
COMPRESION DE VIDEO

©Constantino Pérez Vega
Departamento de Ingeniería de Comunicaciones

6.0 Introducción

En este capítulo se tratan los principios básicos de la compresión de vídeo, orientado principalmente a los estándares adoptados para televisión digital, particularmente MPEG-2. Sin embargo, MPEG-2 tiene aspectos comunes con otros estándares y técnicas de compresión utilizados en teleconferencia o videoconferencia, como H.261 y H.263 del CCIR. El estándar MPEG-2 ha sido adoptado mundialmente para televisión digital, pero no por ello dejan de ser de interés otros estándares utilizados ampliamente en aplicaciones de vídeo en entornos multimedia, como JPEG, MPEG-1 y los ya mencionados H.261 y H.263. Todos ellos comparten algunos principios comunes que se tratan en este capítulo. Se mencionan también otras técnicas de compresión no estándar, pero bastante utilizadas en procesado de imágenes, principalmente en aplicaciones multimedia.

El campo de la compresión de vídeo es muy extenso y no puede tratarse en esta obra con la profundidad y rigor necesarios sin entrar en el campo del procesado digital de imágenes. Aún así, es posible un tratamiento relativamente simple del tema, desde el punto de vista conceptual y suficiente para comprender el papel que desempeñan estas técnicas en el contexto de la televisión digital y aplicaciones multimedia.

Cabe mencionar también que las publicaciones sobre el tema se encuentran principalmente en revistas especializadas y son escasos los textos en el mercado. Una fuente muy útil de información es el *web*, aunque abunda la información superficial y con frecuencia, más de carácter comercial que técnica. También se dan aquí algunas de las referencias más importantes que se pueden consultar en esta fuente.

6.1 Reducción del caudal binario y compresión

Al discutir los procedimientos de reducción del caudal binario para adecuarlo a la capacidad de las redes de telecomunicación en uso, se mencionaron algunas técnicas que, para el caso de transmisión a 140 Mbit/s incluyen la supresión de los intervalos de borrado horizontal y vertical y que son fácil e íntegramente recuperables en el receptor, así como codificación predictiva simple basada en modulación por codificación diferencial de pulsos (DPCM) que, en cierta medida, aprovecha la redundancia entre elementos de imagen adyacentes. Este tipo de codificación se considera, en esta aplicación, como reversible, aunque en realidad elimina parte de la información contenida en la señal original. La información perdida es, en general, poco significativa y si los procesos de codificación-decodificación son pocos, la señal decodificada se aproxima bastante a la original y su calidad es adecuada para ser procesada en el centro de producción.

Por otra parte, la codificación a 34 Mbit/s emplea técnicas en que se elimina una porción importante del contenido de información original dando lugar a una señal *comprimida* que de alguna forma, tiene que reconstruir el decodificador. Esta señal es adecuada para algunas aplicaciones como el suministro a redes de distribución terrestre, pero no lo es para cierto tipo de procesado en el estudio, ya que aún cuando en el decodificador se reconstruya aproximadamente la señal original, sus características iniciales no pueden recuperarse totalmente.

La situación se vuelve aún más compleja si se pretende transmitir la señal digital de vídeo por un sistema de radiodifusión terrestre o cable, en que el ancho de banda de la señal total de televisión

que incluye sonido y otras señales adicionales como teletexto, no pueden ocupar un ancho de banda superior a 6 MHz (NTSC) u 8 MHz (PAL).

En las redes de distribución, que constituyen el eslabón final del sistema y en que las señales destinadas a los espectadores no son objeto de procesamiento adicional excepto quizá de grabación en magnetoscopios domésticos, se pretende básicamente que la imagen y sonido sean aceptables al observador final, para quien la calidad de la señal está determinada por las características perceptuales visuales y auditivas, de modo que es posible aprovechar la respuesta de los sistemas visual y auditivo humano para reducir la cantidad de información transmitida a fin de que las señales sean percibidas subjetivamente como imágenes y sonidos con la misma calidad o cercana, a las de las señales originales. Una señal codificada a 34 Mbit/s puede recuperarse con calidad más que suficiente para un observador normal y si se parte de ella, es necesario un esquema de modulación digital con una eficiencia espectral mínima del orden de 5 bit/Hz para contenerla en un canal de 6 MHz. Esto, si bien es posible, sugiere por una parte la conveniencia de buscar reducciones aún mayores del caudal binario y, por otra, abre la posibilidad de reducir el ancho de banda de transmisión necesario, sobre todo si se tiene en cuenta que la actual congestión del espectro radioeléctrico en las bandas de televisión no es suficiente para satisfacer la demanda de programas de televisión en la sociedad.

En el caso de señales de vídeo, una sola imagen digitalizada en formato 4:2:2 sin compresión requiere una memoria del orden de 700 Kbytes para su almacenamiento, de modo que una secuencia de imágenes a 25 cuadros/s con una duración de un minuto, requeriría de una memoria de aproximadamente 20 Gbits. Tal cantidad de información no puede almacenarse en medios prácticos como discos (CD-ROM) o cintas magnéticas, de manera efectiva a costos razonables y es necesario pensar en técnicas que reduzcan este volumen de datos a valores manejables por los dispositivos actuales. A esto se agrega la necesidad de procesar toda la información en tiempo real o cuasi-real, con la complejidad que ello implica en el hardware y software necesarios. En la tabla 6.1 se muestran algunos valores de los caudales o flujos binarios tanto sin compresión, como comprimidos, para algunas aplicaciones de audio y vídeo.

Aplicación	Caudal binario sin compresión	Caudal binario comprimido
Voz. 8 Kmuestras/s, 8 bits/muestra	64 Kbit/s	2.4 Kbit/s
Audio (voz y/o música). 8 Kmuestras/s, 8 bits/muestra	128 Kbit/s	6 a 64 Kbit/s
Audio estéreo (digital) 46.1 Kmuestras/s, 16 bits/muestra	1.5 Mbit/s	128 a 768 Kbit/s
Videoconferencia 15 cuadros/s. 352 × 240 pixels, con 24 bits/pixel	30.4 Mbit/s	64 a 768 Kbit/s
Vídeo digital (CD-ROM) 30 cuadros/s. 352 × 240 pixels con 24 bits/pixel	60.8 Mbit/s	1.5 a 4 Mbit/s
Vídeo “estándar” NTSC 525 líneas/cuadro, 30 cuadros/s 720 × 480 pixels con 24 bits/pixel	249 Mbit/s	3 a 8 Mbit/s

Tabla 6.1 Valores de los caudales o flujos binarios con y sin compresión

La televisión “convencional” por sí sola no fue el único motor para el desarrollo de las diversas técnicas de compresión de imágenes. La necesidad de almacenar grandes volúmenes de materiales visuales en cinta o disco compacto, como películas, fotografías, radiografías, libros, vídeo juegos, etc., cuyo procesado no tiene que hacerse necesariamente en tiempo real, ha sido un factor importante en la investigación y desarrollo de técnicas de compresión cada vez más eficientes. De hecho, las primeras técnicas de compresión se orientaron al almacenamiento de imágenes fijas, entre ellas la definida en el estándar JPEG y posteriormente, MPEG-1 orientado a imágenes fijas o en movimiento, con sonido asociado, para almacenamiento en CD-ROM. Ello hizo posible almacenar una enciclopedia completa en un solo disco, lográndose caudales binarios para la transmisión de imágenes tan bajos como 1.5 Mbit/s o menos sin deterioro importante en la calidad subjetiva.

6.1.1 Definiciones relacionadas con la compresión de vídeo.

Compresión sin pérdidas. Se refiere a los métodos de compresión en los que la calidad de la señal decodificada es, al menos teóricamente, igual a la calidad de la señal de la fuente.

Compresión con pérdidas. En la práctica del manejo de imágenes y sonido, son relativamente pocas las aplicaciones en que es necesaria una reconstrucción absolutamente fiel de la información de la fuente, ya que intervienen aspectos perceptuales que son aprovechables para reducir la cantidad de información que debe transmitirse y que, se admite, puede ser reconstruida aproximadamente por el decodificador. Por consecuencia, en los sistemas de compresión con pérdidas es admisible la pérdida de cierta cantidad de información que no es relevante para el observador final. En estas condiciones, se pretende que la señal tenga la máxima calidad subjetiva, es decir que resulte aceptable al observador.

El proceso de compresión con pérdidas es irreversible, es decir que no es posible recuperar la información original a partir de la información comprimida. Tal irreversibilidad da lugar a una rápida degradación de la calidad de señal si la compresión se aplica de forma concatenada, es decir, la realización de compresiones sucesivas sobre señales decodificadas.

Relación de compresión. Se define como:

$$c_r = \frac{\text{flujo binario a la entrada del codificador}}{\text{flujo binario a la salida del codificador}}$$

Eficiencia de codificación. Por lo general, se expresa en bits por muestra o en bits por segundo y está limitada por el contenido de información o *entropía* de la fuente. Cuanto mayor sea ésta, más difícil y compleja será la compresión.

Complejidad de codificación. La complejidad del proceso de compresión tiene que ver directamente con la carga o esfuerzo de cómputo para implementar las funciones de codificación y decodificación. Esto afecta tanto al hardware como al software y, por lo general, se mide en función de los requerimientos de memoria y de la cantidad de operaciones aritméticas necesarias, expresada en *millones de operaciones por segundo* (MOPS) o bien de *millones de instrucciones por segundo* (MIPS). La complejidad de codificación también está relacionada con el consumo de potencia del codificador y con el tiempo requerido para realizar el proceso de compresión, al que se designa como *retardo de codificación*.

Retardo de codificación. Todo proceso de compresión requiere de un cierto tiempo que, dependiendo de la aplicación, puede resultar o no, crítico. Hay aplicaciones de compresión que pueden llevarse a cabo “fuera de línea”, es decir en tiempo no real, como el procesado de imágenes médicas o las que generan algunos sistemas de percepción remota. Esto no es posible en televisión,

donde todo el procesado debe hacerse en tiempo real o cuasi-real de modo que el retardo introducido por los procesos de codificación y decodificación no sea perceptible al observador. El retardo de codificación puede reducirse aumentando la potencia de cómputo, lo que aumenta la complejidad de hardware y software y, por consecuencia, del costo.

Calidad de señal. En general, este concepto se aplica a la señal de salida del decodificador y, en realidad, no hay un criterio cuantitativo aceptado universalmente para definirlo. En alguna literatura, se define mediante una relación señal a ruido dada por:

$$SNR = 10 \log_{10} \frac{\text{energía de la señal a la entrada del codificador}}{\text{energía de ruido de la señal}}$$

Entendiéndose aquí por energía de ruido de la señal, como la energía medida para una señal hipotética, igual a la diferencia entre la señal de entrada al codificador y la señal de salida del decodificador. La definición anterior se refiere a niveles medios de energía y, en el caso de imágenes de vídeo, se utilizan los valores pico, lo que da lugar a una relación de señal pico a ruido (PSNR), en lugar de SNR.

Este criterio para definir la calidad de la señal no resulta adecuado desde un punto de vista perceptual en lo que concierne a la señal de televisión ya que los valores altos de SNR o PSNR no siempre se corresponden con señales de alta calidad perceptual. En la práctica, suele preferirse el resultado de sondeos de la opinión de espectadores. Es decir, la calidad de la señal suele medirse en términos subjetivos en la forma mencionada en la sección 1.7.

Tanto en compresión sin pérdidas como con pérdidas, la calidad de los datos de entrada afecta la relación de compresión. Así, el ruido en la adquisición de los datos, errores en el tiempo de muestreo y los propios procesos de conversión analógico digital, afectan la calidad de la señal y reducen la correlación espacial y temporal de las muestras dando lugar a que algunos esquemas de compresión sean muy sensibles a la pérdida de correlación y den peores resultados con señales ruidosas.

Caudal binario a la salida del codificador. Según la forma de implementar el codificador, las características del caudal binario a la salida del codificador pueden dar lugar a los siguientes casos:

- a) *Retardo fijo de codificación y flujo constante de salida.* La eficiencia de codificación varía según las muestras y, por consecuencia, también la calidad de la señal.
- b) *Procesado de muestras en paralelo.* Con esta técnica es posible mejorar la eficiencia de codificación dando lugar a símbolos de longitud fija a la salida, pero a intervalos irregulares dependiendo de la cantidad de información contenida en cada grupo de muestras. También es posible obtener a la salida símbolos de longitud variable a intervalos fijos.
- c) *Eficiencia de codificación fija.* En este caso, la calidad de la señal varía en función de la cantidad de información contenida en las muestras a comprimir. Este enfoque puede resultar adecuado para emplearse en sistemas de comunicaciones en que la velocidad de transmisión no debe exceder un cierto valor, o para aplicarse a medios de almacenamiento de capacidad fija. Este tipo de codificadores se designan como de *flujo o caudal binario constante*.
- d) *Calidad constante y eficiencia de codificación variable.* Si los medios de transmisión o de almacenamiento no restringen la velocidad de transmisión, o bien si es deseable una calidad determinada de señal a la salida del decodificador, puede permitirse que varíe la eficiencia de codificación para mantener constante dicha calidad. Estos esquemas de codificación se designan como de *caudal o flujo variable* y también como de *calidad constante*.

6.2 Compresión basada en la redundancia de información

La principal característica común a todos los sistemas de compresión, es la explotación de la redundancia en la señal, de modo que la función básica del codificador es eliminar dicha redundancia antes de transmitirla o almacenarla y que, en el decodificador deberá reinsertarse de nuevo. Se aprovechan también las características psicofisiológicas de la visión humana que hacen posible que un observador “normal” encuentre totalmente aceptable una imagen reproducida que no sea exactamente igual a la original, ya que el sistema visual humano ofrece cierta tolerancia a la distorsión, dependiendo de las características de la imagen y de las condiciones de visualización. Entre las características de la visión humana que se aprovechan para reducir el caudal de información se encuentra la respuesta limitada a las señales de luminancia y color así como la menor sensibilidad de la visión al ruido en áreas brillantes de la imagen o en altas frecuencias, tal como el que se puede presentar en los bordes de un objeto.

Es fácil observar que en una escena televisada la imagen entre dos cuadros sucesivos varía muy poco. La mayor variación ocurre en las porciones de la imagen en que hay movimiento o cuando hay cambios de escena. Así, en las imágenes televisadas puede hablarse de *redundancia temporal* puesto que los cuadros sucesivos tienen mucha información igual entre ellos. Por otra parte, entre los elementos de imagen que constituyen una línea, ya sea que se les considere horizontal o verticalmente hay también semejanzas que dan lugar a *redundancia espacial*.

