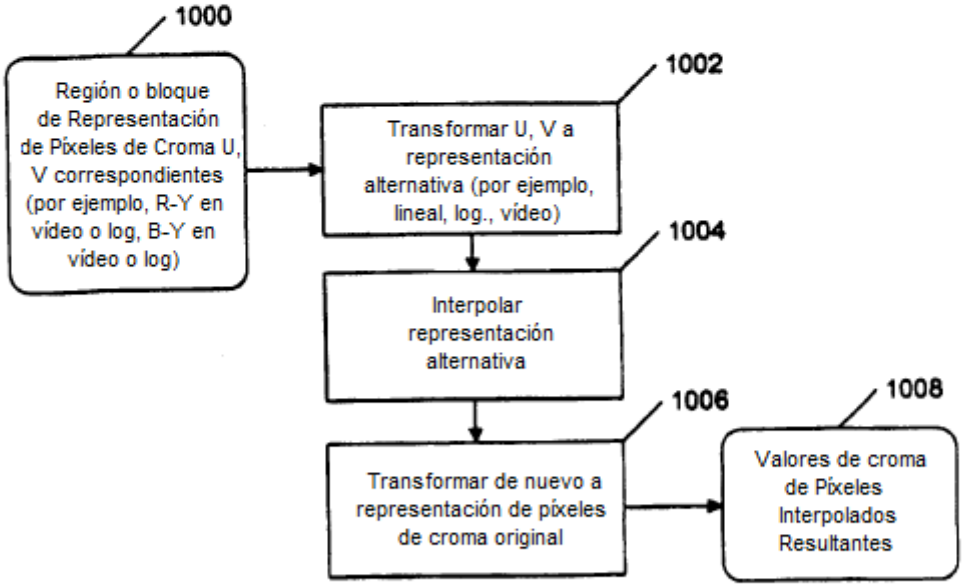
1000
Región o bloque de Representación de Píxeles de Croma U, V correspondientes (por ejemplo, R-Y en vídeo o log, B-Y en vídeo o log)
1002
Transformar U, V a representación alternativa (por ejemplo, lineal, log., vídeo)
1004
Interpolar representación alternativa
1006
Transformar de nuevo a representación de píxeles de croma original
1008
Valores de croma de Píxeles Interpolados Resultantes

FIG. 10

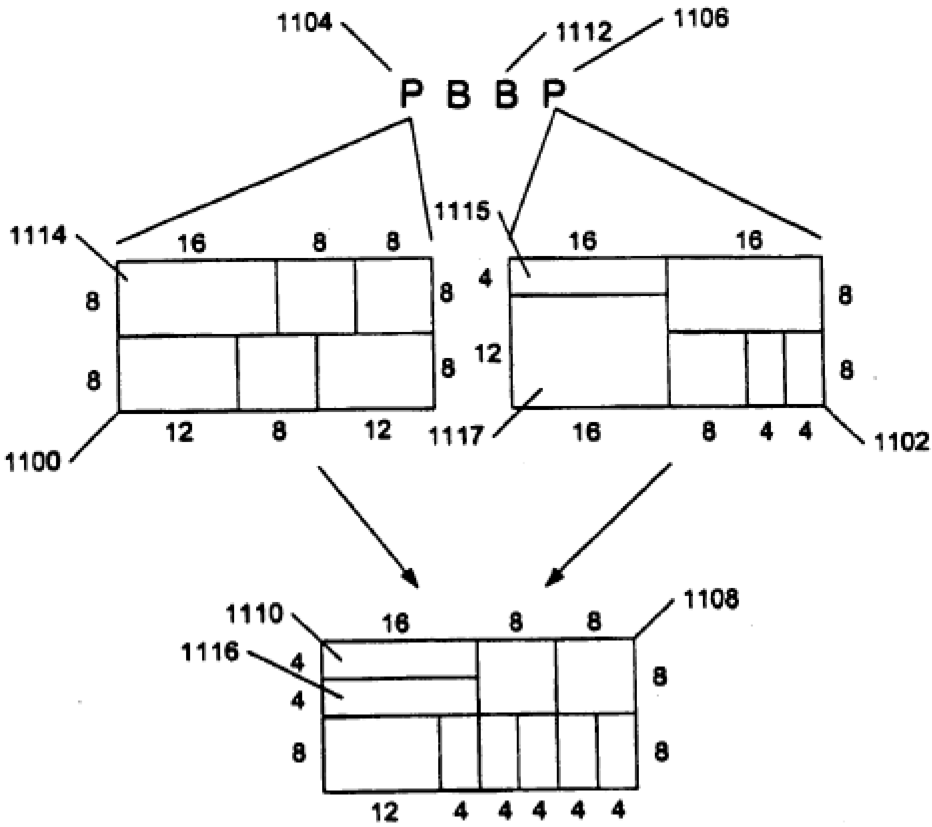
1104
1112
1106
P B B P
1114
16
8
8
8
8
8
8
12
8
12
1100
1115
16
16
4
8
12
8
1117
16
8
4
4
1102
1110
1116
16
8
8
1108
4
4
8
8
8
12
4
4
4
4
4

