



Ingeniería. Investigación y Tecnología

ISSN: 1405-7743

iit.revista@gmail.com

Universidad Nacional Autónoma de México
México

Ochoa Domínguez, H.de J.; Mireles García, J.; Cota Ruíz, J. de D.
Descripción del nuevo estándar de video H.264 y comparación de su eficiencia de codificación con
otros estándares
Ingeniería. Investigación y Tecnología, vol. VIII, núm. 3, julio-septiembre, 2007, pp. 157-180
Universidad Nacional Autónoma de México
Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=40480304>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



Descripción del nuevo estándar de video H.264 y comparación de su eficiencia de codificación con otros estándares

H.de J. Ochoa-Domínguez, J.Mireles-García, J. de D. Cota-Ruiz
Departamento de Ingeniería Eléctrica y Computación
Instituto de Ingeniería y Tecnología de la Universidad Autónoma de Ciudad Juárez
E-mail: hochoa@uacj.mx

(Recibido: septiembre de 2005; aceptado: diciembre de 2006)

Resumen

Los estándares de video han sido desarrollados con el fin de satisfacer una amplia gama de aplicaciones, como son: el almacenamiento digital, transmisión y recepción de multi media, CATV, DVD, video conferencia, indexado de multi-media, cinema digital, entre otras. Los estándares de video logran una alta compresión utilizando varios métodos que explotan las redundancias temporal y espacial. El nuevo estándar de video H.264/MPEG-4 parte 10, no sólo es eficiente para el almacenamiento de video, sino que también proporciona un alto rendimiento en compresión y es más robusto a errores de transmisión que sus antecesores MPEG-2, H.263 y MPEG-4 parte 2. El presente artículo describe al estándar H.264/MPEG-4 parte 10. Primero se describe el algoritmo de codificación y posteriormente se compara su eficiencia de la codificación contra otros estándares anteriores existentes. Las comparaciones demuestran que el H.264 tiene una eficiencia de codificación de aproximadamente 1.5 veces mayor, en cada secuencia de prueba, con relación a otros estándares.

Descriptores: Codificación de video, CAVLC, CABAC, H.264, MPEG4 parte 10, NAL, predicción intra cuadro, predicción inter cuadro, filtro de desbloqueo.

Abstract

Video standards have been developed to fulfill a wide range of applications such as digital storage, multimedia transmission and reception, CATV, DVD, video conferencing, multimedia indexing, digital cinema among others. Video standards develop high compression ratios using several methods to reduce temporal and spatial redundancies. The new video standard, H.264 / MPEG-4 part 10, not only is efficient to store video but also to obtain a higher coding efficiency and error resilience than the previous similar standards MPEG-2, H.263 and MPEG-4 part 2. In this article, we describe the new video standard H.264/MPEG-4 part 10. First, we describe the coding algorithm, and then we compare its coding efficiency versus the previous existing standards. Comparisons show that the new standard has a coding efficiency of about 1.5 times, in each test sequence, as compared to the other standards.

Keywords: Video coding, H.264, Intra coding, inter coding, CAVLC, CABAC, MPEG4 part 10, NAL, deblocking filter.

Introducción

Desde principios de los 90's, el grupo de expertos en codificación de video (VCEG, *Video Coding Expert Group*) de la Unión Internacional de Telecomunicaciones – Sector Telecomunicaciones (ITU-T, *International Telecommunication Union- Telecommunication sector*) y el grupo de expertos para imágenes en movimiento (MPEG, *Moving Picture Expert Group*) de la ISO/IEC, enfocaron sus investigaciones en las diferentes técnicas de codificación de video para diversas aplicaciones. En un inicio, la ITU-T desarrolló el estándar H.261 (ITU-T H.261, 1993) para aplicaciones de video conferencia; por otra parte, MPEG procesó el MPEG-1 (ISO/IEC 11172-2:1993, 1993), cuya intención inicial era para el almacenamiento de video en disco compacto (CD). Posteriormente, el grupo MPEG desarrolló el estándar MPEG-2 como una extensión del MPEG-1, cuya aplicación inicial era televisión digital estándar (SDT) y televisión de alta definición (HDTV) (ISO/IEC JTC1/SC29/WG11, 1994). Al mismo tiempo, la ITU-T lo adoptó como el estándar H.262 (ITU-T H.262, 2000). Estos estándares basan su operación por la división en rebanadas de los cuadros de video de entrada. A su vez, estas rebanadas de video se subdividen en macrobloques, que por último se dividen en bloques. Debido a la necesidad de cubrir un mayor rango de aplicaciones, la ISO/IEC desarrolló el estándar MPEG-4 parte 2, que puede operar de dos formas, la primera se basa en la división de los cuadros de entrada en porciones de video llamadas rebanadas de video, y la segunda, en la segmentación de objetos de video para que el usuario pueda interactuar con ellos o, simplemente para codificar cada objeto por separado (ISO/IEC JTC1/SC29/WG11, 2000). Este estándar consta de varios perfiles que manejan diferentes velocidades binarias compatibles con diferentes aplicaciones. A la par del surgimiento del MPEG-4 parte 2, y para obtener mayor

compresión que en el H.261, la ITU-T desarrolló el estándar H.263 para aplicación en video teléfonos y compatible con el MPEG-4 parte 2 perfil avanzado.

El grupo de expertos de la ITU-T y el grupo de expertos de la ISO/IEC, se unieron para conformar el equipo conjunto de video (JVT, *Joint Video Team*) y trabajar en el desarrollo de un nuevo estándar, con un mejor desempeño, tanto en la calidad de video como en la eficiencia de codificación. El nombre del nuevo estándar es el H.264 o MPEG-4 parte 10 (ITU-T H.264, 2003) y cuenta además con especificaciones simples de su sintaxis, lo cual proporciona una mejor integración con todos los protocolos actuales y arquitecturas múltiples. Esto permite incluir otras aplicaciones, tales como la transmisión de video y video conferencia en redes fijas e inalámbricas y en diferentes protocolos de transporte.

El grupo de expertos de la ITU-T y el grupo de expertos de la ISO/IEC, se unieron para conformar el equipo conjunto de video (JVT, *Joint Video Team*) y trabajar en el desarrollo de un nuevo estándar, con un mejor desempeño, tanto en la calidad de video como en la eficiencia de codificación. El nombre del nuevo estándar es el H.264 o MPEG-4 parte 10 (ITU-T H.264, 2003) y cuenta además con especificaciones simples de su sintaxis, lo cual proporciona una mejor integración con todos los protocolos actuales y arquitecturas múltiples. Esto permite incluir otras aplicaciones tales como la transmisión de video y video conferencia en redes fijas e inalámbricas y en diferentes protocolos de transporte.

El H.264 cuenta con los mismos elementos o bloques funcionales que sus antecesores, ya que también adopta un algoritmo híbrido de predicción y transformación para la reducción de la correlación espacial y de la señal

residual, control de la velocidad binaria o *bit rate*, predicción por compensación de movimiento para reducir la redundancia temporal, así como la codificación de la entropía para reducir la correlación estadística. Sin embargo, lo que hace que este estándar proporcione mayor eficiencia de codificación, es la manera en que opera cada bloque funcional. Por ejemplo, el H.264 incluye predicción intra cuadro (INTRA), característica única de este estándar (Huang, 2005); transformación por bloques de 4x4 muestras, cuyos coeficientes transformados resultan enteros (Wien, 2003), anteriormente, se incluía transformación de 8x8 muestras, referencia múltiple para predicción temporal, tamaño variable de los macrobloques a comprimir, precisión de un cuarto de píxel para la compensación de movimiento, filtro de desbloqueo (List, 1993), codificador de entropía mejorado. Todas estas mejoras vienen acompañadas de un aumento en la complejidad de la implementación.

Los canales inalámbricos suelen agregar mucho ruido a la señal original y no se puede llevar a cabo la recepción perfecta de la misma; cualquier error en la decodificación de un bit puede propagarse a bloques subsecuentes o incluso a cuadros subsecuentes, degradando la calidad subjetiva de la imagen. El H.264 utiliza métodos para incrementar la resistencia a errores. Por ejemplo, utiliza el ordenamiento flexible de macrobloques (FMO), la transmisión de rebanadas redundantes de cuadros de video e incluye el particionamiento de datos, que también es utilizado en los estándares previos.

En el H.264, al igual que en sus antecesores, se definen diferentes perfiles y niveles dentro de cada uno, los cuales especifican restricciones en el tren de bits o *bitstream*. Cada perfil especifica un conjunto de características y los límites del decodificador, aunque los codificadores no requie-

ren de ningún conjunto particular de características de un perfil. Los niveles especifican los límites de los valores que deben tomar los elementos de la sintaxis de la recomendación o estándar. En cada perfil se utiliza la misma definición de niveles, pero las aplicaciones individuales sólo pueden utilizar un nivel diferente en cada uno. Por lo general, la carga de procesamiento del decodificador y la capacidad de memoria para un perfil dado se desprende de los diferentes niveles.