Es posible formar una idea conceptual simple de la compresión si se piensa en una imagen fija, para la que bastaría transmitir únicamente un cuadro completo y hacer que el receptor lo almacene y reproduzca los cuadros sucesivos iguales sin necesidad de continuar recibiendo la misma información del transmisor. El canal de comunicación sólo estaría ocupado en este caso, el tiempo necesario para transportar la información de un cuadro. En el caso de imágenes en movimiento sólo sería necesario transmitir el cuadro inicial y luego, la información correspondiente a las variaciones entre un cuadro y el precedente sin necesidad de repetir la información fija o redundante. Esta idea es la que subyace en el desarrollo de los diversos sistemas de compresión.

La señal digital de vídeo está constituida por una serie de muestras representadas por la información numérica correspondiente a los valores de luminancia y crominancia de cada elemento de imagen. Los valores de las muestras son en cierta medida predecibles, ya que éstas no son totalmente independientes y hay cierta semejanza o correlación entre muestras de elementos de imagen adyacentes a lo largo de una línea, entre líneas adyacentes y también, entre campos o cuadros adyacentes. En una escena de vídeo típica, la correlación entre muestras adyacentes de luminancia o crominancia es elevada, lo que sugiere la posibilidad de representar una imagen en términos de los cambios entre elementos vecinos utilizando únicamente un porcentaje bajo, por ejemplo 10%, del contenido de información original. En la práctica, la sola eliminación de redundancia espacial produce niveles muy modestos de compresión, usualmente inferiores a 10:1, muy lejanos a los requeridos para almacenamiento en CD-ROM, del orden de 120:1 o entre 30:1 y 50:1 para transmisión de imágenes de televisión en canales de 6 a 8 MHz de ancho de banda.¹

Los métodos de compresión de imágenes pueden clasificarse según dos categorías diferentes: una, en que se aplican técnicas que aprovechan la redundancia en los datos, por ejemplo suprimiendo los intervalos de borrado horizontal y vertical, que pueden recuperarse con facilidad en el receptor y otra, en que la compresión se logra mediante alguna transformación que preserve la energía de la señal de imagen y que es capaz de “empaquetar” la información en un mínimo de muestras. Estas dos categorías dan lugar a una clasificación alternativa de las técnicas de compresión en *reversibles* e *irreversibles*. En las primeras no se pierde información en el proceso de codificación-decodificación, en tanto que en las técnicas irreversibles, se pierde cierta cantidad de información durante la codificación, que no es posible recuperar en el decodificador. Estas últimas son las que

¹ Nelson, L. J. “Video Compression”. Broadcast Eng. Vol. 37, N° 11, pp. 42-46. Oct. 1995.

permiten alcanzar mayores niveles de compresión porque reducen considerablemente el caudal binario de la señal original.

Entre las técnicas de codificación reversible de la señal de vídeo, pueden mencionarse las de supresión de los intervalos de borrado horizontal y vertical y la codificación estadística que incluye, entre otras, la codificación vectorial, codificación de recorrido (runlength) y codificación de longitud variable, en tanto que las de codificación irreversible incluyen las de submuestreo de Nyquist, PCM diferencial (DPCM), codificación en subbandas, codificación por transformadas, etc. La mayor parte de las técnicas de compresión empleadas en televisión, tanto para producción y grabación como para transmisión son irreversibles y causan degradación de la imagen reconstruida.

Aún cuando puede considerarse que las técnicas de compresión están todavía en su infancia, se han desarrollado ya varios estándares que comienzan a utilizarse extensamente, tanto para almacenamiento como para transmisión. La mayor parte de dichos estándares, entre ellos los de la familia MPEG², se basan en la *transformada del coseno discreto* (DCT), aunque también se emplean otras técnicas, como las de wavelets³ y fractales⁴.

La compresión mediante wavelets y subbandas emplea una transformada similar a la DCT que ofrece dos ventajas sobre ella. Primero, la degradación es más suave y aunque se pierde detalle al disminuir el caudal binario, no presenta algunos de los problemas inherentes a las imágenes codificadas mediante la DCT. En segundo lugar y probablemente más importante, es la escalabilidad, que permite el almacenamiento de datos con la máxima relación entre la calidad de la imagen y la velocidad de transmisión.

La compresión mediante fractales es promisoria, pero es asimétrica⁵, ya que la codificación requiere un volumen de procesamiento considerablemente mayor que la decodificación, aunque este problema no es único de los fractales y puede reducirse con el futuro desarrollo de circuitos integrados de aplicación específica (ASIC). Los fractales son inherentemente escalables, lo que permite ver una imagen sin degradación a diferentes resoluciones y además, los fractales parecen comportarse mejor con imágenes naturales.

Entre las aplicaciones de las técnicas anteriores, cabe mencionar que el FBI eligió la técnica de compresión mediante wavelets para almacenar su inmenso catálogo de huellas dactilares y las técnicas de fractales se han empleado para almacenar imágenes fijas de alta calidad en CD-ROM. Sin embargo, en la mayoría de las aplicaciones, la elección más clara en lo que atañe a la televisión es alrededor de los estándares MPEG, entre otras razones porque se han desarrollado y comercializado gran cantidad de circuitos integrados y equipos para ellos, cuyos precios están reduciéndose rápidamente.

6.2.1 Aspectos a considerar en la selección del método de compresión

- a) **Compresión sin pérdidas o con pérdidas.** Condicionado por la aplicación, calidad de señal requerida y complejidad de codificación.
- b) **Eficiencia de codificación.** Puede estar condicionada por el ancho de banda del sistema de comunicaciones o por la capacidad de memoria disponible para la señal comprimida deseada. Eventualmente, aún en el caso de compresión sin pérdidas, es posible que no

² Moving Picture Expert Group

³ Un término adecuado en español sería *onditas*. Aquí se ha preferido utilizar el término original en inglés.

⁴ Blumenfeld, S. "Video Compression". Broadcast Eng. Vol. 37, N° 2, pp. 24-32. Feb. 1995.

⁵ Se entiende por compresión o codificación asimétrica, aquella en que el tiempo y complejidad del proceso de codificación difiere considerablemente del de decodificación.

se alcance la eficiencia de codificación deseable, particularmente cuando hay requerimientos mínimos respecto a la calidad de la señal decodificada.

- c) **Complejidad de codificación.** Siempre es deseable que la complejidad sea la mínima posible para una calidad dada de señal.
- d) **Robustez ante errores.** Aunque el tratamiento de errores forma parte más bien de la codificación de canal que de la fuente (compresión), algunos métodos de compresión son más sensibles a errores que otros. El resultado es que la señal decodificada dará lugar a artefactos⁶ visibles que pueden ser más o menos molestos para el observador. El grado en que tales artefactos resultan objetables también varía según la aplicación, por ejemplo, el criterio puede ser muy diferente si se trata de televisión convencional o de transmisión de imágenes médicas, policíacas, científicas, etc.
- e) **Degradación de la señal en procesos concatenados de codificación y decodificación.** Este aspecto es de particular importancia en el entorno de centros de producción.
- f) **Compatibilidad con otros medios.** Los procesos de compresión en la actualidad, no sólo tienen que ver con televisión, sino con computadoras, en lo que se ha dado en llamar sistemas "multimedia".

6.3 Compresión de vídeo en el contexto del sistema digital de televisión.

La compresión de vídeo, y también de audio, pueden definirse, básicamente, como métodos de *codificación de fuente* con los que se pretende obtener la máxima eficiencia en la codificación de la señal analógica original. Sin embargo, desde el punto de vista de transmisión y dado el elevado nivel de compresión utilizado en televisión, es de gran importancia reducir al mínimo la posibilidad de ocurrencia de errores durante la transmisión mediante una *codificación de canal* adecuada, que permita al receptor corregir los errores que sufra la señal como consecuencia de los efectos de los diversos mecanismos de propagación. Así, la señal comprimida en el codificador de fuente debe pasar luego por un codificador de canal, cuyo propósito es proporcionarle la robustez necesaria para hacerla inmune a las inevitables degradaciones que sufrirá en su transporte hasta el receptor.

En el receptor, la señal pasa primero por un *decodificador de canal*, cuya función es detectar y corregir los errores en la señal recibida. La señal corregida se aplica, a continuación, al *decodificador de fuente*, en el que se realiza el proceso de descompresión para obtener, finalmente, una señal digital de vídeo que corresponde a la señal digital original y, de la cual puede obtenerse la señal analógica adecuada para su visualización. Este proceso se ilustra en la figura 6.1

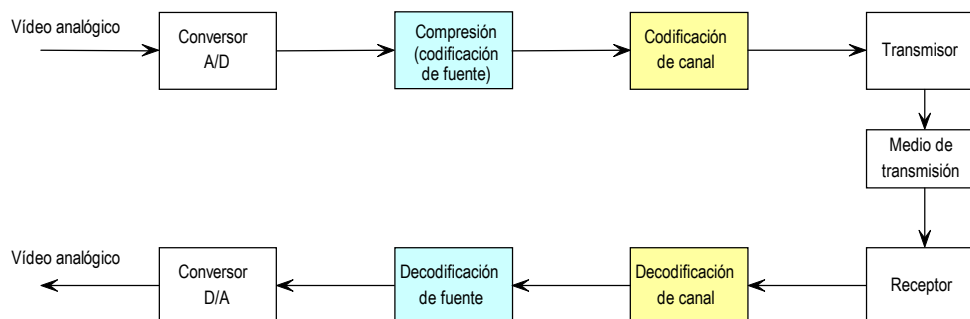


Fig 6.1. Contexto de la compresión en el sistema digital de televisión.

⁶ Se designan como *artefactos* de la imagen a degradaciones de partes de la imagen, ya sea por ruido, interferencia o errores irreversibles en la decodificación.

Los sistemas de compresión mencionados corresponden a la codificación de fuente y no deben confundirse con la codificación de canal que será tratada en el capítulo 8.

6.4 Estándares de compresión de vídeo.

Independientemente de las diversas técnicas de compresión de vídeo que se han desarrollado y de las cuales algunas, como las de wavelets o fractales han encontrado aplicación en campos específicos, aquí son de interés únicamente aquellas destinadas a la compresión de imágenes de televisión, ya sea para su empleo en centros de producción, para transmisión, bien con fines de contribución o distribución y para almacenamiento ya sea en medios magnéticos, CD-ROM u otros. Entre las más importantes se cuentan las siguientes:

- a) *H.261* del CCITT, para aplicaciones en videoconferencia, en que las imágenes tienen escaso movimiento. Las velocidades de transmisión que permite este estándar son, básicamente, 64 Kbit/s, 384 Kbit/s y 1.5 Mbit/s.
- b) *JPEG*. Destinado principalmente a la codificación de imágenes fijas para almacenamiento en CD-ROM o medios magnéticos. Permite elevados niveles de compresión, hasta de 100:1 y está basado en la transformada del coseno discreto (DCT). El algoritmo es simétrico, lo que hace que el tiempo necesario para la codificación sea prácticamente el mismo que el requerido en la decodificación. Al tratar sólo imágenes fijas, no es necesaria la compensación de movimiento y el estándar básico no contempla la codificación de audio.
- c) *CCIR-723*. Es una Recomendación del CCIR (UIT-R) para transmisión a larga distancia de señales de vídeo en componentes, digitalizadas en el formato 4:2:2. Las velocidades de transmisión que contempla este estándar son 34, 45 y 140 Mbit/s.
- d) *MPEG-1*. Es un estándar desarrollado para imágenes en movimiento, basado en la DCT, que emplea compensación de movimiento⁷. Está orientado básicamente al almacenamiento de imágenes en CD-ROM y ofrece calidad equivalente a la del sistema VHS analógico. El estándar enfoca también la compresión de audio. Las velocidades de transmisión son del orden de 1.2 Mbit/s, con audio a 128-334 Kbit/s.
- e) *MPEG-2*. Constituye, de hecho una evolución de MPEG-1 para imágenes de barrido entrelazado, orientado a aplicaciones de televisión con fines de distribución, en que se requiere alta calidad subjetiva. Las velocidades de transmisión con MPEG-2 varían, aproximadamente, entre 3 y 8 Mbit/s. El esquema de compresión de audio es similar al de MPEG-1.

Dependiendo del esquema de modulación utilizado, una señal comprimida con el estándar MPEG-2 puede ocupar un ancho de banda del orden de 1.5 MHz o menos, lo que permite empaquetar hasta cuatro o más canales digitales en el ancho de banda de un canal analógico de 6 MHz. En un futuro cercano, esto permite aumentar la capacidad de los sistemas de transmisión terrestre hasta en cuatro veces y seguramente traerá como consecuencia la reasignación del espectro radioeléctrico destinado a estos servicios.

El estándar MPEG-2 ha sido adoptado internacionalmente para utilizarse en televisión, tanto en los Estados Unidos como en Europa y el resto del mundo. Sin embargo, esto no significa que los sistemas de transmisión de televisión vayan a ser compatibles, ya que se han adoptado estándares

⁷ La compensación de movimiento es una técnica de codificación de los segmentos de un cuadro de la imagen, teniendo en cuenta sus desplazamientos en cuadros sucesivos.

de transmisión diferentes, uno el estadounidense, definido por el ATSC (American Television Standards Committee) y adoptado también en algunos otros países americanos otro, el europeo, designado como DVB (Digital Video Broadcasting) y adoptado prácticamente en el resto de los países en que se están implantando sistemas de televisión digital y finalmente el japonés, designado como ISDB (Integrated Services Digital Broadcasting), similar en muchos aspectos al DVB. Estos estándares de transmisión tienen que ver, entre otros aspectos, con el tipo de modulación del canal de RF. Además, aunque en ambos sistemas la codificación de fuente para el vídeo se realiza en el mismo estándar MPEG-2, la codificación de fuente para el audio es diferente. El sistema DVB realiza la codificación de audio de acuerdo a las recomendaciones de MPG-2, pero el sistema ATSC emplea una codificación totalmente diferente.

6.5 Técnicas de compresión MPEG

Algunas de las principales características que distinguen a los estándares MPEG de otros métodos de compresión son:

- a) MPEG está orientado a vídeo y su formato hace uso de un único espacio de color (Y, C_r, C_b), así como de un rango limitado de resoluciones y relaciones de compresión. Tiene además mecanismos para el manejo del sonido.
- b) Aprovecha el alto grado de correlación entre las imágenes de una secuencia de vídeo, así como la naturaleza predictiva del movimiento.
- c) Proporciona un caudal binario constante mediante el empleo de variables ajustables, haciendo que el formato sea predecible en relación con los requerimientos de ancho de banda.
- d) Especifica la sintaxis para el almacenamiento y transmisión de los datos comprimidos y define el proceso de decodificación. Sin embargo, los estándares no especifican cómo debe llevarse a cabo la codificación, dejando las consideraciones para su implementación a los fabricantes de codificadores. El requisito es que todos los codificadores deben producir caudales binarios MPEG válidos para ser descomprimidos por cualquier decodificador, lo que en la práctica, da lugar a una amplia variedad de implementaciones posibles para los codificadores.