FIG. 11

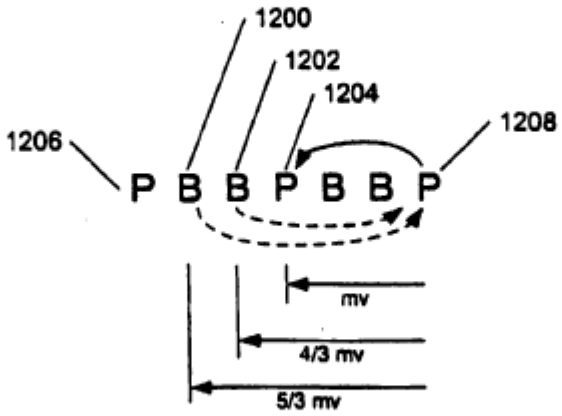
1200
1202
1204
1206
1208
P B B P B B P
mv
4/3 mv
5/3 mv

FIG. 12

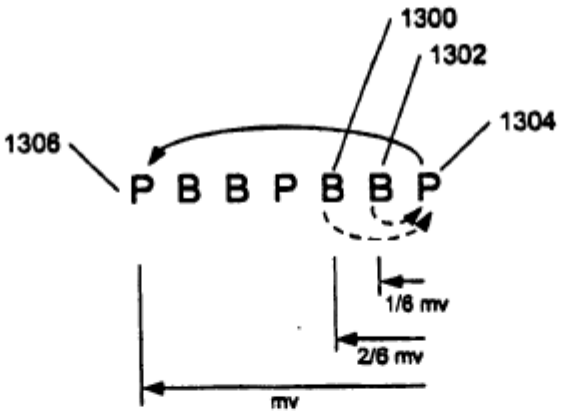
1300
1302
1304
1306
P B B P B B P
1/6 mv
2/6 mv
mv

FIG. 13

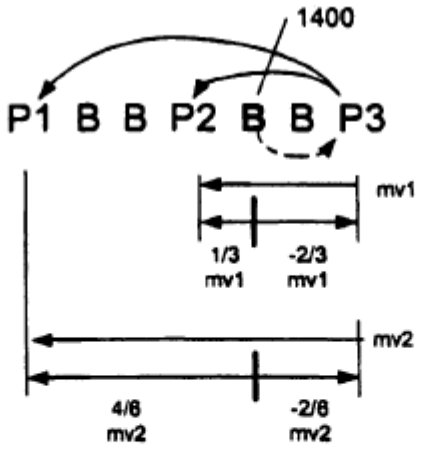
1400
P1 B B P2 B B P3
mv1
1/3
mv1
-2/3
mv1
mv2
4/6
mv2
-2/6
mv2