En la primera versión del H.264 existen tres perfiles, el línea base o *baseline*, el principal o *main* y el extendido o *extended*. El perfil línea base se aplica a los servicios de conversación en tiempo real, como video conferencia y video teléfono. El perfil principal es para aplicaciones de almacenamiento digital de video y datos, así como de transmisión de televisión. El perfil extendido es aplicable también a servicios de multimedia en Internet. La figura 1 muestra la relación que existe entre estos perfiles. La última versión del H.264, define cuatro perfiles altos o superiores, detallados como extensiones del rango de fidelidad (*fidelity range extensions*) para aplicaciones de distribución de contenido (Sullivan, 2004), así como para edición y post procesamiento (*High, High 10, High 4:2:2 y High 4:4:4*). El perfil alto se desarrolló para procesar video de 8 bits con formato de muestreo de la crominancia de 4:2:0 y para aplicaciones que utilizan alta resolución. El perfil *high 10*, se desarrolló para procesar video de hasta 10 bits con formato de muestreo 4:2:0 de los cuadros de entrada, para aplicaciones que utilizan alta resolución y mayor exactitud. El perfil *high 4:2:2*, soporta el formato de muestreo de los cuadros de crominancia de 4:2:2 y hasta 10 bits por muestra de exactitud. El perfil 4:4:4 soporta el formato de muestreo de los cuadros de crominancia 4:4:4 y hasta 12 bits por muestra de exactitud, así como transformación residual

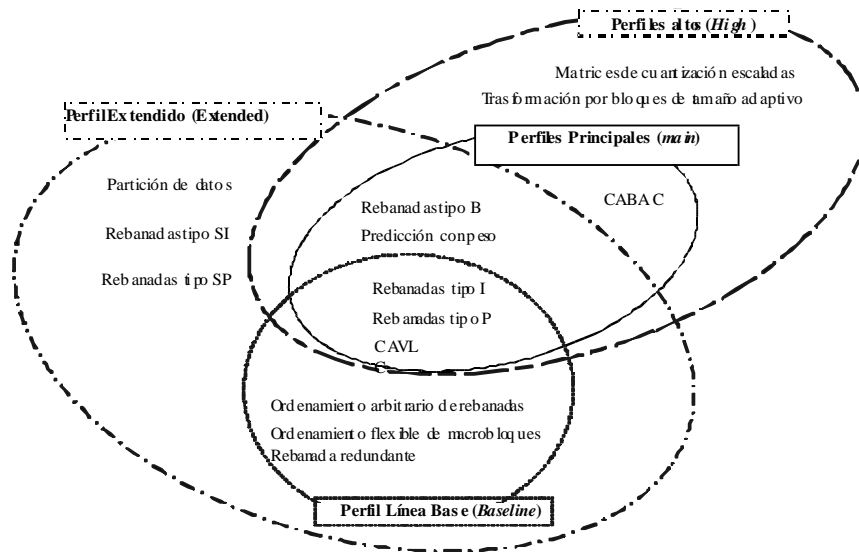


Figura 1. Partes comunes y partes específicas de cada perfil del estándar H.264

entera de los cuadros de color para codificar señales RGB. Las relaciones entre perfiles se describen a continuación:

Partes comunes a todos los perfiles

- Rebanadas tipo I (Rebanadas codificadas utilizando predicción INTRA): Rebanada codificada utilizando la predicción de las muestras decodificadas dentro de la misma rebanada.
- Rebanadas tipo P (Rebanadas codificadas utilizando codificación predictiva en un solo sentido): Rebanada codificada utilizando predicción intercuadro (INTER). Se utilizan como referencia los cuadros previamente decodificados con un vector de movimiento y un índice de referencia para predecir los valores de las muestras de cada bloque.
- Utiliza el codificador CAVLC (*Context-based Adaptive Variable Length Co-*

ding) para la codificación de la entropía.

Perfil línea base (Baseline)

- Ordenamiento flexible de macrobloques: Los macrobloques no necesariamente se ordenan horizontalmente y de izquierda a derecha (*raster scan*). Existe un mapa que asigna los macrobloques a un determinado grupo de rebanadas.
- Ordenamiento arbitrario de rebanadas: La dirección del primer macrobloque de una rebanada puede ser menor que la dirección del primer macrobloque de alguna rebanada anterior dentro del mismo cuadro codificado.
- Rebanada redundante: Esta rebanada pertenece a datos redundantes ya codificados con igual o diferente velocidad binaria en comparación con los mismos datos codificados pertenecientes a la misma rebanada.

Perfil principal (Main)

- Rebanadas tipo B (Rebanadas codificadas utilizando codificación predictiva bidireccional): Rebanada codificada utilizando predicción INTER con referencia de un cuadro previamente decodificado y utilizando a lo más, dos vectores de movimiento e índices de referencia para predecir los valores de las muestras de cada bloque.
- Predicción con peso: Esta es una operación de escalamiento, en la cual se aplica un factor de peso a las muestras resultantes de la compensación de movimiento en rebanadas tipo P o B.
- Utiliza el codificador CABAC (*Context-based Adaptive Binary Arithmetic Coding*) para codificación de la entropía.

Perfil extendido

- Se incluyen todas las partes del perfil línea base.
- Rebanadas tipo SP: Es un tipo de rebanada especial, codificada para cambiar eficientemente de resoluciones entre tramas de video (escalabilidad – capa de bits básicos + capa de ensanchamiento), es similar a la codificación de una rebanada tipo P.
- Rebanadas tipo SI: Igual que el anterior, pero es similar a codificar una rebanada tipo I.
- Particionamiento de datos: Los datos codificados se colocan en particiones separadas en capas diferentes. Se utiliza para reducir errores durante la transmisión y para el empaquetamiento efectivo de los datos.

- Rebanadas tipo B: Se utiliza para predicción bidireccional, incrementa la calidad del video y la velocidad binaria.

- Predicción con peso.

Perfil alto (High)

- Incluye todas las partes del perfil principal (*Main*).
- Transformación por bloques de tamaño adaptivo: DCT de 4x4 y transformación DCT entera para los bloques de luminancia. El tamaño del bloque a codificar dentro de un macrobloque puede ser diferente.
- Matrices de cuantización escaladas: Se utilizan diferentes escalas de acuerdo con frecuencias específicas asociadas con los coeficientes transformados para optimizar la calidad subjetiva del video.

La tabla 1 lista a los perfiles del H.264 y del MPEG-4 parte 2 (video), así como los requerimientos más importantes para cada aplicación.

Los niveles de cada perfil se muestran en la tabla 2. Cada nivel soporta diferentes tamaños de los cuadros de entrada (*QCIF*, *CIF*, *ITU-R 601 (SDTV)*, *HDTV*, *S-HDTV*, *Cinema digital* (Richardson, 2003)). Cada nivel ajusta los límites de las velocidades binarias, tamaño de la memoria para almacenar cuadros de referencia, etc. La tabla 3 muestra los parámetros límite de cada nivel.

Tabla 1. Tabla 1. Requerimientos por aplicación (Richardson, 2003) (SP, ASP, ARTS, FGS, Estudio de perfiles: Simple, avanzado simple, avanzado simple en tiempo real, escalabilidad gran ular fina y perfiles de estudio)

Aplicación	Requerimientos	Perfiles H.264	Perfiles MPEG-4/2
Transmisión de TV	Eficiencia en la codificación, confiabilidad en canales de distribución controlados, transmisión entrelazada de los campos de cada cuadro de vídeo, decodificador de baja complejidad.	Principal	Perfil Simple Avanzado (ASP, <i>Advanced Simple Profile</i>)
Video por cable o Internet	Eficiencia en la codificación, confiabilidad en canales no controlados para redes basadas en paquetes de distribución y escalabilidad.	Extendido	ARTS o FGS
Almacenamiento y reproducción de video	Eficiencia en la codificación, transmisión entrelazada de los campos de cada cuadro de vídeo, decodificador de baja complejidad.	Principal	ASP
Videoconferencia	Eficiencia en la codificación, confiabilidad, baja complejidad del codificador y del decodificador.	Básico	Perfil Simple (SP)
Video a través de redes inalámbricas	Eficiencia en la codificación, confiabilidad, baja complejidad del codificador y del decodificador, bajo consumo de potencia.	Básico	SP
Distribución de video (Estudio)	Características de compresión sin pérdidas o cercanas, transmisión entrelazada de los campos de cada cuadro de vídeo, transcodificación efectiva.	Principal Alto	Estudio

Tabla 2. Niveles, tipos y velocidades de cuadros (SHDTV: Super HDTV, HHR: Alta resolución horizontal, p: progresivo, i: entrelazado)

Niveles	Tipo de cuadros y velocidad de cuadros
1	QCIF @ 15fps (<i>frames per second</i>)
1.1	QCIF @ 30fps
1.2	CIF @ 15fps
1.3	CIF @ 30fps
2	CIF @ 30fps
2.1	HHR @15 or 30fps
2.2	SDTV @ 15fps
3	SDTV: 720x480x30i, 720x576x25i 10Mbps(max)
3.1	1280x720x30p
3.2	1280x720x60p
4	HDTV: 1920x1080x30i, 1280x720x60p, 2Kx1Kx30p 20Mbps(max)

(continúa...)

Tabla 2. Niveles, tipos y velocidades de cuadros (SHDTV: Super HDTV, HHR: Alta resolución horizontal, p: progresivo, i: entrelazado) (...continuación)

Niveles	Tipo de cuadros y velocidad de cuadros
4.1	HDTV: 1920x1080x30i, 1280x720x60p, 2Kx1Kx30p 50Mbps(max)
4.2	HDTV: 1920x1080x60i, 2Kx1Kx60p
5	SHDTV/Cinema digital: 2.5Kx2Kx30p
5.1	SHDTV/Cinema digital: 4Kx2Kx30p

Tabla 3. Límite de los parámetros de cada nivel (los números de nivel no entero se refieren a los niveles intermedios. Los guiones denotan ausencia del límite correspondiente)

Nivel	Vel. max. de procesamiento del macrobloque (MB/s)	Tamaño max. del cuadro (MBs)	Tamaño max. de la memoria para el cuadro decodificado (1024 bytes)	Max. velocidad binaria (1000 bits/s o 1200 bits/s)	Max. CPB size (1000 bits or 1200 bits)	Rango de la componente vertical del vector de movimiento (muestras de luminancia)	Relación de compresión mínima	Número max. de vectores de movimiento por cada dos MBs consecutivos
1	1 485	99	148.5	64	175	[-64,+63.75]	2	-
1.1	3 000	396	337.5	192	500	[-128,+127.75]	2	-
1.2	6 000	396	891.0	384	1 000	[-128,+127.75]	2	-
1.3	11 880	396	891.0	768	2 000	[-128,+127.75]	2	-
2	11 880	396	891.0	2 000	2 000	[-128,+127.75]	2	-
2.1	19 800	792	1 782.0	4 000	4 000	[-256,+255.75]	2	-
2.2	20 250	1 620	3 037.5	4 000	4 000	[-256,+255.75]	2	-
3	40 500	1 620	3 037.5	10 000	10 000	[-256,+255.75]	2	32
3.1	108 000	3 600	6 750.0	14 000	14 000	[-512,+511.75]	4	16
3.2	216 000	5 120	7 680.0	20 000	20 000	[-512,+511.75]	4	16
4	245 760	8 192	12 288.0	20 000	25 000	[-512,+511.75]	4	16
4.1	245 760	8 192	12 288.0	50 000	62 500	[-512,+511.75]	2	16
4.2	491 520	8 192	12 288.0	50 000	62 500	[-512,+511.75]	2	16
5	589 824	22 080	41 310.0	135 000	135 000	[-512,+511.75]	2	16
5.1	983 040	36 864	69 120.0	240 000	240 000	[-512,+511.75]	2	16