El estándar MPEG-1 está enfocado a la codificación de señales digitales de audio y vídeo y al sincronismo de los correspondientes caudales binarios. En principio, este estándar maneja imágenes con barrido no entrelazado y fue desarrollado principalmente para propósitos de almacenamiento en CD-ROM a caudales binarios del orden de 1.5 Mb/s, en tanto que el estándar MPEG-2 está orientado específicamente a televisión y puede manejar imágenes con barrido entrelazado a caudales binarios mayores, codificándolas como campos individuales o como cuadros completos. En MPEG-1 no está definida la relación entre cuadros y campos de la señal analógica de vídeo y se asume que las imágenes bidimensionales pueden obtenerse a partir, bien sea del formato de campo o del de cuadro.

Otros estándares de la misma familia son MPEG-3 y MPEG-4. El primero enfocado a imágenes de alta definición. Sin embargo, las especificaciones básicas de MPEG-3 quedaron absorbidas en MPEG-2 y por consecuencia, se ha abandonado. No ocurre lo mismo con MPEG-4, orientado a conseguir caudales binarios muy bajos, tales como los empleados para teleconferencia y otras aplicaciones relacionadas, que requieren una elevada relación de compresión y gran eficiencia de codificación.

Aún cuando hay numerosos aspectos relativos a la compresión de información y a la codificación de imágenes que por su extensión y detalle no pueden ser tratados aquí, uno de los aspectos claves

para la compresión de imágenes en movimiento es la similitud entre cuadros adyacentes de una secuencia de imágenes. En MPEG-1 la primera operación que se realiza para la compresión de la información es el *diezmado*⁸. El codificador reduce la información tanto horizontal como verticalmente aproximadamente a la mitad eliminando elementos de imagen en ambas direcciones. Las componentes de crominancia (U y V) se someten a un diezmado adicional, basándose en la resolución limitada de la visión humana para las señales de color.

El codificador recibe como entradas las señales digitales de luminancia (Y) y crominancia o diferencia de color (U y V). Si en una secuencia⁹ de imágenes no hay movimiento, ésta puede codificarse como una sola imagen seguida de unos cuantos bits que sirvan para indicar al decodificador que repita la imagen. Ahora bien, es necesaria alguna técnica para codificar la primera imagen de una secuencia sin referencia a otras imágenes, que debe aplicarse también a intervalos regulares en una misma secuencia, es decir, las imágenes de una misma secuencia se codifican ocasionalmente de forma individual, independientemente de imágenes contiguas o vecinas, para evitar la propagación de errores y, eventualmente, facilitar la tarea de edición. Los estándares MPEG tratan, además de las imágenes individuales, secuencias de imágenes necesarias para el proceso de compresión.

Teniendo en cuenta la semejanza entre cuadros adyacentes, puede lograrse un nivel alto de compresión si la mayor parte de los cuadros se codifican como *diferencias* entre cuadros adyacentes, lo que puede llevarse a cabo de varias formas dependiendo del tipo de diferencia. En zonas sin diferencias significativas, la información simplemente se copia del cuadro previo. En zonas con movimiento escaso o moderado, la información se *predice* a partir de los cuadros adyacentes codificando los errores de predicción y finalmente, en algunos casos es necesario aplicar técnicas de compensación de movimiento. Un principio inviolable en la compresión de datos, bien sean de vídeo, audio u otro tipo, es que las predicciones sólo pueden hacerse con la información disponible en el decodificador. Es decir, el decodificador únicamente hace lo que le indica la información que le suministra el codificador a través del canal de comunicación. Por ello, el codificador debe empaquetar la información de manera adecuada para que el decodificador actúe correctamente.

6.6 Jerarquía de los datos de vídeo

La unidad fundamental de proceso en los estándares MPEG es el *bloque*, constituido por un grupo de 8×8 elementos de imagen. Los bloques de información se organizan después en *macrobloques*, formados por cuatro bloques contiguos de luminancia, constituyendo matrices de 16×16 elementos, además de un cierto número de bloques de crominancia. En el formato 4:4:4 hay cuatro bloques de cada componente de crominancia por cada cuatro bloques de luminancia, en el formato 4:2:2 dos bloques de cada componente de crominancia por cada cuatro de luminancia y en el formato 4:2:0, un bloque de cada componente de crominancia. Este diezmado o submuestreo de la información de crominancia es justificable dada la menor resolución cromática de la visión humana. En la figura 6.2 se ilustra la estructura de un macrobloque en el formato 4:2:0.

⁸ Nelson, L. J. "Video Compression". Broadcast Eng. Vol. 37, N° 11, pp. 42-46. Oct. 1995.

⁹ Se entiende aquí por *secuencia*, un conjunto de imágenes sucesivas que por lo general ocurren a intervalos temporales fijos

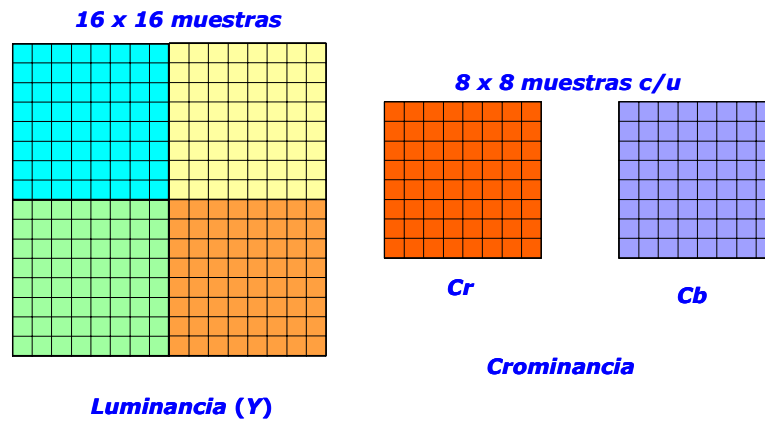


Fig. 6.2 Estructura de un macrobloque en el formato 4:2:0

Una imagen MPEG está formada por segmentos (slices¹⁰), que son secuencias de macrobloques contiguos en el mismo orden de barrido de la imagen y que comienzan en una posición particular sobre ésta, especificada por un *encabezado del segmento* (slice header). La estructura segmentada de la imagen aumenta la flexibilidad en el procesamiento y proporciona un mecanismo conveniente para el manejo de errores. La figura 6.3 ilustra un cuadro de imagen segmentado en que, como puede apreciarse esquemáticamente, los segmentos pueden tener diferentes longitudes.

Cuando ocurre un error en el caudal de datos, el decodificador puede saltar al inicio del segmento siguiente. Como consecuencia de ello, el número de segmentos afecta la eficiencia en la transferencia de los datos, ya que si bien la segmentación favorece la ocultación de errores, por otra parte emplea bits necesarios para la identificación de los segmentos que de otra forma podrían utilizarse para aumentar la calidad de la imagen.

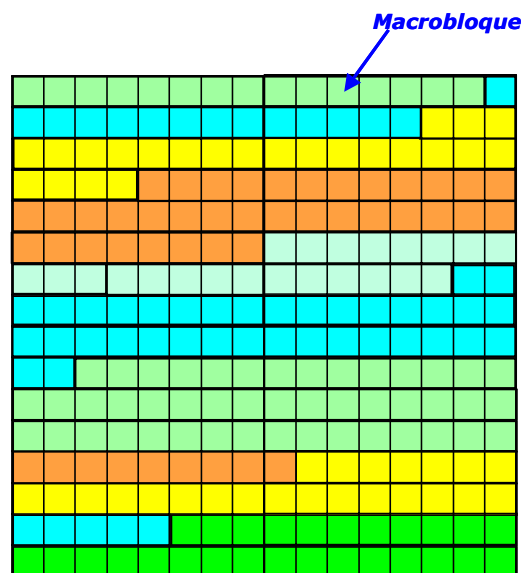


Fig. 6.3 Cuadro de imagen segmentado.

La unidad primaria de codificación de una secuencia de vídeo es la *imagen* o *cuadro*, formada por un cierto número de segmentos que constituyen el *área activa*. Una secuencia de vídeo, se

¹⁰ La traducción literal de este término equivaldría a “tajada” o “rebanada”. Aquí se ha preferido utilizar la palabra *segmento* que, para el caso que nos ocupa, tiene el mismo significado.

compone de uno o más *grupos de imágenes* (GOP¹¹) que, a su vez, están constituidos por una o más imágenes. Esta jerarquía de la información de vídeo en MPEG se ilustra en la figura 6.4.

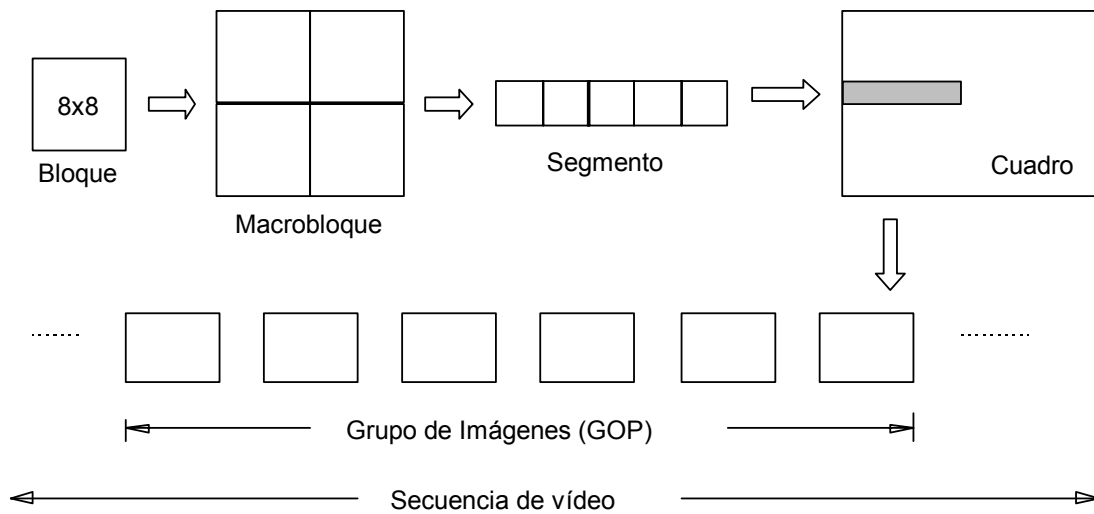


Fig. 6.4. Jerarquía de los datos de vídeo.

6.7 Tipos de imágenes en MPEG

Las imágenes en MPEG se clasifican en tres tipos:

1. **Imágenes I** (intraframe-coded). Son imágenes codificadas de manera totalmente independiente, sin referencia a ninguna otra. Algunos elementos del proceso de compresión explotan únicamente la redundancia espacial de una imagen, ya sea un cuadro o un campo. Estos procesos se designan como *codificación intracuadro* y no se hace uso de la correlación entre cuadros cercanos, gracias a la cual es posible la predicción temporal, a la que se designa como *codificación intercuadro*. En las imágenes I solamente se aplica codificación intracuadro. El empleo periódico de cuadros I facilita la reinicialización del receptor y la adquisición de datos del canal cuando se enciende el receptor o cuando se cambia la sintonía a otro canal. Además, el decodificador puede aprovechar el modo de codificación intracuadro cuando ocurren errores irrecuperables en el canal de comunicación. Al hacer predicción de movimiento, debe disponerse de un cuadro inicial en el decodificador para dar principio a la secuencia o lazo de predicción y por otra parte, los cuadros I proporcionan un medio para que el decodificador recupere rápidamente el sincronismo en caso de pérdida de éste. La periodicidad de las imágenes I puede variar y se selecciona en el codificador.
2. **Imágenes P** (predictive-coded). Son imágenes comprimidas como resultado de la codificación de las diferencias entre la imagen predicha y una referencia, que puede ser una imagen I u otra imagen P y por consecuencia, aprovechan la semejanza entre imágenes adyacentes. Las predicciones pueden realizarse basándose en imágenes I o P precedentes en el tiempo o no realizar ninguna predicción, en cuyo caso esa zona de la imagen se codifica con técnica intracuadro. En los cuadros P la predicción sólo se realiza hacia adelante, es decir, las predicciones se forman solamente con la

¹¹ Group of Pictures

información de cuadros I o P decodificados inmediatamente antes. Los cuadros con predicción hacia adelante permiten aprovechar las técnicas de codificación intracuadro para mejorar la eficiencia global de compresión y la calidad de la imagen, además los cuadros P pueden incluir porciones codificadas sólo con técnicas intracuadro. Cada macrobloque de un cuadro P puede, por tanto, haber sido formado mediante predicción hacia adelante o por codificación intracuadro.

3. Imágenes **B** (bidirectionally predictive-coded). Se comprimen, de manera similar a las imágenes P, codificando la diferencia entre la imagen predicha y una imagen de referencia I o P. Las predicciones se obtienen de las imágenes I o P más cercanas en la secuencia, ya sea pasadas o futuras. Estas imágenes de referencia, I o P, se designan a veces como imágenes o cuadros “ancla”. La predicción de las imágenes B se basa en el hecho de que la información de un cuadro está correlacionada, tanto con la de cuadros en el pasado, como con la de los que ocurran en el futuro. Por consecuencia, si es posible disponer de un cuadro futuro en el decodificador, se puede lograr una predicción mejor. Algunas de las consecuencias de emplear cuadros futuros en la predicción son que, por una parte, los cuadros B no se pueden emplear en la predicción de otros cuadros futuros y además, el orden de transmisión de los cuadros no es el mismo que el orden de presentación al observador, por lo que el codificador y el decodificador deben reordenar los cuadros aumentando, por consecuencia, la latencia total. Los cuadros I y P se transmiten fuera de secuencia respecto a los cuadros B a fin de que el decodificador pueda disponer de ambos cuadros “ancla” para efectuar la predicción. Los cuadros B se emplean para aumentar la eficiencia de compresión y la calidad de la imagen percibida por el observador cuando la latencia de codificación no es un factor importante. Con el empleo de estos cuadros se aumenta también la eficiencia de codificación, tanto para señales de barrido entrelazado como progresivo. El número de cuadros B entre dos imágenes I o P de referencia se determina en el codificador.

Las imágenes de un grupo se organizan en forma similar a la ilustrada en la figura 6.5. El número de cuadros P y B entre dos cuadros I es variable y depende del proceso de codificación.