FIG. 14

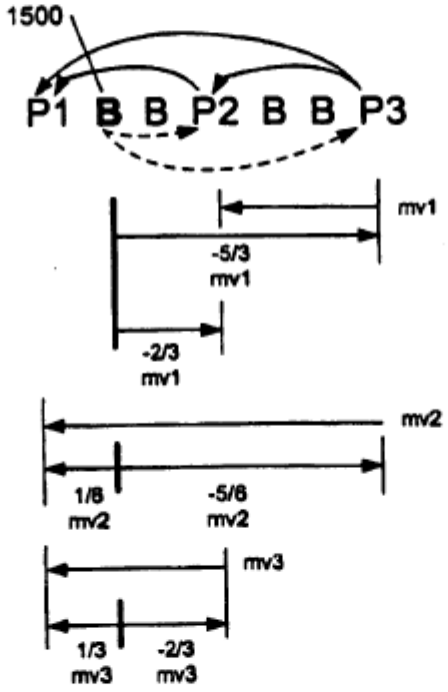
1500
P1 B B P2 B B P3
mv1
-5/3
mv1
-2/3
mv1
mv2
1/6
mv2
-5/6
mv2
mv3
1/3
mv3
-2/3
mv3

FIG. 15

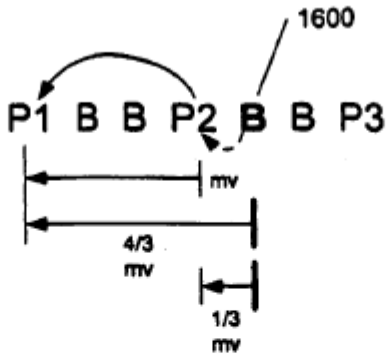
1600
P1 B B P2 B B P3
mv
4/3
mv
1/3
mv

FIG. 16

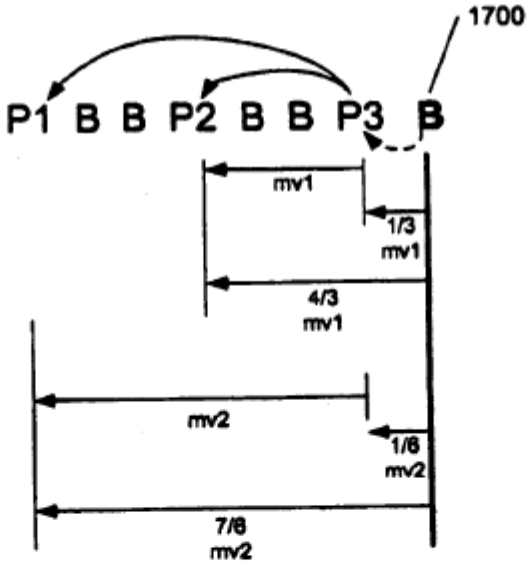
1700
P1 B B P2 B B P3 B
mv1
1/3
mv1
4/3
mv1
mv2
1/6
mv2
7/6
mv2