Estructura de las capas del H.264

El H.264 se compone de dos capas, la capa de la red de abstracción (NAL, *Network Abstraction Layer*) y la capa de codificación del video (VCL, *Video Coding Layer*). NAL abstrae los datos para hacer compatible al tren de bits de salida del codificador con casi todos los canales de comunicación o medios de almacenamiento. Esta unidad de red, especifica los datos en un formato de bytes (*byte-stream*) o de paquetes. El formato de bytes define

patrones de bytes o de bits utilizados en el estándar H.320 o en el MPEG-2. El formato de paquetes, define paquetes de datos identificables por protocolos de transporte para aplicaciones de RTP/UDP/IP (Wenger, 2003).

La capa VCL constituye el núcleo de los datos codificados. Ésta consiste en la secuencia de video a codificar, cuadros o campos dentro de la secuencia de video con tres arreglos de muestras (luminancia, crominancia o RGB), rebanadas dentro de cada

cuadro y macrobloques dentro de cada rebanada, así como bloques dentro de cada macrobloque. También, el H.264 soporta búsqueda (escaneo) progresiva y entrelazada, la cual puede mezclarse dentro de la misma secuencia. El perfil línea base se limita únicamente a búsqueda progresiva. Los cuadros se dividen en rebanadas. Una rebanada de video es una secuencia de macrobloques que pueden tener distintos tamaños (tamaños flexibles). En el caso de grupos de rebanadas, la posición de un macrobloque se determina por medio de un mapa que representa al grupo de rebanadas. El mapa indica a cuál

grupo de rebanadas pertenece el macrobloque (Ghanbari, 2003).

Algoritmo de codificación de video

La figura 2 muestra el diagrama a bloques del algoritmo de codificación del H.264. El codificador debe seleccionar entre codificación INTRA o INTER. En codificación INTRA se utilizan varios modos de predicción para reducir la redundancia espacial de un solo cuadro. La codificación INTER es más eficiente y se utiliza en la codificación tipo P o B (predictiva o bidireccional) de cada bloque de

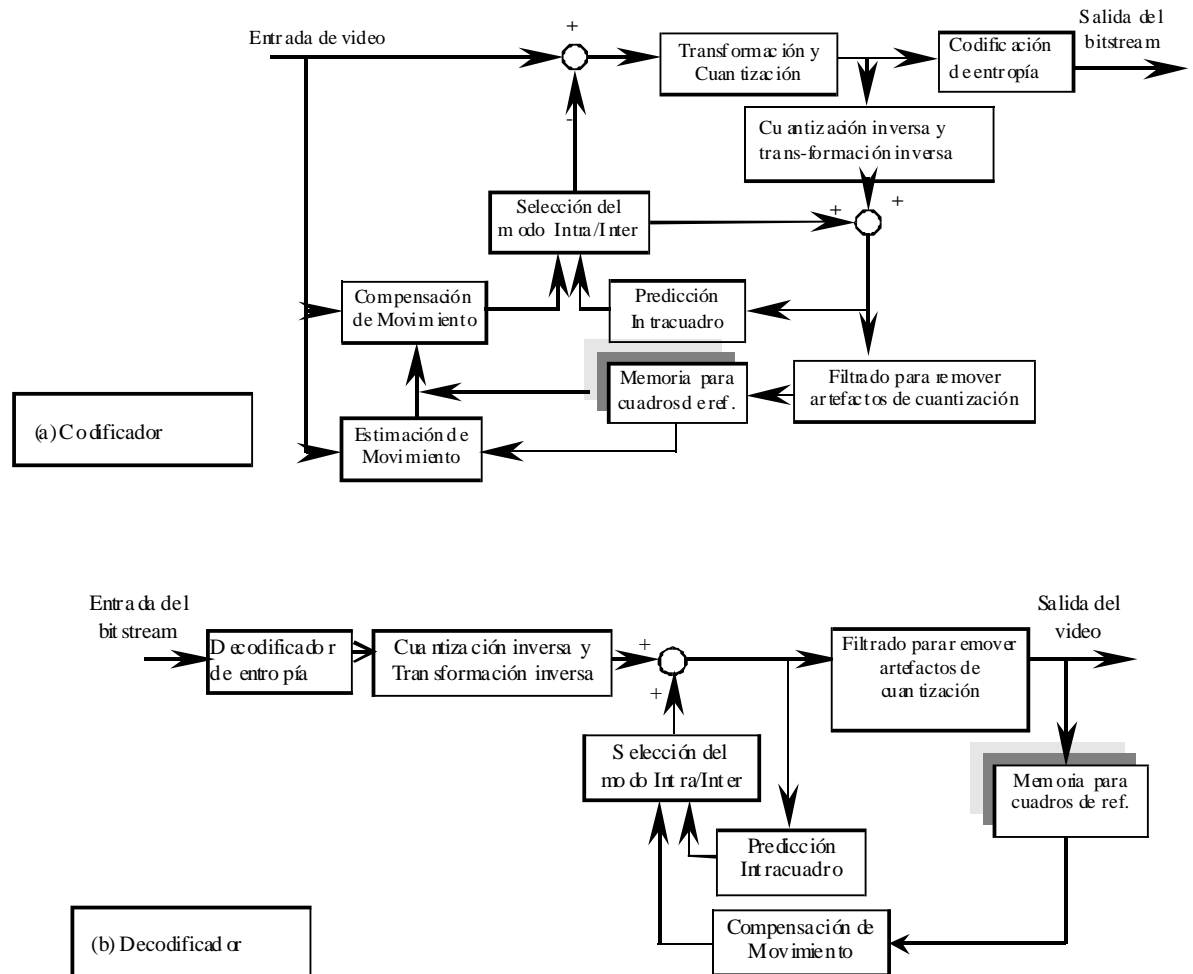


Figura 2. Diagrama a bloques del CODEC del H.264, (a) codificador, (b) decodificador

muestras. En esta codificación se utilizan como referencia los cuadros decodificados previamente. La codificación INTER utiliza vectores de movimiento para reducir la redundancia temporal entre cuadros. La predicción se obtiene después de filtrar el bloque anterior reconstruido. El filtro reduce los artefactos o distorsiones introducidos en las orillas de un bloque, debido a la cuantización. Los vectores de movimiento y el modo de predicción INTRA pueden tener varias especificaciones, dependiendo del tamaño de los bloques a codificar. Antes de ser cuantizado, el error o predicción residual se comprime aún más utilizando una transformada, la cual remueve la correlación espacial del bloque. Finalmente, los vectores de movimiento y los coeficientes cuantizados se codifican utilizando codificadores de entropía tales como el CAVLC (*Context-Adaptive Variable Length Codes*) o el CABAC (*Context-Adaptive Binary Arithmetic Codes*) (Marpe, 2003).

Predicción intracuadro (modo INTRA)

Los estándares anteriores llevan a cabo la predicción INTRA, codificando independientemente cada macrobloque (MB), como si se tratara de la codificación de una imagen (Ochoa, 2006), ya que se necesita que tengan la menor distorsión posible para que sirvan de referencia a la codificación INTER. La codificación INTRA, también se aplica dentro

de rebanadas codificadas en modo INTRA o en macrobloques que contienen una corrección temporal inaceptable, aunque se haya utilizando compensación de movimiento (macrobloques INTER con una distorsión por encima de un nivel de referencia específico). Este efecto provoca que se incremente el número de bits en el modo INTRA, haciendo imposible reducir la velocidad binaria.

El H.264 utiliza el concepto de predicción INTRA para codificar bloques o macrobloques de referencia y reducir la cantidad de bits codificados. Para codificar un bloque o macrobloque en modo INTRA, se forma un bloque de predicción basado en un bloque reconstruido previamente dentro del mismo cuadro y sin filtrar. Posteriormente, se codifica la señal residual (error) entre el bloque actual y la predicción, disminuyendo considerablemente la cantidad de bits que representan al bloque actual. El bloque de luminancia bajo predicción, puede formarse por subbloques de 4x4 muestras o por todo el bloque de 16x16 muestras. Para cada bloque de luminancia de 4x4, se selecciona un modo de predicción de nueve modos existentes. Existen cuatro modos de predicción para bloques de luminancia de 16x16 muestras. Solo existe un modo de predicción para cada bloque de crominancia de 4x4 muestras.

La figura 3 muestra la predicción de un bloque de luminancia de 4x4 muestras. [a, b,

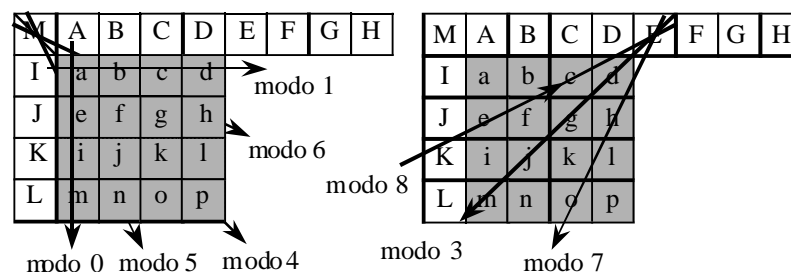


Figura 3. Direcciones de los modos de predicción INTRA 4 x 4 (Vertical: 0, horizontal: 1, DC: 2, diagonal izquierda y hacia abajo: 3, diagonal derecha hacia abajo: 4, vertical derecha: 5, horizontal hacia abajo: 6, vertical izquierda: 7, horizontal hacia arriba: 8)

..., p] son las muestras bajo predicción del bloque actual y [A, B, ..., M] las muestras del bloque previamente reconstruido. Las muestras se utilizan de acuerdo con las direcciones de predicción de cada modo, las flechas indican la dirección de predicción (ITU-T H.264, 2003).