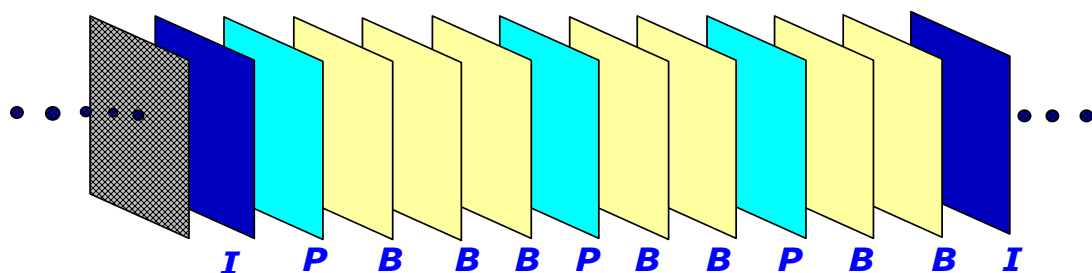


Figura 6.5. Grupo de Imágenes.

Las imágenes P y B se codifican como diferencias, por lo que solamente es necesario actualizar la información correspondiente a las áreas en que ocurran cambios. Este proceso se designa como *llenado o relleno condicional* (conditional replenishment).

Los grupos de imágenes (GOP) pueden tratarse como cerrados o abiertos. En el primer caso, las imágenes P y B sólo se predicen a partir de otras imágenes del mismo grupo. En grupos abiertos la predicción puede realizarse con otras imágenes fuera del grupo.

6.8 Composición del caudal de datos de vídeo

Las técnicas MPEG constituyen una sintaxis que permite seleccionar la frecuencia de ocurrencia de los cuadros I y P, así como el número de imágenes bidireccionales (B) en el codificador, teniendo en cuenta la necesidad de acceso aleatorio en los procesos de edición. Para que el decodificador pueda realizar su función adecuadamente, es necesario disponer de imágenes previas y futuras en la secuencia que le sirvan de referencia para efectuar la predicción. Para ello, es necesario reordenar los cuadros antes de transmitirlos de modo, que al inicio de cada grupo de imágenes haya dos cuadros de referencia. Supóngase que el orden original del grupo de imágenes al ser codificado es:

BIBBPBBPBBBPB

El orden del grupo para transmisión debe empezar siempre con dos cuadros a partir de los cuales pueda efectuarse la predicción bidireccional; así, un orden posible sería:

IPBBBPBBPBBBPB

6.9 Arquitectura del sistema

De acuerdo a lo anterior y, con una perspectiva del sistema, un caudal binario MPEG puede considerarse constituido básicamente por dos capas: una del sistema y otra de compresión en que la primera proporciona la envoltura necesaria a las capas de compresión. La estructura del sistema puede resumirse fácilmente para el caso del decodificador en la forma mostrada en la figura 6.6.

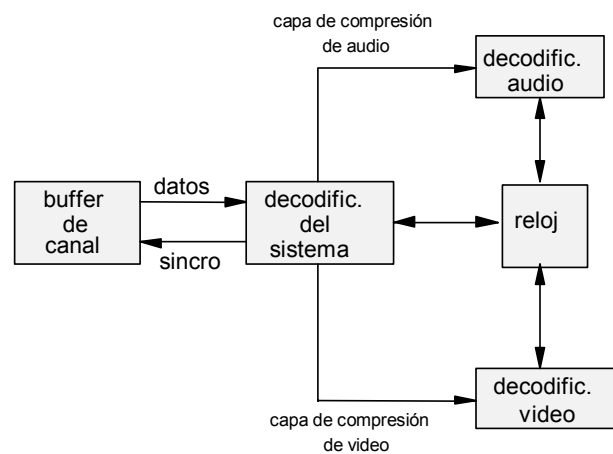


Fig. 6.6. Estructura del proceso de decodificación en el sistema MPEG

Las capas de compresión contienen los datos que deben alimentarse a los decodificadores de audio y vídeo en tanto que la capa del sistema proporciona los medios de control para el demultiplexado de las capas de compresión y de sincronismo necesario en la reproducción.

El caudal binario MPEG consiste de una secuencia de envolturas¹² que, a su vez, se subdividen en paquetes, en la forma que se muestra en la figura 6.7.

¹² El término empleado en inglés es *pack* que puede traducirse como *empaquetado*. Aquí se ha preferido utilizar el término *envoltura* para evitar confusión con *paquete* (*packet*) que tiene un significado diferente en MPEG.

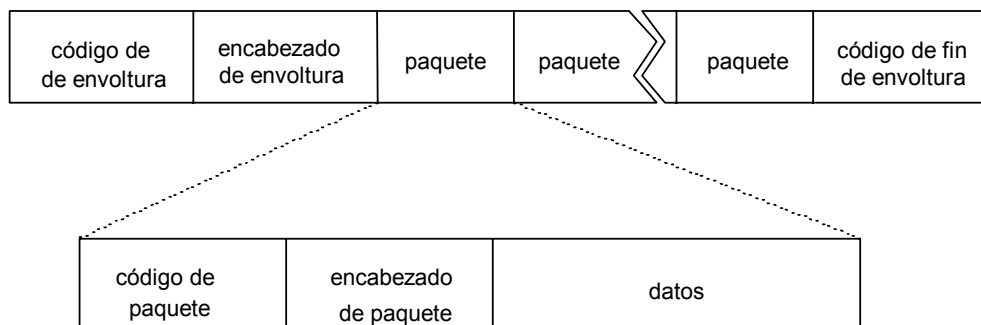


Fig.6.7. Envoltura de la capa del sistema y estructura de los paquetes MPEG.

Cada “envoltura” consiste de una secuencia de 32 bits que contiene el código de inicio y el encabezado, seguido de uno o más paquetes de datos que a su vez, contienen un código de inicio y un encabezado también con 32 bits, seguidos de datos comprimidos de audio o vídeo. El decodificador del sistema identifica el tipo de información sin necesidad de decodificar los datos y canaliza éstos hacia los decodificadores respectivos, junto con la información de sincronismo, extraída del decodificador del sistema.

6.10 Codificación basada en transformadas.

La codificación basada en transformadas o, simplemente, codificación por transformadas, es un método muy extendido empleado en la reducción del caudal binario de una señal y constituye uno de los procesos básicos en los estándares MPEG. El principio de esta técnica consiste en cambiar el dominio de la variable original a otro dominio como en el caso de la transformada de Fourier y por analogía con ésta, al dominio de la variable transformada se le designa como dominio espectral o dominio de frecuencia, aunque las componentes espectrales no tengan en general, el mismo significado habitual de ciclos por unidad de tiempo.

Aunque quizá la transformada más conocida es la de Fourier, hay otras transformadas que se utilizan ampliamente en el terreno de procesamiento de señales, entre las que pueden citarse la de Hilbert, Hadamard, Hartley, seno discreto, coseno discreto, Karhunen-Loeve, etc., y algunas de ellas son particularmente interesantes en el dominio digital. En el caso de una imagen, los niveles de crominancia y luminancia de los elementos de imagen de un cuadro se almacenan como valores numéricos en una matriz bidimensional. Esta matriz no es otra cosa que un conjunto de números que pueden manipularse de formas diversas y, entre ellas, pueden ser transformados a otro dominio mediante la aplicación de alguna transformada como las mencionadas antes. Las respectivas transformadas son, para este caso, bidimensionales y la transformación da como resultado un número de coeficientes igual al número de elementos de la matriz original. Se dice, en estas condiciones, que la matriz transformada representa los valores de *frecuencia espacial* de la señal original. En la figura 6.8 se representa esquemáticamente la información en los dominios espacial y transformado para un bloque de 8x8 elementos de imagen o pixels, que corresponden a una porción de la imagen original. Si se aplica a los valores de ese bloque una transformada, se obtiene una nueva matriz de 8x8 coeficientes. La aplicación de la transformada inversa sobre los valores de este bloque transformado, darían como resultado la matriz en el dominio espacial. El proceso de transformación es, estas condiciones *reversible*, ya que la aplicación de la transformada inversa no produce pérdida de información de la matriz original. Mediante la caracterización transformada, los coeficientes pueden considerarse que indican el grado de correspondencia de un patrón particular de luminancia, en el dominio de frecuencia, con un campo determinado de la imagen en el dominio espacial. La *imagen transformada* puede someterse posteriormente a diversos procesos que serán tratados en las secciones subsiguientes.

Una característica importante de las diversas transformadas es su propiedad de *compactación*. Los coeficientes de una transformada representan también la energía de una señal a las diversas frecuencias; sin embargo, no todos los coeficientes son igualmente significativos y algunas transformadas dan lugar a coeficientes muy poco significativos en altas frecuencias o, por decirlo de otra forma, compactan la mayor parte de la energía de la señal sólo en unos pocos coeficientes. En el caso de la imagen transformada bidimensional en el bloque de la figura 6.8, la zona de bajas frecuencias corresponde a la parte superior izquierda de la matriz, en tanto que las frecuencias más altas corresponden al extremo inferior derecho. El término de la matriz en la celda superior izquierda se designa como término de corriente continua, por similitud con el caso eléctrico.

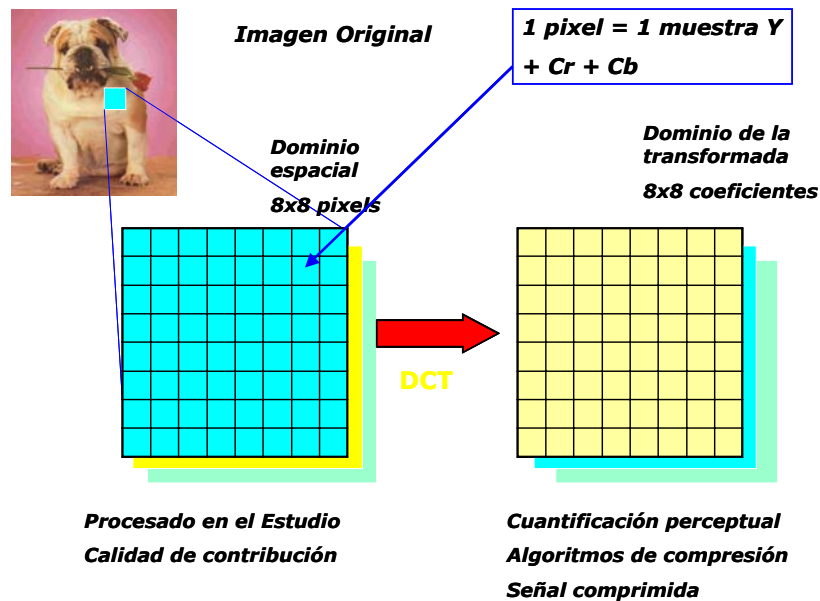


Fig. 6.8. Información en los dominios espacial y transformado.

En la figura 6.9 se pueden apreciar las propiedades de compactación de cuatro transformadas: Fourier, coseno discreto, Hadamard y Karhunen-Loeve. En ellas las zonas blancas corresponden a los coeficientes significativos y las zonas más oscuras a coeficientes de valor muy pequeño o nulo. Esta es una propiedad muy importante que, como se verá más adelante, permite prácticamente despreciar estos coeficientes sin pérdida importante de información desde el punto de vista perceptual.

Desde el punto de vista de cómputo, la transformada de Fourier presenta el inconveniente de que sus componentes espectrales son complejos, lo que aumenta el número de operaciones necesarias para su procesamiento y por otra parte, sus propiedades de compactación en el sentido de concentrar la energía en pocos componentes espectrales no es muy buena. De las transformadas anteriores, la que mejores propiedades de compactación ofrece es la de Karhunen-Loeve, sin embargo, requiere de una carga de cómputo considerable, lo que la hace poco adecuada para su aplicación en televisión en que las imágenes deben manejarse a gran velocidad en tiempo real. Por otra parte, la transformada del coseno discreto (DCT), aunque algo menos eficiente que la de Karhunen-Loeve, plantea menos dificultades de cómputo y es en la que se basan los estándares JPEG y MPEG.

Como no todas las imágenes tienen las mismas características estadísticas¹³, la transformada óptima no es constante y depende del contenido instantáneo de información en la imagen a codificar, con lo que se hace necesario recalcular la matriz transformada óptima para cada bloque a transmitir. Este es un principio que se aplica por ejemplo, en la transformada de Karhunen-Loeve (KLT). Esta

¹³ Whitaker, J. "Transform Coding". Broadcast Engineering, Vol. 39, N° 2, pp.14-16. Feb. 1997.

transformada es eficiente en términos de rendimiento, pero no es adecuada en la práctica para manejar imágenes en tiempo real o cuasi real, ya que obliga a investigar las características de cada imagen y calcular la matriz transformada óptima. Además, esta matriz debe proporcionarse al decodificador en cada cuadro, ya que es necesaria en el proceso de decodificación para recuperar la información original.

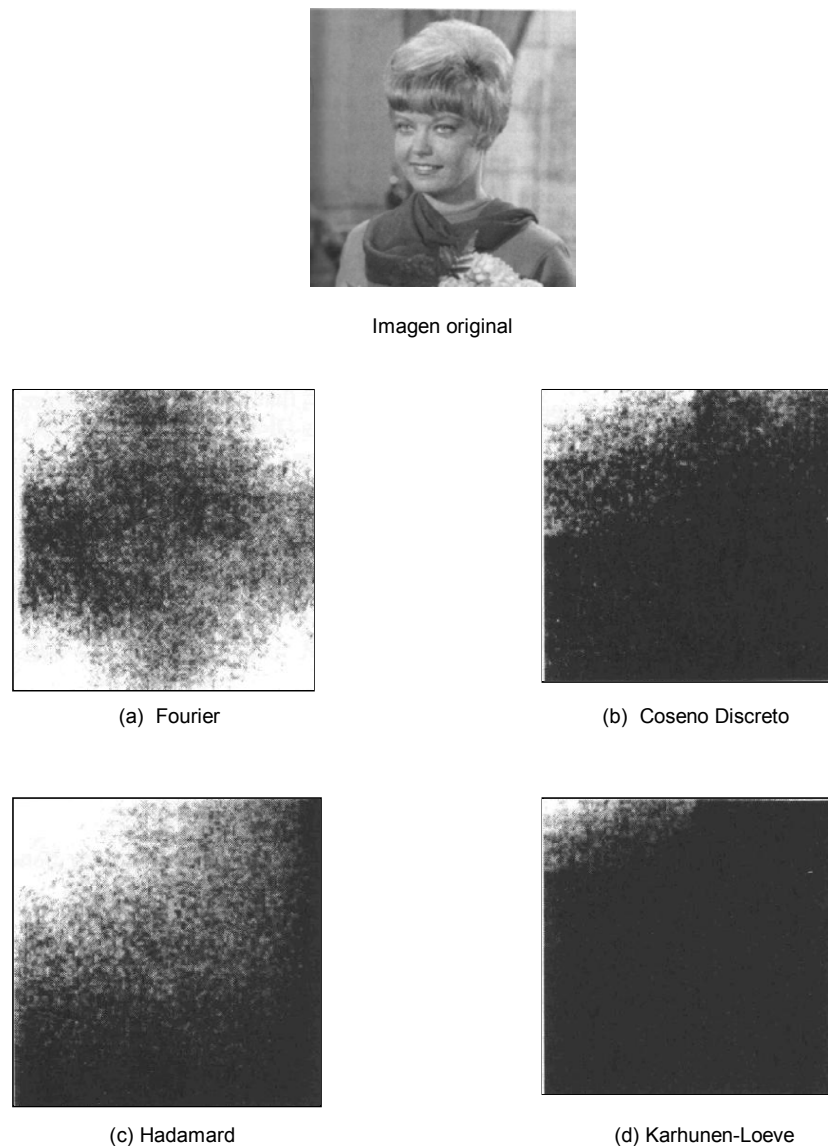


Fig. 6.9. propiedades de compactación de algunas transformadas

La transformada del coseno discreto por otra parte, emplea una matriz constante y es adecuada para una gran variedad de imágenes. A veces se le designa como “transformada rápida de Karhunen-Loeve” o KLT rápida. Tiene semejanza con la transformada de Fourier pero a diferencia de ésta, los coeficientes transformados son reales. Su principal ventaja es que decorrelaciona eficientemente los elementos de imagen, es decir, convierte de modo eficiente los valores de los elementos de imagen estadísticamente dependientes en coeficientes independientes. Como consecuencia de esta

propiedad, la DCT “empaqueta” la energía de la señal contenida en un bloque, en un número reducido de coeficientes y permite una gran variedad de implementaciones rápidas.