FIG. 17

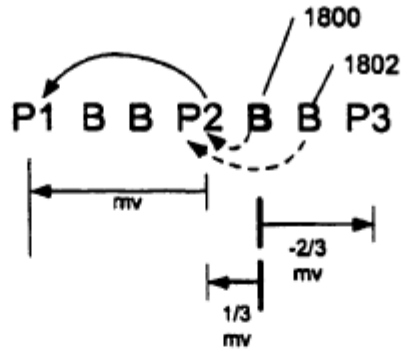

FIG. 18

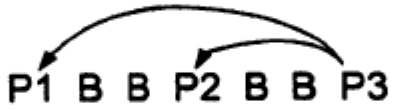

FIG. 19

FIG. 20

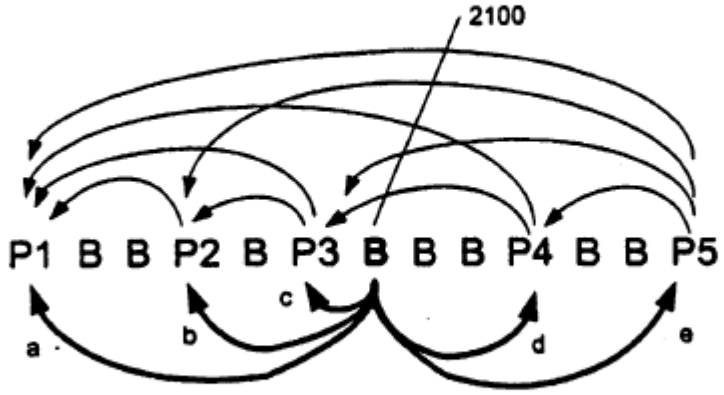

FIG. 21

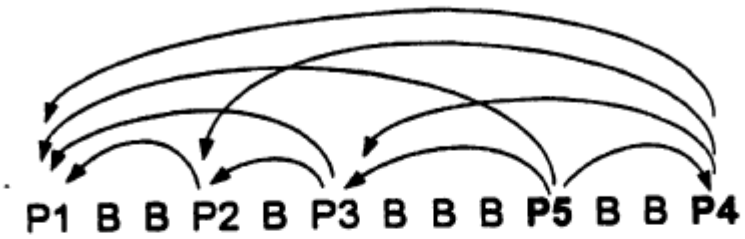

FIG. 22

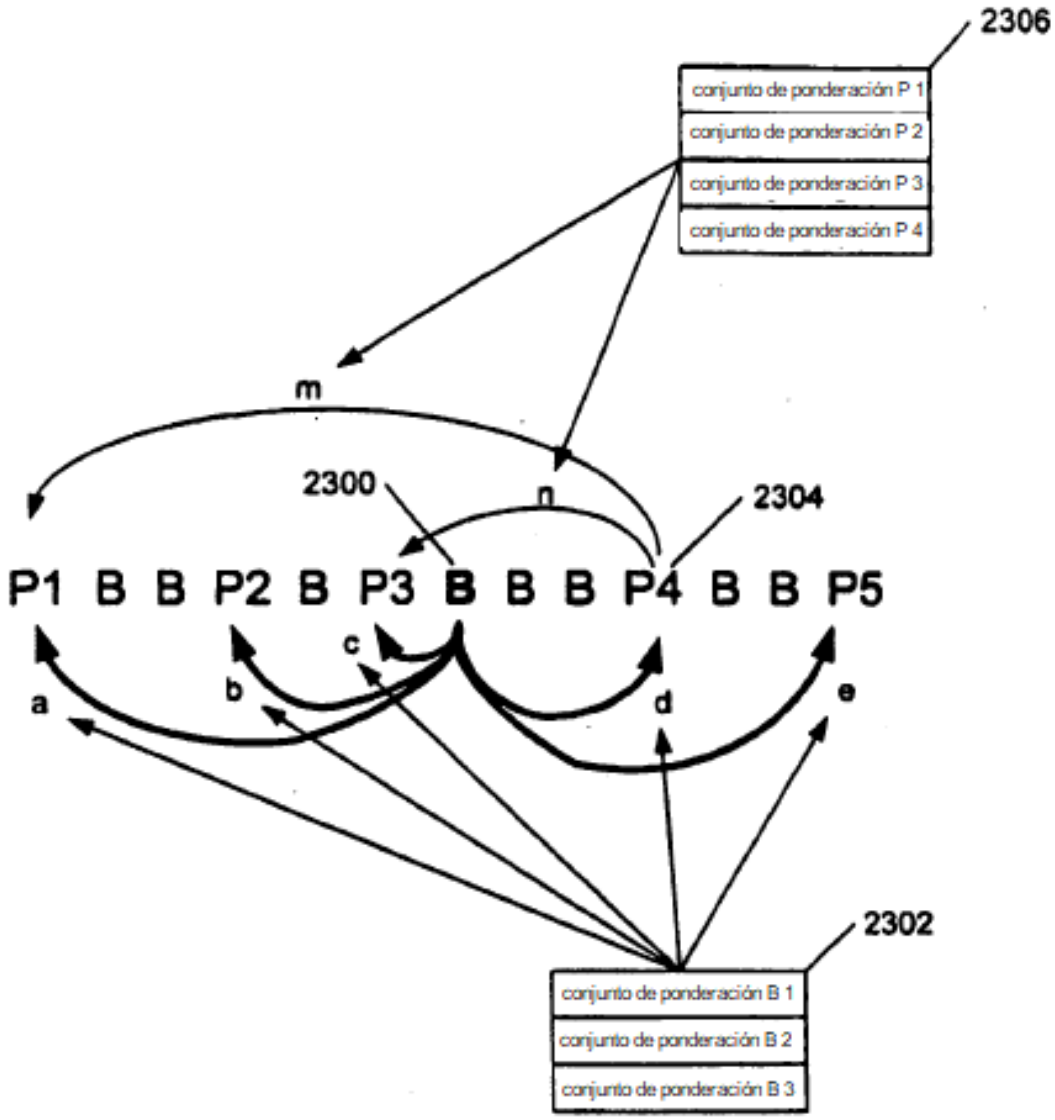
2306
conjunto de ponderación P 1
conjunto de ponderación P 2
conjunto de ponderación P 3
conjunto de ponderación P 4
m
n
2300
2304
P1 B B P2 B P3 B B B P4 B B P5
a
b
c
d
e
2302
conjunto de ponderación B 1
conjunto de ponderación B 2
conjunto de ponderación B 3