En el modo 0 (vertical) y el modo 1 (horizontal), las muestras predecidas se forman por extrapolación de las muestras superiores [A, B, C, D] y de las muestras de la izquierda [I, J, K, L] respectivamente. En el modo 2 (DC), todas las muestras predecidas se forman por medio de las muestras [A, B, C, D, I, J, K, L]. En el modo 3 (diagonal izquierda y hacia abajo), modo 4 (diagonal derecha hacia abajo), modo 5 (vertical derecha), modo 6 (horizontal hacia abajo), modo 7 (vertical izquierda) y modo 8 (horizontal hacia arriba), las muestras predecidas se forman dando cierto peso al promedio de las muestras A-M. Por ejemplo, las muestras 'a' y 'd' son predecidas redondeando el resultado $\text{round}(I/4 + M/2 + A/4)$ y $\text{round}(B/4 + C/2 + D/4)$ en el modo 4. También, por medio de $\text{round}(I/2 + J/2)$ y $\text{round}(J/4 + K/2 + L/4)$ en el modo 8. El modo seleccionado por el codificador será aquel que minimice el error o residuo entre el bloque a codificar y su predicción (ITU-T H.264, 2003).

Existen únicamente 4 modos para la predicción de los componentes de luma de un bloque de 16x16. Para los modos 0 (vertical), 1 (horizontal) y 2 (DC) las predicciones son similares a los casos de 4x4. Para el modo 4 (plano), se utiliza una función lineal que se ajusta a las muestras superior e izquierda.

Cada bloque de 8x8 muestras de la componente de crominancia, en un macrobloque, se predice de las muestras de los bloques superior o izquierdo previamente codificados y reconstruidos. Los cuatro modos de

predicción son similares a los modos de predicción de los bloques de 16x6 de luminancia, excepto que el orden de los modos es diferente, modo 0 (DC), modo 1 (horizontal), modo 2 (vertical) y modo 3 (plano).

Predicción intercuadro (modo INTER)

La predicción INTER, la estimación del movimiento y la compensación del movimiento son tres factores que ayudan a reducir la redundancia o correlación temporal. En el H.264, el cuadro actual puede particionarse en macrobloques o bloques más pequeños. La compensación de movimiento realizada con bloques más pequeños, incrementa la ganancia de la codificación, a costa de incrementar el número de datos necesarios para representar la compensación. En el proceso de codificación INTER, se pueden procesar bloques de hasta 4x4 muestras de luminancia, utilizando una exactitud en los vectores de movimiento de hasta un cuarto de muestra. El proceso de predicción INTER de un bloque, involucra la selección de cuadros reconstruidos previamente, los cuales están almacenados en la memoria del sistema. Los vectores de movimientos se codifican utilizando una técnica diferencial. A diferencia de otros estándares, en el H.264 las rebanadas codificadas bidireccionalmente se pueden utilizar como referencia para la codificación INTER de otros cuadros (ITU-T H.264, 2003).

Predicción en tamaños variables de bloques

Un macrobloque de luminancia de 16x16 muestras, puede dividirse en pequeños bloques de hasta 4x4, como se muestra en la figura 4. Existen cuatro casos: 16x16, 16x8, 8x16 y 8x8 para los macrobloques de 16x16. También existen cuatro casos para un macrobloque de luminancia de 8x8: 8x8, 8x4, 4x8,

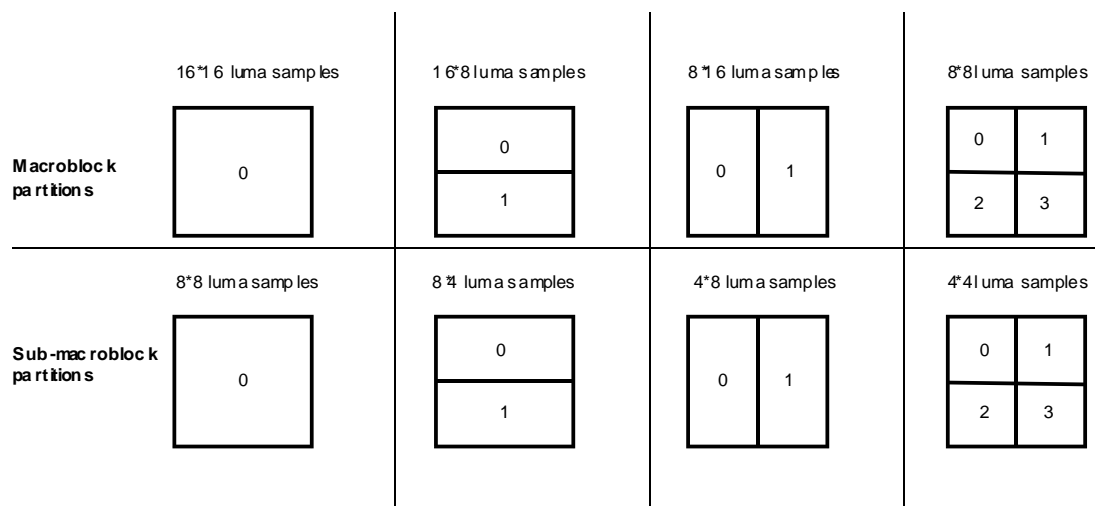


Figura 4. Particiones de macrobloques y sub macrobloque

4x4. Entre más pequeño sea el bloque bajo predicción, mayor será el número de bits para representar los vectores de movimiento y otros datos tales como el tipo de partición; sin embargo, si se utilizan particiones pequeñas se puede reducir considerablemente el error o residuo resultante de la compensación de movimiento. La selección del tamaño de la partición depende de las características del video de entrada. En general, una partición grande es beneficiosa cuando las muestras del bloque son homogéneas y una partición pequeña cuando no existe homogeneidad (detalles) (Wien, 2003).

Interpolación para la predicción fraccional o sub muestra (Sub-pel, Sub Picture Element)

La predicción en modo INTER de cada partición de un macrobloque codificado, se lleva a cabo en un área del mismo tamaño del cuadro de referencia. En este caso, la compensación de movimiento se hace con respecto a muestras enteras. En ocasiones, el movimiento no se lleva a cabo en muestras enteras sino fraccionales, por lo que hay que

interpolar el área de referencia. El desplazamiento u *offset* entre las dos áreas (vectores de movimiento) tiene una resolución en fraccional (sub muestras), tanto para la luminancia como para la croma (Figura 5). Si el proceso de interpolación elegido incrementa la resolución del área de referencia (área de búsqueda) al doble, se dice que se tiene una compensación de movimiento con exactitud de $\frac{1}{2}$ muestra, y si se aumenta cuatro veces, se dice que se tiene una compensación de movimiento con exactitud de $\frac{1}{4}$ de muestra. En el H.264 la compensación de movimiento puede realizarse con una exactitud de hasta $\frac{1}{4}$ de muestra (ITU-T H.264, 2003).

La compensación de movimiento fraccional ($\frac{1}{2}$, $\frac{1}{4}$) resulta en una mayor compresión y mejor calidad de la imagen que la compensación de movimiento entera, a expensas del incremento de la complejidad en el codificador. En aplicaciones para velocidades binarias altas, así como altas resoluciones, se ha comprobado que la predicción fraccional de $\frac{1}{4}$ de muestra es mucho más eficiente que la de $\frac{1}{2}$ muestra, en cuanto a compresión y calidad de la imagen se refiere.

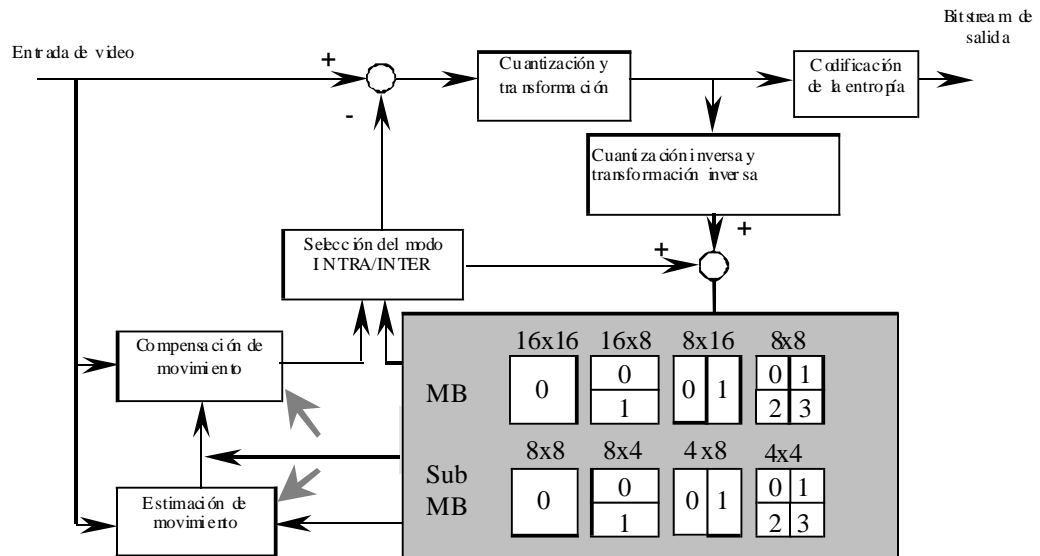


Figura 5. Diagrama a bloques de codificador enfatizando la estimación del movimiento y la compensación de movimiento