La transformada que se aplica en el caso de bloques de elementos de imagen no es unidimensional, ya que se tienen dos variables espaciales, una en dirección horizontal y otra vertical. Se dice entonces que la transformada es bidimensional o plana y una de sus propiedades es la separabilidad, lo que significa que se puede aplicar primero a una de las variables y luego a la otra.

Para fines de procesamiento de la señal en el estudio o centro de producción así como para transmisión con fines de contribución, la señal debe conservar la máxima calidad y por tanto, el nivel de compresión al que se somete es muy pequeño o nulo, por lo que en estas aplicaciones a veces no se transforma el dominio de la información original. Cuando la transmisión se destina al público en general, lo que interesa es la calidad subjetiva de la imagen y entonces es permisible aplicar un nivel de compresión relativamente elevado. Una expresión usual de la transformada bidimensional del coseno discreto es:

$$F(u, v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \left[\frac{(2x+1)u\pi}{16} \right] \cos \left[\frac{(2y+1)v\pi}{16} \right] \quad (6.1)$$

donde x e y son los índices de los elementos de imagen en un bloque de 8×8 , u y v los índices de los coeficientes de la DCT para ese bloque y:

$$\begin{aligned} C(w) &= \frac{1}{\sqrt{2}} \quad \text{para } w = 0 \\ &= 1 \quad \text{para } w = 1, 2, \dots, 7 \end{aligned} \quad (6.2)$$

De acuerdo a lo anterior, un conjunto de 8×8 números $f(x, y)$ son los datos de una fórmula matemática, cuyo resultado es otro conjunto de 8×8 números diferentes, $F(u, v)$. La transformada inversa¹⁴ (IDCT) está dada por:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos \left[\frac{(2x+1)u\pi}{16} \right] \cos \left[\frac{(2y+1)v\pi}{16} \right] \quad (6.3)$$

En el decodificador se aplica la transformada inversa para reconstruir, aproximadamente, los conjuntos de números transformados en el codificador. La aproximación en la reconstrucción se controla de antemano durante la codificación con el fin de minimizar los efectos visuales que resultan de las imprecisiones en los coeficientes, al mismo tiempo que se reduce la cantidad de datos que deben transmitirse. En la figura 6.10 se muestra un diagrama de bloques de un conjunto codificador-decodificador basado en la DCT.

En la codificación intracadro, la DCT se aplica a todos los datos “en bruto” de la imagen, sin tener en cuenta la información de otros cuadros y transformando la luminancia y crominancia por separado. Puesto que la información de crominancia está submuestreada horizontal y verticalmente, cada bloque de 8×8 datos de crominancia (C_r o C_b), corresponde a un macrobloque de 16×16 datos de luminancia.

¹⁴ Es necesario que la transformada inversa (IDCT) cumpla el estándar de especificaciones del IEEE para la implementación de la transformada inversa del coseno discreto de 8×8 . *Std. 1180-1990*, December 6, 1990.

En la codificación intercuadro, las diferencias o errores de predicción entre dos cuadros diferentes se agrupan en bloques de 8x8 y se aplica una transformación espacial mediante la DCT a estos bloques de valores de diferencia.

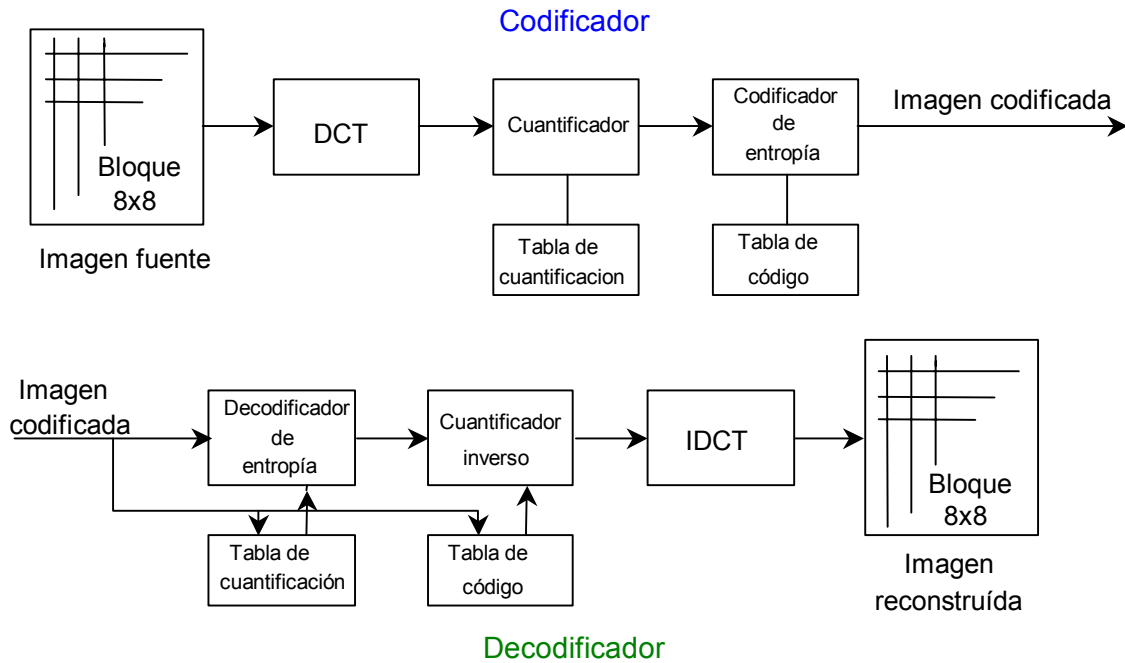


Fig. 6.10. Diagrama de bloques de un conjunto codificador-decodificador intracuadro basado en la DCT

6.11 Lazo de predicción en el codificador

La codificación intercuadro es más compleja que la intracuadro ya que se trata de codificación predictiva que requiere de la información de cuadros previos y, para las imágenes B en MPEG, también de cuadros futuros. Una posible forma de comprender la forma en que se combinan los diferentes elementos algorítmicos para lograr la compresión de vídeo es examinando el lazo de predicción que constituye la parte medular del sistema de compresión y que se muestra esquemáticamente en el diagrama de bloques de la figura 6.11.

El lazo de predicción realiza la función de estimar, o predecir, los valores de la imagen del siguiente cuadro que se codificará en la secuencia de imágenes sucesivas que constituyen un programa de televisión. Esta predicción se basa en información previa disponible en el lazo y obtenida de imágenes anteriores. La transmisión de la información comprimida predicha funciona debido a que la misma información empleada para la predicción está disponible también en el decodificador del receptor, asumiendo que la transmisión es sin errores.

La resta de los valores predichos de los de la nueva imagen a codificar constituye el elemento medular de la codificación predictiva. La meta es realizar la predicción de manera tal que la predicción de los nuevos valores que resulta de la función de resta al inicio del lazo sea cero o muy cercana a cero la mayor parte del tiempo. La predicción de las diferencias se realiza separadamente para la luminancia y para las dos componentes de crominancia antes de cualquier otro proceso subsecuente. En la codificación intracuadro no se efectúa ninguna predicción.

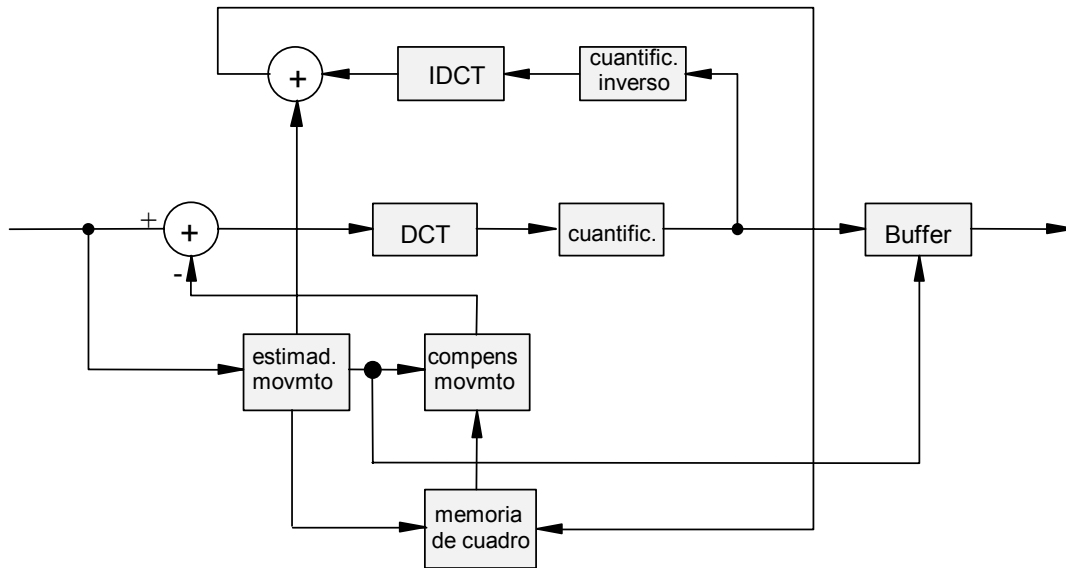


Fig. 6.11. Lazo de predicción del codificador

El hecho de utilizar cuadros pasados o futuros introduce una tercera variable en el proceso de compresión y se puede hablar entonces de una transformación tridimensional o *cúbica*, que impone requerimientos importantes en la capacidad de memoria. Así para una DCT de $8 \times 8 \times 8$ son necesarias, por lo menos, siete memorias de cuadro y se recurre a una alternativa más simple designada como DCT *híbrida* que también permite la codificación de imágenes con objetos en movimiento de forma eficiente. Esta técnica emplea modulación diferencial de pulsos codificados (DPCM) en que, en lugar de codificar cada imagen individualmente, se codifica la diferencia entre dos cuadros sucesivos, compensada en movimiento. La DPCM es, esencialmente una forma de codificación predictiva de las diferencias entre muestras y puede aplicarse en codificación intercuadro aprovechando la redundancia temporal entre la imagen de entrada al codificador y la imagen almacenada. También, puede aplicarse a la codificación intracuadro aprovechando la redundancia espacial de la imagen.

En el modo intracuadro, las diferencias se calculan utilizando los valores de dos elementos de imagen adyacentes en el mismo cuadro. En el modo intercuadro la diferencia se calcula utilizando la diferencia entre los valores del mismo elemento de imagen en cuadros sucesivos. En ambos modos de operación el valor del elemento de imagen resultante se predice utilizando los valores reconstruidos de los elementos de imagen vecinos, previamente codificados y este valor se resta luego del valor original para formar el valor del elemento correspondiente en la imagen diferencial.

6.11.1 Cuantificador

La información de un bloque de elementos de imagen se transforma mediante la DCT en un nuevo conjunto de coeficientes espectrales del bloque o de las diferencias entre bloques. Una vez realizada esta transformación el codificador puede ajustar selectivamente la precisión de los coeficientes transformados, representando los de los componentes espectrales de alta frecuencia espacial con menor precisión que los de baja frecuencia, ya que la resolución de la visión humana es inferior para las altas frecuencias espaciales. Este proceso recibe también el nombre de cuantificación y no debe confundirse con la cuantificación que se realiza en un conversor analógico-digital. Los coeficientes de la DCT se cuantifican dividiéndolos entre un número entero positivo designado como *valor de cuantificación* y redondeando el cociente al entero más cercano. Esta operación se ilustra esquemáticamente en la figura 6.12. La división se realiza no en forma matricial, sino dividiendo cada elemento de la matriz transformada por el elemento correspondiente de la matriz de cuantificación.

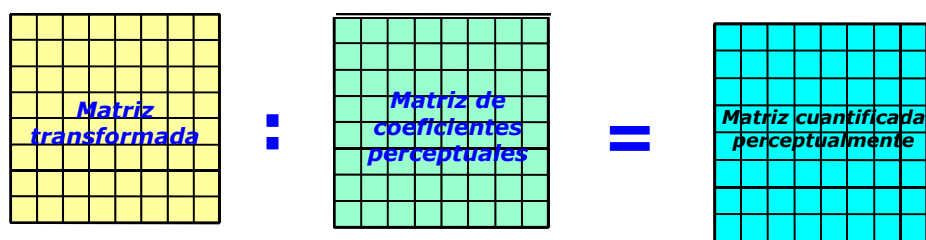


Fig. 6.12. Cuantificación

Cuanto mayores sean los valores de los elementos de la matriz de cuantificación o coeficientes perceptuales, menor será la precisión del coeficiente DCT cuantificado. Esta cuantificación, según se mencionó, está ponderada en términos de la respuesta visual humana y pueden ser, por ejemplo para codificación intracadro, los mostrados a continuación, únicamente a fin dar una idea, no son únicos posibles dependiendo del codificador.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Los coeficientes aplicados a los bloques de crominancia suelen ser mayores, lo que da como resultado menor resolución, de nuevo como consecuencia de las características de la visión humana en la percepción cromática. La tabla aplicada en el caso de codificación intercuadro es plana, con un valor fijo de 16 para todos los coeficientes. La razón de ello es que, en el caso intercuadro, lo que se codifica son cambios diferenciales entre dos imágenes.