FIG. 23

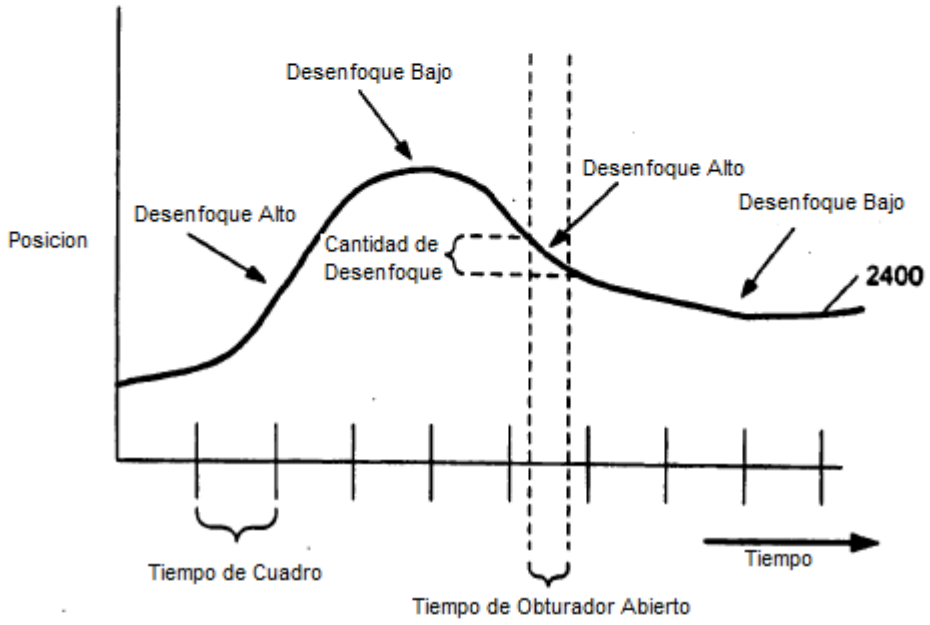
Desenfoque Bajo
Desenfoque Alto
Desenfoque Alto
Desenfoque Bajo
Posicion
Cantidad de
Desenfoque
2400
Tiempo de Cuadro
Tiempo
Tiempo de Obturador Abierto

FIG. 24