En las muestras de luminancia, la exactitud de $\frac{1}{2}$ muestra se genera filtrando las muestras vecinas con un filtro FIR, cuyos coeficientes son (1, -5, 20, 20, -5, 1)/32. Esto significa que cada muestra intermedia se obtiene de la suma de 6 muestras vecinas enteras, cada muestra con su correspondiente peso. Por ejemplo, en la figura 6, la muestra intermedia 'b' se calcula de la muestras E, F, G, H, I y J utilizando la formula 1.

$$b = \text{round}((E - 5F + 20G + 20H - 5I + J) / 32) \quad (1)$$

Similarmente, la muestra h se puede calcular de las muestras verticales A, C, G, M, R y T. Una vez calculadas las muestras con exactitud de $\frac{1}{2}$, se calculan las muestras con exactitud de $\frac{1}{4}$ por medio de interpolación bilineal entre muestras de exactitud entera y de $\frac{1}{2}$. Por ejemplo, la muestra 'a', con exactitud de $\frac{1}{4}$ de muestra, se puede calcular de la muestra entera G, la muestra 'b' con exactitud $\frac{1}{2}$, se calcula utilizando la ecuación 2.

$$a = \text{round}((G + b) / 2) \quad (2)$$

En el caso del formato de video 4:2:0, se requiere una exactitud de $\frac{1}{8}$ de muestra en las componentes de crominancia (correspondientes a $\frac{1}{4}$ de muestra en luminancia). Las muestras a $\frac{1}{8}$ de exactitud se interpolan utilizando interpolación bilineal entre muestras enteras de los cuadros de crominancia (ITU-T H.264, 2003).

Filtro de desbloqueo (remueve la distorsión debida a la cuantización en los bordes de los bloques)

El proceso de codificación involucra macrobloques con distintas características, algunos con mayor correlación que otros. Para mantener una cierta velocidad binaria, los bloques INTRA o INTER se cuantizan utilizando diferentes cuantizadores, los cuales introducen distorsión o artefactos indeseables alrededor de los bloques reconstruidos. En estándares anteriores, el filtro de desbloqueo era sólo una recomendación (opcional) del estándar, en el H.264, el filtro de desbloqueo es parte obligatoria del mismo. El filtro utilizado

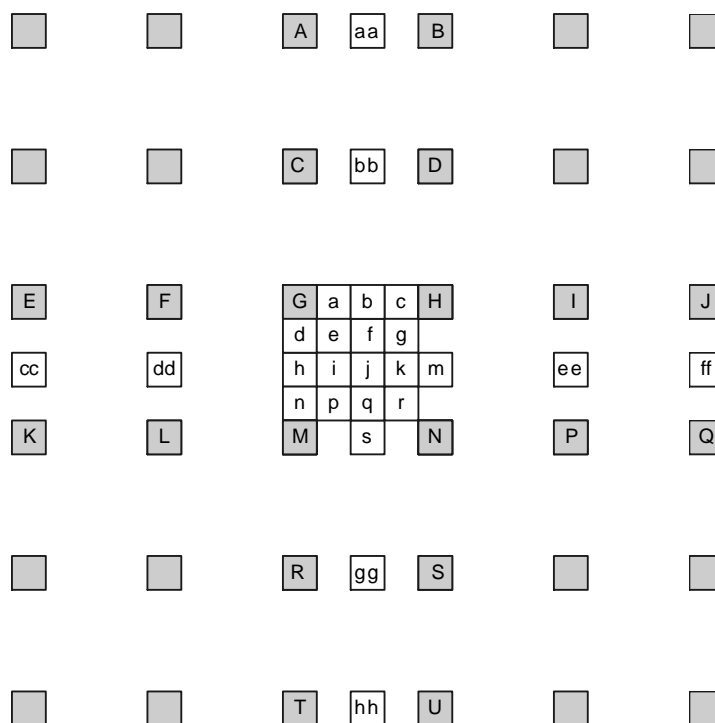


Figura 6. Muestras enteras (bloques sombreados con letras mayúsculas) y posiciones fraccionales (bloques blancos con letras minúsculas) para interpolación de un cuarto de muestra

en el H.264 reduce la distorsión en los bordes del bloque y evita que el ruido acumulado debido a la codificación se propague. El MPEG-1 y MPEG-2 no utilizan este filtro, debido a la complejidad de su implementación. Por otra parte, la distorsión en estos estándares se reduce utilizando compensación de movimiento con exactitud de $\frac{1}{2}$ muestra.

En el H.264, el filtrado se aplica a los bordes de bloques de 4x4 muestras de un macrobloque como se muestra en la figura 7. El filtro procesa por separado los bloques de luminancia y crominancia. El proceso de filtrado para bloques de luminancia se lleva a cabo en cuatro bordes de 16 muestras cada uno y el proceso de filtrado de luma en dos bordes de 8 muestras. La figura 7 (a) muestra el proceso de filtrado horizontal (afecta bordes verticales) y la figura 7(b) el proceso de filtrado vertical (afecta bordes horizontales).

El filtro se aplica adaptivamente en varios niveles:

- A nivel rebanada: El filtrado se puede ajustar a las características individuales de la secuencia de video.
- A nivel de borde de bloque: El filtrado se vuelve independiente de la decisión INTRA/INTER de las diferencias de movimiento y de la presencia de residuos codificados en los dos bloques participantes en el filtrado.
- A nivel muestra: El efecto de filtrado se puede anular dependiendo de los valores de las muestras y de los umbrales del cuantizador.

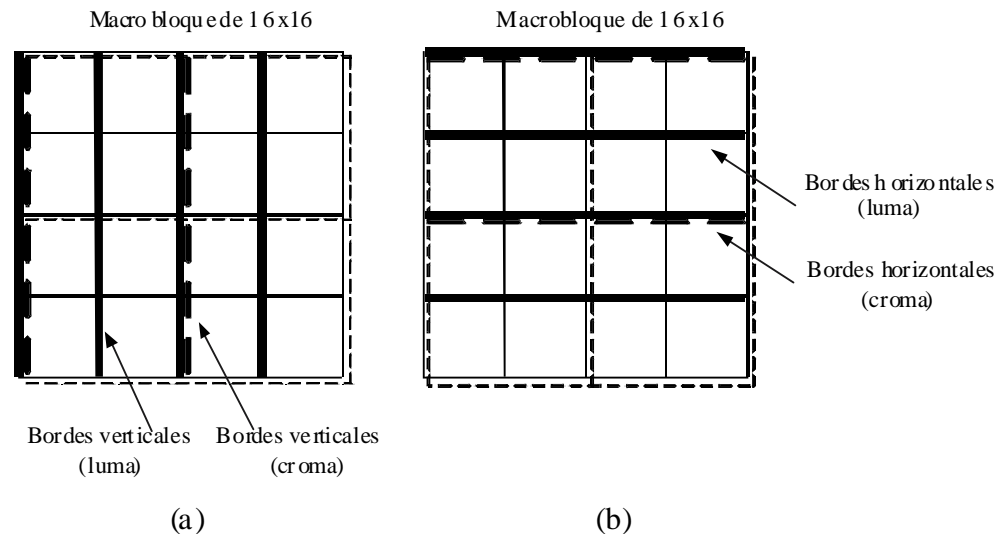


Figura 7. Bordes del macrobloque a filtrarse (los bordes de luminancia se muestran con líneas sólidas y las de crominancia con líneas entre cortadas o guiones)

Transformación y cuantización

El H.264 se basa en la codificación de macrobloques, utiliza transformadas para remover la redundancia espacial (Ochoa, 2006). Los cuadros de entrada al codificador, así como los residuos, resultado de la predicción, contienen una correlación o redundancia espacial alta. Después de la predicción INTER cuadro, o predicción espacial basada en las muestras de cuadros previamente decodificados con respecto al cuadro actual (bajo predicción), el residuo o predicción resultante se divide en bloques de 4x4 u 8x8 muestras, los cuales se convierten al dominio de la frecuencia y los coeficientes resultantes se cuantizan.

El H.264 utiliza una transformada de tamaño adaptivo de 4x4 y de 8x8 (perfiles altos); los estándares anteriores utilizaban únicamente la transformada DCT de 8x8. Una transformada de 4x4 es mucho menos compleja, ya que necesita menos multiplicaciones para llevar a cabo una transformación. También, si el codificador decide

procesar un bloque de menor tamaño, las distorsiones en los bordes del mismo bloque se reduce significativamente.

Para mejorar la eficiencia de la compresión, el H.264 emplea una estructura de transformación jerárquica, esto es, los coeficientes de DC de bloques vecinos de 4x4 de luminancia se agrupan en bloques de 4x4 y se aplica de nuevo la transformada Hadamard.

Existe una alta correlación entre coeficientes transformados de DC de bloques vecinos, cuyas muestras son muy similares. Por lo tanto, el estándar especifica la transformada Hadamard para los coeficientes de DC de bloques transformados de luma de 16x16, y la transformada Hadamard de 2x2 para los coeficientes de DC de bloques transformados de crominancia.

En algunas aplicaciones se requiere reducir el tamaño de los pasos de los cuantizadores para aumentar la relación señal a ruido pico (PSNR, *Peak Signal to Noise Ratio*) a niveles que se consideran visualmente sin

pérdidas. Para logra esto, el H.264 extiende el rango de los pasos de cuantización (QP) en dos octavas, teniendo que redefinir las tablas de cuantización y permitiendo que QP pueda variar de 0 a 51.

En general, la transformación y la cuantización requieren de varias multiplicaciones, esto eleva la complejidad de su implementación. Por lo tanto, para lograr una implementación más simple, el proceso de transformación exacta se modifica para reducir el número de multiplicaciones, resultando en una transformada entera, la cual integra en proceso de transformación, cuantización y escalamiento. Este proceso se llama transformación entera con post escalamiento

A continuación se describen los pasos para lograr la transformada directa DCT entera, post-escalamiento, así como la transformada inversa.