El nivel de degradación subjetiva de la imagen causado por la cuantificación de los coeficientes tiende a depender de la naturaleza de la imagen codificada, ya que para una imagen dada las distorsiones en algunas zonas pueden ser menos notorias que en otras. El sistema de codificación de vídeo permite ajustar el nivel de cuantificación para cada macrobloque a fin de reducir en lo posible el número de bits, empleando cuantificación más basta. Así, los coeficientes perceptualmente importantes, los de baja frecuencia, pueden cuantificarse con un grado más fino que los de alta frecuencia. La forma más simple de hacerlo es mediante una ponderación de los coeficientes basándose en su importancia visual, que se refleja en los valores de la matriz anterior.

6.11.2 Cuantificación adaptativa

La sintaxis de codificación de vídeo permite especificar las matrices de cuantificación para cada imagen con el fin de lograr mayor eficiencia en la codificación, por lo que a los códigos de longitud variable (VLC) de los coeficientes cuantificados se les asocia una cierta función de distribución de probabilidad. Aunque no es posible cambiar la distribución del código de longitud variable para ajustarse a la distribución real de los datos, las matrices del cuantificador pueden modificarse para

aproximar el ajuste de la distribución de los datos a la distribución del VLC. Durante la codificación de los datos de un cuadro puede calcularse la varianza en cada banda de frecuencia espacial tanto para datos intracuadro como intercuadro, lo que proporciona un método para elegir los valores de las matrices del cuantificador aplicando límites superiores e inferiores por banda, a fin de asegurar un funcionamiento razonable en todos los casos.

La transmisión de las matrices de cuantificación consume bits del caudal binario comprimido que, si se transmiten con cada imagen pueden representar del orden del 0.3% del ancho de banda del canal. Aunque esta cifra es modesta, puede reducirse actualizando la matriz del cuantificador con menos frecuencia, o bien, cuando la diferencia entre la matriz deseada en el cuantificador y la que está siendo más recientemente utilizada se vuelve significativa.

6.12 Codificación por entropía

Un efecto importante de la cuantificación de los coeficientes transformados, es que muchos de ellos serán redondeados a cero después de la cuantificación. De hecho, un método inicial de controlar el caudal de datos codificados es controlar el nivel de resolución (pasos de cuantificación), ya que una cuantificación con menor resolución aumenta el número de coeficientes con valor cero. A los valores cuantificados o a grupos de ellos, se les asignan palabras de código para generar el caudal binario de salida.

Hipotéticamente, los valores cuantificados podrían representarse simplemente utilizando palabras de código con longitud fija o uniforme; con este enfoque, cada valor cuantificado estaría representado con el mismo número de bits. Sin embargo, puede obtenerse mayor eficiencia en términos de caudal binario utilizando codificación por entropía, mediante la cual se aprovechan las propiedades estadísticas de la señal a codificar. Una señal, ya sea el valor que representa a la luminancia de un elemento o bien un coeficiente transformado, tiene una cierta cantidad de información o entropía, de acuerdo a la probabilidad de que ocurran los diferentes valores o eventos posibles. Por ejemplo, un evento que ocurre con poca frecuencia contiene mucha más información que otro que ocurre frecuentemente. Si se tiene en cuenta este hecho el caudal binario promedio puede reducirse de manera apreciable.

Uno de los esquemas más comunes de codificación por entropía es el *código de Huffman*, en el que es posible generar una tabla de código que se aproxime a la longitud promedio de descripción de los eventos cuando se conoce la distribución de probabilidad de éstos. A los eventos con mayor probabilidad de ocurrencia se les asignan palabras de código más cortas, en tanto que los eventos menos probables se expresan con palabras de mayor longitud.

Según se mencionó, en la codificación de vídeo la mayoría de los coeficientes cuantificados en la región de alta frecuencia espacial tendrán valor cero, quizá con unos pocos coeficientes, por lo general dispersos, cuyo valor sea diferente de cero. Para aprovechar esta propiedad, se modifica la matriz de los coeficientes transformados en forma de una secuencia o vector unidimensional, mediante un proceso de barrido de los coeficientes de la matriz que puede ser en zigzag, en la forma indicada en la figura 6.13. El tipo de barrido puede ser también de otras formas alternadas, pero debe especificarse antes de codificar cada imagen y puede variar de cuadro a cuadro.

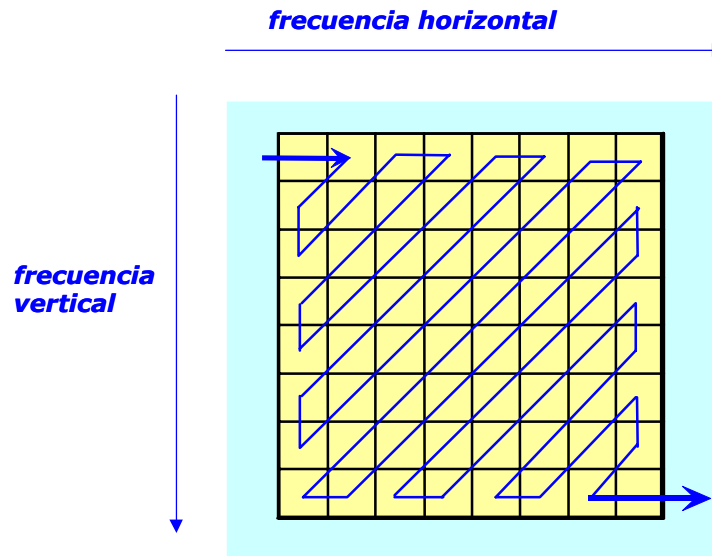


Fig 6.13. Recorridos en zigzag de los coeficientes

El barrido de los coeficientes en esta forma da lugar a que los coeficientes más importantes, diferentes de cero, queden agrupados al inicio de la secuencia y a continuación de ellos, habrá cadenas más o menos largas de ceros que pueden representarse con mayor eficiencia mediante *codificación de recorrido* (runlength coding). Con esta forma de codificación el número de ceros consecutivos que preceden a un coeficiente diferente de cero puede codificarse por entropía, ya sea de forma separada o conjuntamente con el valor distinto a cero. El recorrido de la matriz en esta forma separa la mayoría de los coeficientes cero de los distintos de cero en grupos, mejorando la eficiencia de la codificación de recorrido. Se emplea un carácter especial de *fin de bloque* (EOB) para indicar el caso en que todos los coeficientes restantes del vector son cero, con lo que se aumenta la eficiencia y se consigue un grado importante de compresión.

6.13 Buffer de canal

Al codificar por entropía, el caudal de salida del codificador es variable y función de la estadística de la señal de vídeo. Puesto que el caudal binario permitido a la entrada del canal de transmisión es menor que el valor de pico del caudal generado por el codificador de longitud variable, es necesario un *buffer de canal* en el codificador en el que se requiere, además, asegurar que no ocurra contención (*underflow*) ni desbordamiento (*overflow*).

El control del buffer obliga a un mecanismo de realimentación en el algoritmo de compresión mediante el cual es posible variar tanto la resolución de la cuantificación, como las resoluciones espacial, temporal y de color, de acuerdo con las necesidades instantáneas del caudal binario. Si éste disminuye considerablemente, puede efectuarse una cuantificación más fina para aumentarlo. Si aumenta, se reduce la resolución de la cuantificación.

El valor de pico del caudal binario a la salida del codificador de longitud variable fluctuará considerablemente y con frecuencia, excederá el ancho de banda máximo asignado al canal de transmisión, aunque el valor medio del caudal no puede, por definición, exceder el valor pico del ancho de banda. Esta situación se maneja por el buffer de canal en cada codificador, que constituye el almacén temporal del caudal binario codificado. El estándar de televisión digital americano especifica una capacidad de 8 Mbits para el buffer de canal.

En el sistema de codificación se define un “buffer modelo” como referencia para los fabricantes de codificadores y decodificadores, a fin de garantizar la interoperabilidad entre equipos de diferente manufactura. El codificador debe controlar su producción de bits de modo tal que el buffer modelo no se vea contenido ni desbordado.

Para evitar la contención o desbordamiento del buffer de canal, el codificador debe ser capaz de tomar medidas adecuadas en función de la ocupación del buffer y de la complejidad de la imagen. Cuando el codificador necesita reducir el caudal binario, puede hacerlo aumentando el nivel de cuantificación, lo que producirá degradación en la imagen. Por el contrario, cuando es posible aumentar el caudal binario, el nivel de cuantificación se reduce y la degradación de la imagen es menor. En realidad, la contención del buffer es permisible en el caso de caudales de pequeño retardo que no tengan cuadros B. En esta situación, la contención puede ocurrir con imágenes muy complejas que requieren un número particularmente grande de bits para su codificación. El resultado de la contención es la repetición de una o más imágenes, que se presentan en lugar de las imágenes omitidas o saltadas en el codificador.

6.14 Concatenación de secuencias

Una secuencia codificada de vídeo comienza con un *encabezado* o *header* y puede contener alguna secuencia repetida de encabezados y una o más imágenes codificadas. La secuencia se termina mediante un *código de fin de secuencia*. En el encabezado de la secuencia se especifican varios parámetros que deben mantenerse constantes mientras dura la secuencia. Estos *parámetros de nivel de secuencia* incluyen, entre otros, los siguientes:

- Resolución horizontal y vertical.
- Frecuencia de cuadro
- Relación de aspecto
- Formato de croma
- Nivel y perfil
- Indicador de barrido (entrelazado o progresivo)
- Verificador de tamaño del buffer
- Máximo caudal binario

Se contempla el hecho de que con frecuencia, será necesario cortar caudales binarios con fines de edición, inserción de anuncios y otros fines habituales, tanto en la producción de programas como en la cadena de distribución. Si uno o más parámetros de nivel son diferentes para los dos caudales binarios (el que se corta y el que se pretende insertar), es necesario insertar un código de fin de secuencia para cortar el primer caudal y debe existir un nuevo encabezado de secuencia al principio del segundo caudal. Esta situación se conoce como *concatenación de secuencias*.

En tanto que el estándar MPEG-2 especifica el comportamiento de los decodificadores de vídeo en el caso de procesar una sola secuencia, no impone ningún requisito para el manejo de secuencias concatenadas. La especificación de la decodificación en el primer caso es factible porque el estándar MPEG-2 impone restricciones a la formación y codificación de secuencias individuales que prohíben el desbordamiento del buffer de canal y la codificación de la misma paridad de campo para dos campos consecutivos. El estándar MPEG no prohíbe estas situaciones en la unión entre dos secuencias codificadas y por tanto, no especifica el comportamiento del decodificador en este caso.

En el estándar de televisión digital americano se recomienda, pero no se impone, la producción de secuencias concatenadas “bien constreñidas”. Las secuencias concatenadas “bien constreñidas” se definen como las que tienen las siguientes características:

- a) El buffer extendido del decodificador nunca se desborda y sólo puede contenerse en el caso de caudales binarios de pequeño retardo. El término “buffer extendido” se refiere aquí a la extensión natural del buffer modelo del decodificador MPEG-2, en el caso de decodificación continua de secuencias concatenadas.
- b) Cuando se especifica la paridad de campo en dos secuencias codificadas concatenadas, la paridad del primer campo en la segunda secuencia es opuesta a la del último campo de la primera secuencia.
- c) Dondequiera que se inserte una secuencia de una imagen de barrido progresivo entre dos de barrido entrelazado, el número exacto de cuadros progresivos debe ser tal que se preserve la paridad de las secuencias entrelazadas, como si no hubiera ocurrido concatenación.

6.15 Compensación de movimiento

La compensación de movimiento es un técnica aplicada en la compresión de las imágenes P y B, que mejora el factor de compresión por un factor de aproximadamente 3 respecto a la codificación intracuadro¹⁵ eliminando la redundancia temporal de la imagen. Si hay movimiento en la secuencia de vídeo se consigue mejor predicción codificando las diferencias relativas a las áreas de la imagen que se desplazan respecto a la imagen codificada. A este proceso se le designa como *compensación de movimiento* y a la determinación de los vectores de movimiento en el codificador, se le llama *estimación de movimiento*.

Los algoritmos de compensación de movimiento se aplican a nivel de macrobloque y, cuando se aplica a cuadros P produce dos tipos de información:

- a) *Vectores de movimiento*, que representan la diferencia entre la información del cuadro de referencia y el macrobloque codificado. Los vectores de movimiento describen la magnitud y dirección del movimiento en un macrobloque y se transmiten al decodificador como parte del caudal binario total, de forma que el decodificador “sabe” qué área de la imagen de referencia fue utilizada para cada predicción. Esto se ilustra en la figura 6.14.
- b) *Términos de error*, que representan las diferencias entre los valores predichos y los resultados reales.

¹⁵ Baron, S. and Wilson, W.R. “MPEG Overview” ITU/SMPTE Tutorial Digital Terrestrial Television Broadcasting”. Society of Motion Picture and Television Engineers (SMPTE). pp. 28-35. White Plains, NY. 1994.

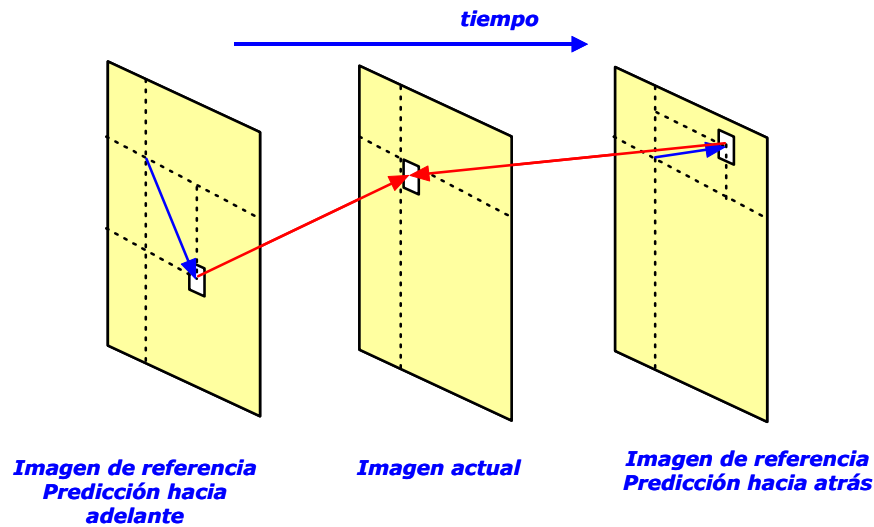


Fig. 6.14 Vectores de movimiento

Cuando no es posible generar una imagen P empleando compensación de movimiento, como en el caso en que un objeto aparezca súbitamente en una escena, el macrobloque se codifica de la misma forma que en los cuadros I, es decir, con técnicas intracuadro.