Proceso de codificación

Paso 1. Transformada entera directa

Para la transformación DCT exacta de un bloque de luminancia F de 4×4 , utilizamos

$$X = HFH^T \quad (3)$$

Donde H es la siguiente matriz

$$H = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix} \quad (4)$$

Las variables a, b, c y d tienen los siguientes valores.

$$a = \frac{1}{2}, \quad b = \sqrt{\frac{1}{2}} \cos\left(\frac{\pi}{8}\right), \quad c = \sqrt{\frac{1}{2}} \cos\left(\frac{3\pi}{8}\right) \quad (5)$$

A fin de facilitar la implementación de (5), c se aproxima a 0.5 y b se modifica para asegurar ortogonalidad, quedando como sigue:

$$a = \frac{1}{2}, \quad b = \sqrt{\frac{2}{5}}, \quad c = \sqrt{\frac{1}{2}} \quad (6)$$

Se pueden evitar algunas multiplicaciones en el proceso de transformación integrando el proceso de cuantización. Por lo tanto, la ecuación (1) se puede modificar antes de cuantizar de la siguiente forma:

$$X = \overline{H} F \overline{H}^T \otimes SF \quad (7)$$

donde

$$\overline{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad (8)$$

$$SF = \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix} \quad (9)$$

El símbolo \otimes denota multiplicación elemento por elemento de las matrices correspondientes.

Paso 2. Post escalamiento y cuantización

A la matriz SF se le aplica una cuantización, utilizando un cuantizador (Q_{step}), para después obtener una nueva matriz Y de la siguiente forma:

$$Y_{ij} = X_{ij} \text{round}\left(\frac{SF_{ij}}{Q_{step}}\right) \quad (10)$$

El H.264 define un total de 52 valores para Q_{step}

Proceso de decodificación

Paso 1. Cuantización inversa y preescalamiento

La señal Y recibida en el decodificador se escala utilizando el valor de Q_{step} y se utiliza SF^{-1} como parte de la cuantización y transformación inversas de la siguiente forma:

$$X_{ij} = Y_{ij} \cdot Q_{step} \cdot SF_j^{-1} \quad (11)$$

Paso 2. Transformación inversa entera

Recuperamos la señal F' de la siguiente forma:

$$F = \bar{H}_v^T X' \bar{H}_v \quad (12)$$

Donde la matriz de transformación inversa es:

$$\bar{H}_v = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/2 & -1/2 & -1 \\ 1 & -1 & -1 & 1 \\ 1/2 & -1 & 1 & -1/2 \end{bmatrix} \quad (13)$$

Posteriormente se aplica a bloques de 4x4 de coeficientes de DC de luminancia la transformada Hadamard de la ecuación 12. Esto constituye una transformación jerárquica.

$$\hat{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad (14)$$

A los coeficientes de DC de los bloques de crominancia (formato 4:2:0), se les aplica la siguiente transformada.

$$\hat{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (15)$$

Para los coeficientes de DC de los bloques de crominancia en los formatos 4:2:2 y 4:4:4 también se utiliza la transformada Hadamard adaptada.

La siguiente matriz de la DCT entera se utiliza únicamente en perfiles altos.

$$\bar{H} = \begin{bmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 12 & 10 & 6 & 3 & -3 & -6 & -10 & -12 \\ 8 & 4 & -4 & -8 & -8 & -4 & 4 & 8 \\ 10 & -3 & -12 & -6 & 6 & 12 & 3 & -10 \\ 8 & -8 & -8 & 8 & 8 & -8 & -8 & 8 \\ 6 & -12 & 3 & 10 & -10 & -3 & 12 & -6 \\ 4 & -8 & 8 & -4 & -4 & 8 & -8 & 4 \\ 3 & -6 & 10 & -12 & 12 & -10 & 6 & -3 \end{bmatrix} \quad (16)$$

Codificación en entropía

En estándares anteriores (MPEG-1, 2, 4, H.261 y H.263), la codificación de la entropía se basa en tablas previamente definidas, las cuales contienen los códigos de longitud variable (VLCs, *variable length code*) (Flierl, 2003), donde el conjunto de palabras de código en las tablas se basa en distribuciones de probabilidad de datos obtenidos en secuencias de video genérico, en lugar de utilizar la codificación Huffman o aritmética exacta para la codificar la secuencia en cuestión.

El H.264 utiliza diferentes VLCs a fin de igualar el símbolo que representa un dato de video, con un código basado en las características del contexto en el que se encuentra el símbolo. Todos los elementos de la sintaxis se codifican utilizando el código Exp-Golomb, excepto los datos residuales (Golomb, 1966). A fin de leer los datos residuales (coeficientes transformados y cuantizados) se utiliza una búsqueda en zigzag o una búsqueda alternada (campos de

cuadros de video no entrelazados). Para codificar los datos residuales, se utiliza un método más sofisticado llamado código de longitud variable adaptivo basado en el contexto (CAVLC, *Context Based Adaptive Variable Length Code*). En los perfiles principal y alto, también se utiliza otro método llamado CABAC, pero es más complejo que el CAVLC.

Código adaptivo de longitud variable basado en el contexto (CAVLC)

Después de la transformación y la cuantización, la probabilidad de que el valor de un coeficiente sea cero o ± 1 es muy alta. Por lo tanto, se codifica el número total de ceros y ± 1 que ocurren. Para los demás coeficientes sólo se codifica su nivel.

Por ejemplo, suponiendo que los siguientes coeficientes se van a codificar utilizando CAVLC

orden	0	1	2	3	4	5	6	7	8	9...	16
coef	c0,	c1,	c2,	1,	1,	0,	0,	-1,	0,	0,...	0

Las reglas que utilizaría el CAVLC son:

- Paso 1: Se utiliza una palabra de código de una tabla para expresar el número de coeficientes diferentes de cero, 6 (orden 0, 1, 2, 4, 5, 7) con magnitud uno, 3 (orden 4, 5, 7).
- Paso 2: Se forma una palabra de código para indicar los signos de los unos

en orden inverso de la siguiente forma, - (orden 7), + (orden 5) y +(orden 4).

- Paso 3: Se utiliza una palabra de código para cada nivel de los coeficientes restantes en orden inverso, esto es, una palabra de código para c2 (orden 2), c1, y c0.
- Paso 4: Se utiliza una palabra de código para indicar en número de ceros, 2 (orden 3, 6)
- Paso 5: Las palabras de código resultantes se codifican utilizando run-length en orden inverso, esto es, una palabra de código para 1 (orden 6-5), 0 (orden 4), 1 (orden 3-2).

Código aritmético binario adaptivo basado en el contexto (CABAC)

CABAC utiliza la codificación aritmética a fin de obtener una buena compresión. El modelo de probabilidad se actualiza con cada símbolo como se muestra en la figura 7 (Marpe, 2003).

Paso 1. Binarización. Proceso por el cual un símbolo no binario (coeficiente transformado, vector de movimiento, etc.) se mapea a una secuencia binaria única antes de aplicar la codificación aritmética.

Paso 2 Modelado del contexto. Un modelo de un contexto es un modelo de probabilidades para uno o más elementos de un símbolo binarizado. El modelo de probabi-

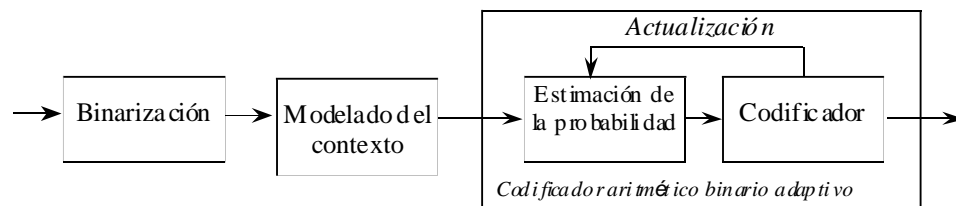


Figura 8. Diagrama a bloques del CABAC

lidad se selecciona de tal forma, que dicha selección depende sólo de elementos de la sintaxis previamente codificados.

Paso 3 Codificación binaria aritmética. Se utiliza el código aritmético para codificar cada elemento de acuerdo con la selección del modelo de probabilidad.

Rebanadas de video tipo B

La predicción bidireccional contribuye a reducir la correlación temporal, ya que utiliza como referencia más cuadros almacenados en memoria. Los estándares que manejan cuadros tipo B, utilizan el modo bidireccional, lo cual sólo permite la combinación de una señal predecida previamente con la predicción subsiguiente. Una señal predecida, se forma por la señal subsiguiente, codificada como INTER, una señal perteneciente a un cuadro anterior y otra señal que se forma tomando el promedio lineal de dos señales predecidas en donde se utiliza la compensación de movimiento.

El H.264 incluye no sólo predicción hacia adelante/atrás, sino también la predicción adelante/adelante y atrás/atrás (Flierl, 2003). Se pueden utilizar dos cuadros anteriores para la predicción de una región de video, justo antes de que la escena cambie y dos referencias posteriores justo después de que la escena cambió. A diferencia de los estándares previos, en el H.264, las rebanadas predecidas bidireccionalmente pueden utilizarse como referencia de otros cuadros que se codificarán en modo INTER. También, se agrega la predicción con peso cuando existen transiciones entre las diferentes escenas del video.

Predicción con peso

Los estándares anteriores consideran con igual peso a todos los cuadros o imágenes

que sirven como referencia para la predicción. Por ejemplo, una señal bajo predicción se obtiene promediando con igual peso las señales de referencia. Sin embargo, la predicción de las señales en transición gradual de escena a escena, precisa de diferentes pesos. La transición gradual es muy popular en cines, transiciones de difuminadas hacia el color negro (las muestras de luminancia de la escena difuminan gradualmente aproximándose a cero y las de crominancia se aproximan a 128) o desde el color negro.