La codificación de la compensación de movimiento en los cuadros B puede hacerse utilizando imágenes de referencia tanto previas como futuras y puede dar lugar a cuatro posibles tipos de codificación:

- a) *Intracuadro*, en que no se tiene compensación de movimiento.
- b) *Predicción hacia adelante*, en que el cuadro previo I o P más cercano sirve de referencia.
- c) *Predicción hacia atrás*, en que el cuadro futuro I o P más cercano sirve de referencia.
- d) *Predicción bidireccional*, en la que se usan dos imágenes de referencia, la previa I o P más próxima y la futura I o P también más próxima.

La predicción hacia atrás puede emplearse para definir áreas cubiertas que no aparezcan en imágenes previas.

6.16 Decodificador de vídeo

El decodificador de vídeo, cuyo diagrama general de bloques se ilustra en la figura 6.15 y que no debe confundirse con el decodificador de canal, contiene los elementos necesarios para invertir el proceso de codificación. El caudal binario se almacena inicialmente en un buffer de canal cuya salida alimenta a un decodificador de longitud variable (VLD).

El decodificador de longitud variable reconstruye los bloques de 8×8 coeficientes cuantificados durante la codificación de recorrido/amplitud y los distribuye adecuadamente, de acuerdo al tipo de barrido utilizado en la codificación. A la salida del VLD se tienen los coeficientes de predicción de error cuantificados mediante la DCT. El cuantificador inverso convierte los coeficientes cuantificados a su valor original y luego les aplica la transformada inversa del coseno discreto (IDCT) para obtener los valores de luminancia de los elementos de imagen originales o bien los errores de predicción.

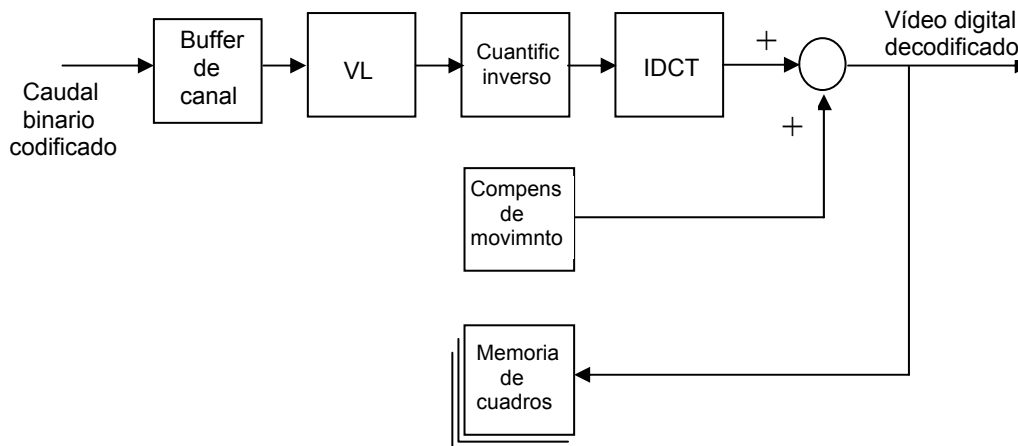


Fig. 6.15. Decodificador de vídeo

En el caso de predicción intercuadro el decodificador utiliza los vectores de movimiento recibidos para llevar a cabo la misma predicción que realizó el codificador. Los errores de predicción se suman con los resultados de la predicción de compensación de movimiento para producir, finalmente, los valores de luminancia de los elementos de imagen, es decir, el vídeo en componentes, decodificado.

Aunque no se muestra explícitamente en el diagrama de bloques de la figura 6.10, cuando ocurren errores de transmisión que no hayan podido recuperarse en el decodificador de canal, el decodificador debe ser capaz de enmascararlos para minimizar su efecto. Una forma relativamente simple de hacerlo es interpolar el valor del elemento erróneo, asignándole un valor intermedio entre los de los elementos de imagen previos y posteriores.

6.17 Jerarquía de codificación en MPEG-2

El estándar MPEG-2 puede considerarse como una “caja de herramientas”, que permite al usuario seleccionar diferentes tipos de codificación, perfiles y diversos niveles de resolución, muestreo y complejidad de procesamiento. El estándar es escalable hasta el nivel de alta definición e incluye además, provisiones para la compresión de audio multicanal. MPEG-2 fue desarrollado como un estándar cuya capa del sistema es independiente del medio (cable, satélite, etc.) en lo que respecta a su multiplexado y transporte. Se dice que un decodificador MPEG-2 está en conformidad con un nivel y perfil dados, si puede decodificar adecuadamente todos los elementos sintácticos del caudal binario según están especificados para ese perfil y nivel. Adicionalmente, un decodificador debe ser compatible con todos los caudales binarios de orden inferior en la matriz. En la tabla 6.2 se muestran los perfiles y niveles de codificación de MPEG-2; los valores que aparecen en blanco no han sido definidos.

Perfiles → Niveles ↓	Simple	Principal	SNR	Escalable Espacial I,P,B 4:2:0	Alto I,P,B 4:2:0/4:2:2
Alto <1920×1152 60 cuadros/s		<80 Mb/s			<100 Mb/s
Alto - 1440 <1440×1152 60 cuadros/s		<60 Mb/s		<60 Mb/s	<80 Mb/s
Principal <720×576 30 cuadros/s	<15 Mb/s	<15 Mb/s	<15 Mb/s		<20 Mb/s
Bajo <352×288 30 cuadros/s		<4 Mb/s	<4 Mb/s		

Tabla 6.2 Jerarquías de codificación en MPEG-2.

Puede decirse que hay dos tendencias en la codificación de MPEG-2. Una se inclina por el perfil simple al nivel principal (SP@ML) y otra por el perfil principal al nivel principal (MP@ML). La primera defiende el argumento de que, sin cuadros B, el costo de los decodificadores puede reducirse considerablemente, ya que la decodificación de los cuadros B requiere de una considerable cantidad de memoria porque deben almacenarse en memoria cuadros tanto previos como futuros para efectuar la predicción. La segunda tendencia, que es la mayoritaria, postula que cuando se hace referencia a MPEG-2, de hecho se está haciendo referencia al perfil y nivel principales. En la práctica el comportamiento de los codificadores SP y MP puede considerarse excelente y para un caudal binario dado, es difícil notar diferencias en la calidad subjetiva de las imágenes decodificadas. Sin embargo, el problema es de compatibilidad al plantearse la cuestión de si un decodificador para SP@ML es capaz de decodificar un caudal MP@ML.

6.18 Resumen de los modelos de codificación MPEG

La descripción detallada de los modelos (algoritmos) de codificación MPEG queda fuera del alcance de este texto. Un tratamiento bastante detallado del tema se tiene en la referencia [1], en la que se ha basado una buena parte del material anterior. Nos limitaremos aquí a ofrecer un breve resumen de las características más importantes.

- Los coeficientes de la DCT se codifican sin pérdidas, de modo que el decodificador los puede reconstruir con precisión.
- Se utiliza una codificación aproximadamente óptima, basada en el código de Huffman para generar las tablas de los códigos de longitud variable necesarios para ello.
- Los coeficientes se ordenan de acuerdo a una secuencia unidimensional en zigzag. Cuanto mayor sea el número de coeficientes de valor cero, mayor será el nivel de compresión.
- El modelo o algoritmo de codificación utiliza símbolos que permiten la codificación eficiente de la DCT cuando la mayoría de los coeficientes son cero. Un símbolo muy utilizado es el de fin de bloque (EOB).
- Se aplican tablas distintas de cuantificación para los casos intracadro e intercadro y, en esta última, el redondeo se realiza de forma diferente. Las

características de los datos son diferentes en ambos casos, por lo que se emplean tablas de código distintas y diferentes conjuntos de símbolos.

La codificación en MPEG es altamente asimétrica es decir, irreversible, ya que los datos de entrada al codificador no pueden recuperarse a partir de los datos decodificados. Los decodificadores son relativamente simples, en tanto que los codificadores son muy complejos y deben identificar áreas de movimiento, determinar los vectores óptimos de movimiento, controlar tanto el caudal binario como el almacenamiento temporal (buffering) para determinar cuando modificar los niveles de cuantificación, determinar cuando es posible repetir un bloque y, finalmente, determinar cuando aplicar técnicas de codificación intracadro o intercadro.

6.19 Formato SIF (Source Input Format)

La sintaxis del flujo binario en MPEG-1 permite tamaños de imagen hasta de 4095×4095 pixels. Sin embargo, muchas de las aplicaciones en que se emplea vídeo comprimido MPEG-1 están optimizadas para el formato de entrada de fuente o SIF, derivado del formato especificado en la Recomendación 601 del CCIR (4:2:2). De acuerdo a la Rec. 601, una fuente de vídeo cromático tiene tres componentes, una de luminancia (Y) y dos de crominancia (C_b y C_r). Para el formato CCIR 601 hay dos opciones para la resolución espacial de la imagen. La primera, para los sistemas NTSC, utiliza 525 líneas por cuadro y 60 cuadros por segundo. El cuadro de luminancia tiene 720×480 pixels activos y, cada cuadro de crominancia tiene 360×480 pixels activos. (*Nótese que se trata de MPEG-1, para barrido secuencial*). La segunda opción, para los sistemas PAL, utiliza 625 líneas por cuadro a 50 cuadros por segundo. Aquí la luminancia tiene 720×576 pixels activos y las componentes de crominancia, 360×576 pixels activos. Para conseguir flujos binarios tan bajos como 1.5 Mb/s, MPEG-1 define (pero no requiere), el formato de entrada de fuente, SIF. Las secuencias SIF tienen una resolución de 360×240 pixels por cuadro a 30 cuadros por segundo, o 360×288 pixels por cuadro a 25 cuadros por segundo. En ambos casos la resolución de las componentes de croma es la mitad de la de luminancia, tanto en las direcciones horizontal como vertical. Este formato se designa como 4:2:0. Las imágenes SIF pueden obtenerse fácilmente a partir de las imágenes CCIR 601 mediante filtrado y submuestreo.

Frecuencia de cuadro →	30	25
CCIR 601		
Y	720×480	720×576
Cb, Cr	360×480	360×576
SIF		
Y	360×240	360×288
Cb, Cr	180×120	180×244
Area de pixels significativos para SIF		
Y	352×240	352×288
Cb,Cr	176×120	176×144

Tabla 6.3 Pixel significativos en MPEG-1

En MPEG-1, las componentes de color Y,Cr y Cb están siempre intercaladas, es decir ...CrYCbYCr... y, por otra parte, se define como unidad mínima de codificación el *macrobloque*, que consiste de cuatro bloques de luminancia de 8×8, un bloque de 8×8 de Cb y otro de 8×8 de Cr. La dimensión máxima de un macrobloque es de 16 pixels y cada imagen se divide en una serie de macrobloques de izquierda a derecha y de arriba a abajo, lo que obliga a que tanto la resolución horizontal como vertical sea múltiplo de 16. Si no es así, el codificador añade pixels de relleno a la derecha o en la parte inferior de cada imagen que luego son descartados en el decodificador. Puesto que la resolución horizontal del formato SIF no es divisible por 16, se puede duplicar el último

pixel de cada línea ocho veces, de modo que el número total de pixels por línea es de 368. En una implementación alternativa, se descartan cuatro pixels del extremo izquierdo y otros cuatro del extremo derecho de cada línea, con lo que la imagen se reduce ahora 352 pixels por línea a lo que se designa como *área de pixels significativos*, cuyos valores se reflejan en la tabla 6.3. Como no sería realista esperar que cualquier decodificador pudiera soportar todas las opciones de codificación, MPEG-1 define los *parámetros restringidos o constreñidos del caudal binario*, es decir, el flujo binario que todo decodificador compatible con MPEG-1 debe soportar. Estos parámetros no son un requisito del estándar y se muestran en la tabla 6.4.

Parámetro de codificación	Valor máximo
Tamaño horizontal de imagen	768 pixels
Tamaño vertical de la imagen	576 pixels
Macrobloques	396
Flujo de pixels (pixel rate)	396×25 macrobloques/s
Flujo de imágenes	30 cuadros/s
Rango de los vectores de movimiento	±64 pixels, a resolución de medio pixel
Tamaño del buffer de entrada	327,680 bits.
Tasa binaria	1.856 Mbit/s

Tabla 6.4 Flujo binario de los decodificadores MPEG-1

6.20 Otros estándares de compresión

6.20.1. Estándares de compresión para facsímil

Bajo los auspicios de la Unión Internacional de Telecomunicaciones (UIT o ITU de las iniciales en inglés), en el seno del Comité Consultivo Internacional de Telefonía y Telegrafía (CCITT, designado actualmente como ITU-T, se desarrollaron estándares para la codificación y transmisión de señales sobre redes públicas de telecomunicaciones. Los primeros estándares, para compresión sin pérdidas se desarrollaron para aplicaciones de facsímil¹⁶ (fax), en que las imágenes exploradas son bitonales, es decir, los elementos de imagen toman solamente uno de dos posibles valores que pueden representarse mediante un bit.

En cualquier imagen bitonal, es decir, blanco y negro, se tienen zonas grandes que son, o bien totalmente blancas o totalmente negras, de modo que en una línea no es necesario incluir todos los bits correspondientes a cada elemento de imagen y basta codificar únicamente el recorrido, es decir el número de bits iguales a lo largo de una porción de la línea explorada. Esta codificación de recorrido permite conseguir un nivel apreciable de compresión similar a la que se hace en MPEG al explorar un bloque en zig-zag. La codificación de recorrido se combina luego con una codificación de Huffman y este es el esquema básico del estándar de codificación para facsímil. Este estándar está reflejado en dos Recomendaciones (ITU-T Rec. T.4¹⁷ e ITU-T Rec. T.6¹⁸) de la Unión Internacional de Telecomunicaciones. En la Recomendación T.4 se tienen dos posibles enfoques:

¹⁶ Hunter, R. and Robinson, A.H. "International digital facsimile coding standards. Proc IEEE 68(7):854-867, Jul. 1980

¹⁷ ITU-T (CCITT). Standardization of Group 3 facsimile apparatus for document transmission. Recommendation T.4. ISO, 1980.

¹⁸ ITU-T (CCITT). Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus. Recommendation T.6. ISO, 1984.