El H.264 utiliza el método de predicción con pesos para los macrobloques de una rebanada tipo *P* o tipo *B*. Una señal de predicción *p* para una rebanada tipo *B*, se obtiene de dos señales de referencia (*r1* y *r2*) utilizando diferentes pesos.

$$p = w1 \times r1 + w2 \times r2 \quad (17)$$

Donde *w1* y *w2* son los pesos.

Los pesos se determinan en el codificador de dos maneras diferentes, implícitamente y explícitamente. En la determinación implícita, los pesos los calcula el decodificador basado en la distancia temporal entre cuadros de referencia. Si la distancia temporal entre la referencia y el cuadro actual es cercana, los pesos son pequeños, y si la distancia temporal es grande, los pesos son más grandes. En la determinación explícita los pesos se envían al decodificador en el encabezado.

Slices tipo SP y SI

En los estándares de codificación de video anteriores, el cambio entre trenes de bits es posible, únicamente en cuadros tipo *I*, reconstruyendo cuadros *I* a intervalos fijos, lo que permite el acceso aleatorio de los cuadros para reproducción rápida del video (*fast forward/backward*). Sin embargo, el problema

de usar cuadros *I* es que se requiere de un número de bits mucho mayor y no se explotan las redundancias temporales.

En el H.264 se utilizan rebanadas de video tipo SP y SI para cambiar el tren de bits (Ghanbari, 2003). La figura 9 muestra cómo utilizar los cuadros SP para cambiar entre diferentes trenes de bits. Suponiendo que existen dos trenes de bits, $P(1,k)$ y $P(2,k)$, correspondientes a la misma secuencia de video pero con diferentes velocidades binarias. Dentro del tren de bits codificado, los cuadros SP se colocan en una posición a la cual se permita un cambio de tren de bits.

En caso de cambiar del tren de bits superior $P(1,3)$ al inferior $P(2,3)$, se genera el cuadro SP $S(3)$ que permite producir un cuadro decodificado $P(2,3)$ utilizando $P(1,2)$ de otro tren de bits, aun cuando haya compensación de movimiento.

También se puede utilizar una rebanada de video tipo SI de manera similar al SP, pero la predicción se formaría utilizando el modo de predicción INTRA de 4x4 (Karczewicz, 2003).

Resistencia a errores

El particionamiento de los datos es un método muy popular para incrementar la resistencia a errores del sistema. Los datos son particionados de acuerdo con su significancia dentro del tren de bits. Posteriormente, se transmiten primero los datos con mayor prioridad para reducir el error medio cuadrático de la secuencia a un mínimo, posteriormente se transmiten los datos menos significativos. También, la codificación en capas (escalable) aumenta la resistencia a errores (transmisión de bits más significativos primero). Durante la codificación escalable espacial o temporal se pueden recuperar datos perdidos en otras capas.

Este nuevo estándar incrementa la resistencia a errores de transmisión por medio de la contribución de rebanadas tipo S, el ajuste de los parámetros de codificación, el ordenamiento flexible de los macrobloques y el uso de rebanadas redundantes (Wenger, 2003).

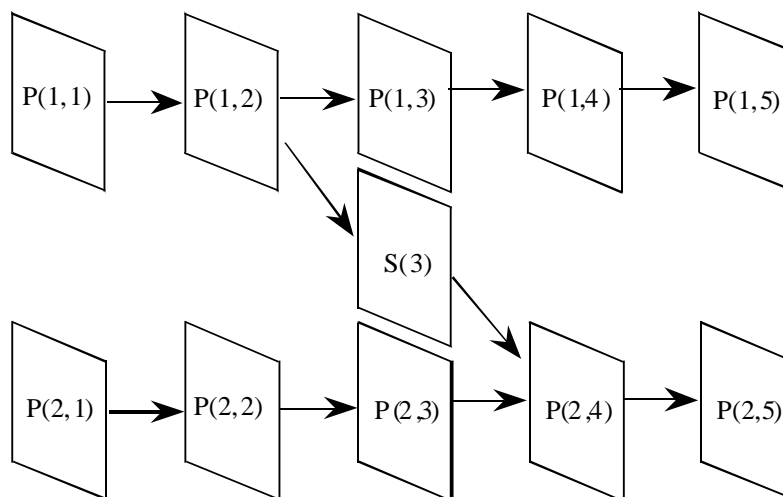


Figura 9. Cambio en la velocidad del tren de bits utilizando cuadros tipo SP

Comparación del esquema de codificación del H.264 con otros esquemas

La tabla 8 muestra en análisis comparativo de los algoritmos de codificación de video MPEG-2 y MPEG-4 parte 2 contra el H.264 (ISO/IEC JTC1/SC29/WG11 *Verification Tests on AVC*, 2003).

Tabla 8. Comparación de los estándares MPEG-2, MPEG-4 parte 2 y H.264 / MPEG-4 parte 10

Característica	MPEG-2	MPEG-4 part 2	MPEG-4 part 10/H.264
Tamaño del Macrobloque	16x16 (Modo cuadro) 16x8 (Modo campo)	16x16	16x16
Tamaño de bloques	8x8	16x16, 16x8, 8x8	16x16, 8x16, 16x8, 8x8, 4x8, 8x4, 4x4
Predicción INTRA	No	Dominio de la transformada	Dominio espacial
Transformada	DCT de tamaño 8x8	DCT-Wavelet de tamaño 8x8	8x8, 4x4 DCT entera 4x4, 2x2 Hadamard
Cuantización	Cuantización escalar con tamaño constante entre pasos del cuantizador	Cuantización vectorial	Cuantización escalar con tamaño entre pasos de 12.5% de la velocidad binaria
Codificación de la entropía	VLC	VLC	VLC, CAVLC, CABAC
Exactitud de las muestras de la imagen	1/2 muestra	1/4 muestra	1/4 muestra
Cuadros de referencia	1 cuadro	1 cuadro	Múltiples cuadros
Modo de predicción bidireccional	adelante / atrás	adelante / atrás	adelante / atrás atrás / atrás backward / backward
Predicción con peso	No	No	Si
Filtro de desbloqueo	No	No	Si
Tipo de cuadros	I, P, B	I, P, B	I, P, B, SI, SP
Perfiles	5 perfiles	8 perfiles	7 perfiles
Acceso aleatorio	Si	Si	Si
Resistencia a errores	Particionamiento de datos, FEC para transmisión de paquetes por importancia	Sincronización, particionamiento de datos, extensión de encabezados, VLCs reversibles	Particionamiento de datos, ajuste de parámetros, orden flexible de macrobloques, rebanadas redundantes, rebanadas tipo SP y SI
Velocidad de transmisión	2-15 Mbps	64kbps - 2Mbps	64kbps - 150Mbps
Complejidad del codificador	Mediana	Mediana	Alta
Compatibilidad con estándares previos	Si	Si	No

Eficiencia de codificación

La tabla 9 muestra las comparaciones del H.264 perfil Línea Base (BP) del H.264 contra el Perfil Simple del MPEG-4 parte 2 (SP) para secuencias con definición de multimedia (MD). Los números en las tablas indican la mejora en la eficiencia de la codificación. Por ejemplo, > 2x significa que la eficiencia del H.264 fue más de dos veces la eficiencia del MPEG-4 parte 2 Perfil Simple (ISO/IEC JTC1/SC29/WG11, 2000) y para la velocidad binaria de la secuencia indicada. La letra T significa que el H.264 fue transparente en la velocidad binaria para la secuencia dada. Estadísticamente el H.264 BP tuvo una mejora en la eficiencia de la codificación de 2 veces o más, en 14 de 18 casos probados.

La tabla 10 muestra la comparación del H.264 Perfil Principal (MP) y el MPEG-4 Parte 2 Perfil Simple Avanzado (ASP) para

secuencias MD (ITU-T H.264, 2003). Estadísticamente el H.264 BP tuvo una eficiencia de codificación de 2 veces o más que el MPEG-4 parte 2 ASP, en 18 de 25 casos probados.

La tabla 11 muestra la comparación entre el H.264 Perfil Principal (MP) y el MPEG-2 para definición estándar (SD) (ITU-T and ISO/IEC, 1994). Estadísticamente el H.264 MP, tuvo una eficiencia de codificación de 1.5 veces o más que el MPEG-2HiQ, en 8 de 12 casos probados, de los cuales, en 3 casos el H.264 mostró una mejora en la eficiencia de la codificación de 2 veces o más y en 1 caso mostró una mejora de 4 veces mayor.

Cuando se comparó contra el modelo de prueba 5 del MPEG-2 (MPEG-2 TM5), estadísticamente, el H.264 Perfil Principal tuvo una eficiencia de codificación de 1.8 veces o más en 9 de 12 casos probados, de los cuales, en 2 casos mostró una mejora de 4 veces mayor.

Tabla 9. Comparación del H.264 Perfil Línea Base y el MPEG-4 Parte 2 Perfil Simple para secuencia con definición MD

Secuencia	Velocidad binaria [kbps] para Q CIF				Velocidad binaria [kbps] para CIF			
	24	48	96	192	96	192	384	768
Foreman	> 1 x	2x	2x	T	2x	> 2x	T	T
Paris	> 1 x	2x	2x		2x	2x	T, 2x	T
Head		> 2x	2x		2x		T	T
Zoom	> 1 x	1x	2x		2x			

Tabla 10. Comparación del H.264 Perfil Principal y el MPEG-4 Parte 2 Perfil Simple Avanzado para secuencias con definición MD

Secuencia	Velocidad binaria [kbps] para Q CIF				Velocidad binaria [kbps] para CIF			
	24	48	96	192	96	192	384	768
Football	2x / 1 x	2x	2x		> 1x	> 1x	1x	> 1x
Mobile	2x / 1 x	2x	2x		> 2x	4x	> 2x	T
Husky	2x	2x	> 1x		2x	2x	2x	
Tempete	2x	2x	> 2x	T	2x	2x	T, 2x	T

Tabla 11. Comparación del H.264 MP contra el MPEG-2 para definición estándar (SD)

Secuencia	Velocidad binaria [Mbps] para el MPEG-2 HiQ					Velocidad binaria [Mbps] para MPEG-2 TM5				
	1.5	2.25	3	4	6	1.5	2.25	3	4	6
Football	> 1.5x	> 1.3x	1.3x	1.5x		2x	1.8x	1.3x	1.5x	
Mobile	4x	2.7x	2x	T	T	> 4x	> 2.7x	> 2x	T	T
Husky	> 1.5x	1.3x	1x / 1.3x	1.5x		2.7x / 2x	1.8x	2x	> 1.5x	
Tempete	T, 2x	T	T	T	T	T, 4x	T	T	T	T

La tabla 12 muestra la comparación entre el H.264 Perfil Principal y el MPEG-2 para alta definición (HD). Estadísticamente, el H.264 Perfil Principal tuvo una eficiencia de codificación de 1.7 veces mayor que el MPEG-2 en 7 de 9 casos probados, de los cuales, en 3 casos mostró mejoras de hasta 2 veces y en 1 caso de hasta 3.3 veces.

Cuando se comparó el modelo de prueba 5 del MPEG-2 (MPEG-2 TM5), estadísticamente, el H.264 Perfil Principal, tuvo una eficiencia de codificación 1.7 veces o mayor en 8 de 9 casos, de los cuales, en 4 casos mostró una mejora de 2 veces mayor.

Codificador de audio

El H.264/MPEG-4 parte 10, está siendo adoptado por muchas compañías de video tales

como Phillips, Polycom, Ligos, Broadcom, Netvideo, Motorola, STMicroelectronics entre otras. El estándar se refiere únicamente a la codificación de video, dejando la libertad a las compañías de elegir sus propios codificadores de audio. Por ejemplo, la empresa europea DVB está considerando adoptar AAC-SBR (Advanced Audio Coding - Spectral Band Replication) llamado formalmente AAC plus, mientras que la empresa americana ASTC utilizará AC-3 plus de los laboratorios Dolby Labs.

Además del H.264, China ha desarrollado su propio estándar de video llamado AVS (Audio Video Standard de China), el cual es un algoritmo similar, con algunas diferencias en los tamaños de las transformadas, codificadores de entropía, filtros de desbloqueo, filtros de interpolación para la compensación de movimiento, etc.

Tabla 12. Comparación del H.264 MP y el MPEG-2 para HD

Secuencia		Velocidad binaria [Mbps] para el MPEG-2 HiQ			Velocidad binaria [Mbps] para el MPEG-2 TM5		
		6	10	20	6	10	20
720 (60p)	Crew	1.7x	2x	T	1.7x	2x	T
	Harbour	T, 3.3x	T	T	T, 1.7x	T	T
	Stockholm		1x			2x	
1080 (30i)	Pan						
	New Mobile & Calendar		T, 2x	T		T, 2x	T
1080 (25p)	River Bed	> 1.7x	> 1x	T	> 1.7x	> 1x	T
	Vintage Car	1.7x	T, 2x	T	1.7x	T, 2x	T

Conclusiones

El presente artículo describe al nuevo estándar de compresión de video H.264/MPEG-4 parte 10, el cual supera a los estándares de video anteriores. Las principales diferencias se encuentran en la estructura de los bloques funcionales del algoritmo. Por ejemplo, compensación de movimiento en bloques de tamaño variable, interpolación para exactitud fraccional, filtro de desbloqueo adaptivo, rebanadas tipo SI y SP, mayor resistencia a los errores que los estándares anteriores, transformada de 4×4 , predicción con carga, CABAC, CAVLC y predicción direccional para codificación INTRA.

La tendencia actual de las empresas dedicadas al video se está moviendo a favor de la implementación del nuevo estándar para las diferentes aplicaciones como son, cámaras digitales, transmisión de video sobre redes IP, teléfonos celulares, transmisión de video satelital, entre otras.

El siguiente paso que está dando el grupo de expertos, es coordinar los términos de las licencias para que las compañías manufactureras paguen las regalías correspondientes por el codificador-decodificador y los términos de las licencias.

Referencias

- Flierl M. and Girod B. (2003). Generalized B Picture and the Draft H.264/AVC Video-Compression Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 587-597, Julio.
- Ghanbari M. (2003). *Standard Codecs: Image Compression to Advanced Video Coding*. Hertz, UK: IEE.
- Golomb S.W. (1966). Run-Length Encoding. *IEEE Trans. on Information Theory*, IT-12, pp. 399-401, Dic.
- Huang Y.W. (2005). Analysis, Fast Algorithm, and VLSI Architecture Design for H.264/AVC Intra Frame Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15, pp. 378-401, Marzo.
- ISO/IEC 11172-2:1993 (1993). Information Technology, Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1,5 Mbit/s-Part 2: Video.
- ISO/IEC JTC1/SC29/WG11, "ISO/IEC 13818-2: (1994). Information Technology-Generic Coding of Moving Pictures and Associated Audio Information: Video, ISO/IEC.
- ISO/IEC JTC1/SC29/WG11 (2003). Report of the Formal Verification Tests on AVC (ISO/IEC14496-10 | ITU-T Rec. H.264)", MPEG2003/N6231, diciembre.
- ISO/IEC JTC1/SC29/WG11, ISO/IEC 14496:2000-2: (2000). Information on Technology-Coding of Audio-Visual Objects-Part 2: Visual, ISO/IEC.
- ITU-T H.262 (2000). International Telecommunication Union, Recommendation: Generic Coding of Moving Pictures and Associated Audio Information: Video, ITU-T.
- ITU-T H.264 (2003). International Telecommunication Union, Recommendation: Advanced Video Coding for Generic Audio visual Services, ITU-T.
- ITU-T H.261 (1993). International Telecommunication Union, Recommendation ITU-T H.261: Video Codec for Audio visual Service at px64 kbits, ITU-T.
- Karczewicz M. and Kurceren R., (2003). The SP- and SI Frames Design for H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 537-544, Julio.
- List P. (2003). Adaptive Deblocking Filter. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 614-619, Julio.

- Marpe D. (2003). Context-Based Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 620-635, Julio.
- Ochoa H. and Rao K.R. (2005). A New Modified Hybrid DCT-SVD Coding System for Color Images. *WSEAS Transaction on Circuits and Systems*, Vol. 4, pp. 1246-1253, Octubre.
- Ochoa H. and Rao K.R. (2006). A New Modified Version of the HDWTSVD Coding System for Monochromatic Images. *WSEAS Transaction on Systems*, Vol. 5, pp. 1190 – 1195, Mayo.
- Richardson I. E.G. (2003). *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*, Wiley.
- Sullivan G., Topiwala P. and Luthra A. (2004). The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions. *SPIE Conference on Applications of Digital Image Processing XXVII*, Vol. 5558, pp. 53-74, agosto.
- Wenger A. (2003). H.264/AVC Over IP. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 645-656, julio.
- Wien M. (2003). Variable Block-Size Transform for H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, pp. 604-613, julio.

Semblanza de los autores

Humberto de J. Ochoa-Domínguez Estudió su licenciatura en el Instituto Tecnológico de Veracruz, su maestría en el Instituto Tecnológico de Chihuahua y su doctorado en la Universidad de Texas en Arlington, apoyado por Fulbright y PROMEP. Trabajó en el grupo de investigaciones de multimedia de Nokia, en Irving Texas, y actualmente es profesor investigador en la Universidad Autónoma de Ciudad Juárez. Ha impartido diversos cursos tutoriales de codificación de video y procesamiento de señales en Singapur, Tailandia, Estados Unidos y la República de Malta. En 1998, recibió el premio Chihuahua por el trabajo "Sistema para la Clasificación de Mamografías Digitales en Normales y Anormales Mediante el Análisis de Textura y Detección de Micro calcificaciones. Ha ofrecido conferencias en diversos foros de multi media y procesamiento de señales en varias partes del mundo y ha publicado en diferentes Journals y revistas arbitradas de reconocimiento nacional e internacional. Asimismo, es miembro de la Sociedad de Procesamiento de Señales del al IEEE .

José Mireles-García. Estudió su licenciatura en ingeniería industrial con especialidad en electrónica, así como su maestría en electrónica en el Instituto Tecnológico de Chihuahua. Obtuvo su doctorado en ingeniería eléctrica en la Universidad de Texas en Arlington (UTA), apoyado por PROMEP. Actualmente es profesor investigador en la Universidad Autónoma de Ciudad Juárez y profesor de investigación adjunto de las Universidades de Texas en Arlington y de El Paso. Es miembro del Sistema Nacional de Investigadores (SNI,) de la Sociedad de Control y de la SMC de la IEEE. Asimismo, ha sido integrado en la lista de Strathsmore's Who's Who los años 2002 y 2003. Fue visitante en la UTA de Agosto 2002 a Julio 2003, en cuyo periodo también trabajó en el Instituto de Robótica de UTA en Fort Worth TX como asistente de investigación, y participó en la recepción de fondos binacionales NSF-CONACyT para desarrollo de trabajo en sistemas de eventos discretos. Sus áreas de interés son el diseño, fabricación y prueba de los Sistemas Microelectromecánicos (MEMS,) robotización y automatización y sistemas de eventos discretos. Ha impartido conferencias en diversos foros de robótica, automatización y MEMS en varias partes del mundo y ha publicado en diferentes Jour nals y revistasarbitradas de reconocimiento nacional e internacional, así también ha participado en la organización de diversos foros nacionales e internacionales.

Juan de Dios Cota-Ruíz Estudió su licenciatura en el Instituto Tecnológico de Sonora y su maestría en el Instituto Tecnológico de Chihuahua. Trabajó en el área de diseño de equipo de prueba para la empresa SMTC de Chihuahua y RCA Thomson. Actualmente trabaja como profesor investigador en el Departamento de Ingeniería Eléctrica y Computación de la Universidad Autónoma de Ciudad Juárez en el área de Instrumentación y Procesamiento de Señales. Ha publicado en diferentes revistas arbitradas del país.