- *Código de Huffman modificado* (MH). En esta alternativa, la imagen se trata como una secuencia de líneas de barrido e inicialmente se realiza una descripción del recorrido de cada línea. Se aplica luego codificación de Huffman a la descripción (*recorrido, valor*) y se utiliza una codificación separada, también de Huffman para distinguir entre los recorridos de porciones blancas y negras. La tabla del código de Huffman es estática, es decir, no cambia de una imagen a otra. Al final de cada línea se inserta un código de *fin de línea* (EOL), principalmente para fines de detección de errores.
- *Código de lectura modificado* (MR). En este caso, los valores de los elementos de una línea previa, se usan como predictores de la línea actual. A continuación se realiza una descripción de recorrido y código estático de Huffman similar al caso anterior. También se utiliza un código de fin de línea. Para evitar la propagación de errores, la codificación MR se mezcla periódicamente con la MH.

En la Recomendación T.6 la técnica de codificación se conoce como MMR (Modified Modified Read). Este código es una simplificación del código MR, en que los mecanismos de protección contra errores se eliminan para aumentar la relación de compresión.

Las relaciones de compresión que se consiguen con los estándares T.4 y T.6 están entre 20:1 y 50:1 para documentos comerciales comunes. Sin embargo, para imágenes bitonales con técnica de medios tonos, la relación de compresión disminuye considerablemente, ya que los recorridos se reducen a sólo unos pocos elementos de imagen. En este caso el código de Huffman no resulta adecuado y es preferible utilizar un codificador aritmético adaptativo.

6.20.2 Estándard JBIG

Para comprimir con mayor eficiencia, tanto las imágenes de medios tonos como los documentos comerciales se desarrolló el estándar JBIG (Joint Binary Experts Group), designado como ISO/IEC IS 11544 y Recomendación T.82 de la ITU-T. En este estándar se incluye un modelador y un codificador aritmético. El modelador se emplea para estimar las probabilidades de símbolo que serán usadas luego por el codificador aritmético. Para imágenes de medios tonos, las relaciones de compresión con JBIG son de 2 a cinco veces mayores que con el estándar T.6. Un inconveniente de JBIG es que las implementaciones de software en computadoras de propósito general son de dos a tres veces más lentas que para los estándares T.4 o T.6.

6.20.3 Estándares para teleconferencia de vídeo

En bastantes aplicaciones de tipo científico, médico, industrial y comercial, es necesario transmitir imágenes con escaso movimiento, ya sea un orador o representaciones gráficas con escasa o nula animación, lo que en el caso de imágenes digitales, permite niveles de compresión relativamente altos, mayores que para las señales de televisión estándar. Esto dio lugar a que a principios de la década de 1980 se produjeran los primeros estándares por el CCITT, reflejados en las Recomendaciones H.120 y H.130. Dichos estándares proporcionaban los lineamientos para transmitir señales de televisión PAL o NTSC a tasas binarias del orden de unos pocos Mbits/s integrando la conversión entre ambos sistemas en el proceso de codificación.

Los estándares anteriores solamente encontraron aceptación en Europa. En los Estados Unidos y Japón se continuó trabajando sobre técnicas de codificación de vídeo que permitían conseguir mejor calidad de señal a tasas binarias inferiores a 2 Mbits/s. A finales de la década de 1980 el trabajo conjunto de fabricantes y operadores de telecomunicaciones dio lugar al estándar H.320

para videoconferencia sobre redes conmutadas del tipo ISDN. De hecho, el estándar H.320 comprende una serie de recomendaciones que incluyen las mostradas en el diagrama general de bloques de la figura 6.16.

- **H.261.** Es un algoritmo de compresión de vídeo a tasas binarias desde 64 Kbits/s hasta 1.92 Mbits/s.
- **G.722, G.726 y G.728.** Son algoritmos para compresión de audio a tasas desde 16 Kbits/s hasta 64 Kbits/s.
- **H.221.** Especifica la estructura de cuadro para el multiplexado de audio, vídeo y datos en un flujo binario único.
- **H.230 y H.242.** Especifican los protocolos de enlace (handshaking) entre los equipos que satisfacen el estándar H.320.
- **H.233.** Esta recomendación permite a los fabricantes seleccionar entre tres métodos de cifrado en su equipo H.320. El DES, usado en Estados Unidos, SEAL en Japón y BCRYPT en el Reino Unido. No se ha resuelto aún el problema de cómo pasar las llaves de cifrado entre una localidad y otra.

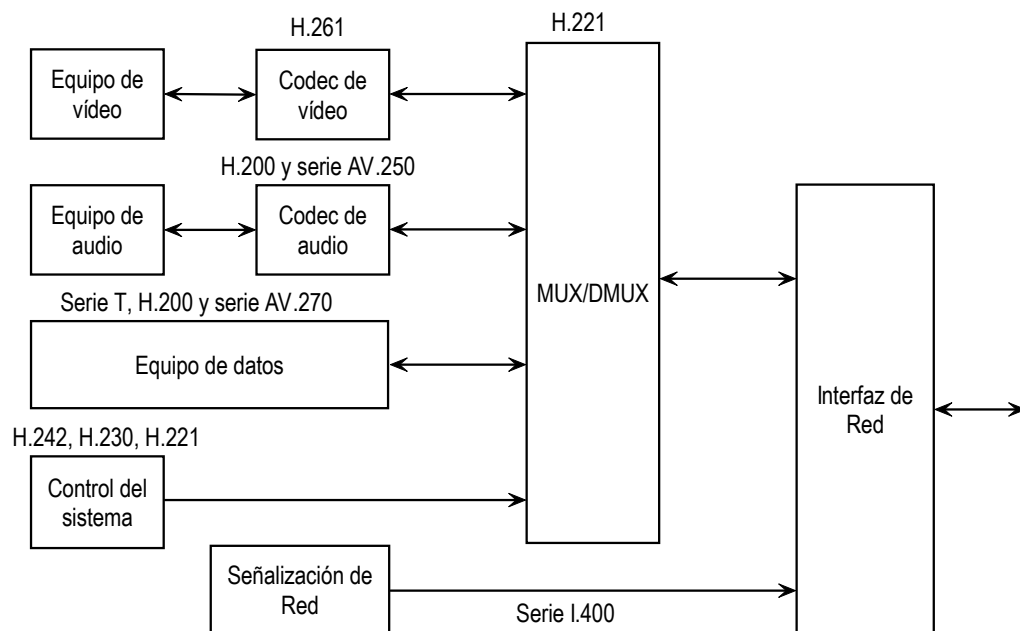


Fig. 6.16. Diagrama genérico de bloques del estándar H.320.
(Basado en *Image and Video Compression Standards*. 2nd Ed.
V. Bhaskaran and K. Konstantinides. Kluwer Academic Pub.1997)

6.20.4 Estándar de codificación de vídeo H.261

El estándar H.261 es la parte del estándar H.320 relativo a la codificación de vídeo. El H.230 fue ratificado en Ginebra en diciembre de 1990. En esa época aún no se había desarrollado la televisión digital y un estándar para videoconferencia que fuese compatible con los sistemas PAL y NTSC debía permitir la *interoperabilidad* entre los dos sistemas diferentes. Con este fin, se adoptó un nuevo *formato común intermedio*, designado como CIF (Common Intermediate Format). Tanto los sistemas de 525 como de 625 líneas, NTSC o PAL, deben, por consecuencia, incluir módulos de pre y postprocesado para convertir a CIF o de éste al sistema específico.

6.20.5 Formato CIF

Es un formato de barrido *no entrelazado*, basado en 352 elementos por línea (pixels), 288 líneas no entrelazadas por cuadro y 30 cuadros por segundo. Estos valores representan la mitad de las líneas activas tanto de una imagen de 625 líneas, 25 cuadros por segundo, como de una de 525 líneas y 30 cuadros por segundo. En estas condiciones, los sistemas de 625/25 sólo necesitan efectuar una conversión de la frecuencia de cuadro y los de 525/30, convertir el número de líneas.

Las señales de color se codifican de acuerdo a la Recomendación 601 del CCIR, con una componente de luminancia y dos de crominancia. Las componentes de crominancia se submuestrean por un factor de 2, tanto horizontal como verticalmente, de modo que se tienen 176 elementos (pixels) por línea y 144 líneas por cuadro para estas componentes. La relación de aspecto es 4:3.

Para aplicaciones de muy baja velocidad de transmisión, además del CIF, los codificadores de vídeo pueden utilizar también un formato de “un cuarto de CIF” (QCIF), con la mitad del número de elementos por línea y de líneas necesarias en el formato CIF. Desde el punto de vista de implementación, todos los codificadores y decodificadores deben ser capaces de trabajar con QCIF, siendo opcional el funcionamiento con CIF.

6.20.6 Codificación H.261

El algoritmo de codificación H.261 utiliza una combinación de DCT y codificación diferencial. Los elementos principales son semejantes a los de MPEG: transformada DCT, cuantificación y codificación de longitud variable.

El procedimiento de codificación es similar al utilizado en JPEG y MPEG, excepto que el cuadro de vídeo se convierte primero a un cuadro CIF y se almacena como tal en una memoria de cuadro. En esta etapa se puede realizar también algún preprocesado y filtrado de ruido. La DCT también trabaja sobre bloques de 8×8 elementos. La combinación de cuatro bloques de luminancia y un bloque de cada una de las componentes de crominancia constituyen un macrobloque.

La predicción de cuadros se lleva a cabo de manera semejante a MPEG-1, excepto que en H.261 solamente se utilizan cuadros I y P. La organización de los macrobloques puede ser de tipo intracuadro o intercuadro, en que la codificación diferencial permite al codificador funcionar bien sobre los macrobloques de entrada (modo *intra*), o sobre los macrobloques diferenciales entre el cuadro actual y el anterior (modo *inter*). En H.261 no es necesario codificar y transmitir todos los macrobloques. Cuando la tasa binaria es baja, macrobloques, y aún hasta tres cuadros completos pueden omitirse. El criterio para la selección de los macrobloques o cuadros que se omiten, así como el mecanismo de control para seleccionar codificación intracuadro o intercuadro no forman parte del estándar y pueden variar dinámicamente, dependiendo de la complejidad de la señal de entrada y de las restricciones sobre la tasa binaria de salida. El codificador puede incluir, de manera opcional, la estimación y compensación de movimiento.

El flujo binario de salida se dispone jerárquicamente en cuatro capas:

- El cuadro o imagen constituye la capa superior.
- Cada cuadro se divide en grupos de bloques (GOB). Un grupo de bloques es, o bien 1/12 de un cuadro CIF o 1/3 de un cuadro QCIF.
- Cada GOB se subdivide en 33 macrobloques.
- Cada macrobloque consiste de seis bloques de 8×8 elementos de imagen, de los cuales cuatro son de luminancia y dos de crominancia (un bloque por cada componente).

6.20.7 Decodificación H.261

En la figura 6.17 se ilustra esquemáticamente el decodificador de vídeo H.261. Después de la corrección de errores que, en este caso es opcional, la entrada de datos de vídeo comprimidos pasa a un buffer y es procesada a continuación por el decodificador de longitud variable.

Los datos decodificados son separados de acuerdo al contenido de las tramas y luego procesados mediante un cuantificador inverso y una transformada inversa del coseno discreto. Dependiendo del modo de transmisión (intracadro o intercadro), los macrobloques de un cuadro previo pueden también agregarse a los datos actuales para configurar los datos reconstruidos. Finalmente, los datos CIF decodificados pueden convertirse al formato adecuado para su visualización.

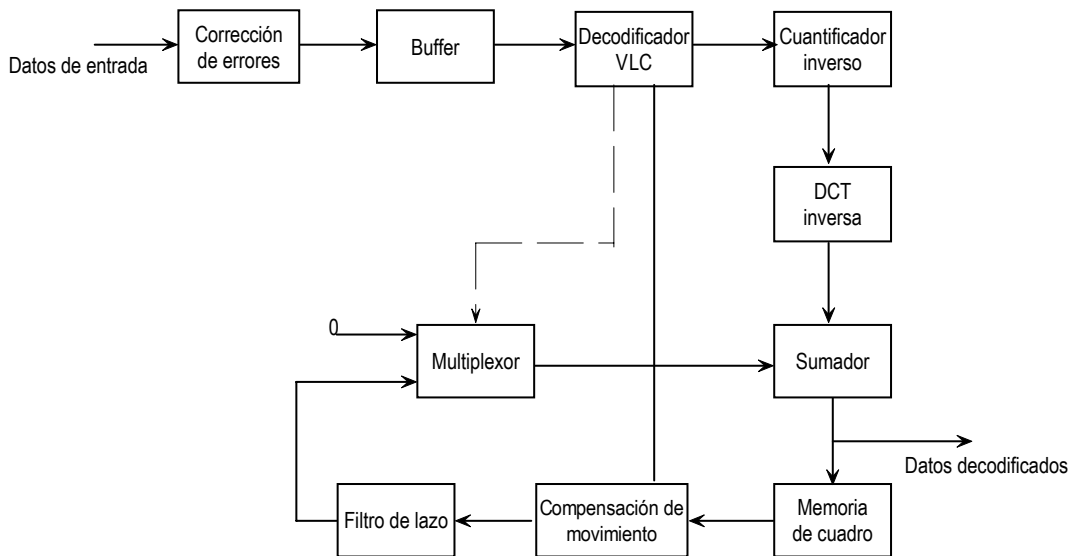


Fig. 6.17. Diagrama de bloques de un decodificador H.261.

6.20.8 Estándar H.324 para Teleconferencia de Vídeo

El estándar H.324 es un estándar para transmisión de audio, vídeo y datos a baja velocidad, principalmente para aplicaciones de teleconferencia, orientado a multimedia y que, de manera similar al H.320, comprende varios estándares. En la figura 6.18 se ilustra esquemáticamente la estructura de este estándar, así como la de los estándares que lo conforman.

El estándar H.324 es similar al H.320 descrito anteriormente. Sin embargo, se incluyen cuatro recomendaciones nuevas:

- **H.263.** Especifica el algoritmo de codificación de vídeo para comunicaciones de baja tasa binaria.
- **G.273.1.** Codificador de voz y audio, de doble tasa binaria (5.3 y 6.3 Kbit/s) para aplicaciones multimedia.
- **H.223.** Protocolo de multiplexado.
- **H.245.** Protocolo de control que puede usarse para especificar las capacidades de codificación y decodificación del transmisor y del receptor.

Esta característica puede utilizarse para permitir diversos modos de codificación. Los procedimientos en H.245 también están pensados para su utilización en videoconferencia sobre redes ATM (Recomendación 310), o en redes de área local (LAN), sin ancho de banda garantizado (Recomendación H.323). Estas características permiten la interconexión de terminales H.324 con sistemas de videoconferencia basados en ATM o LAN.

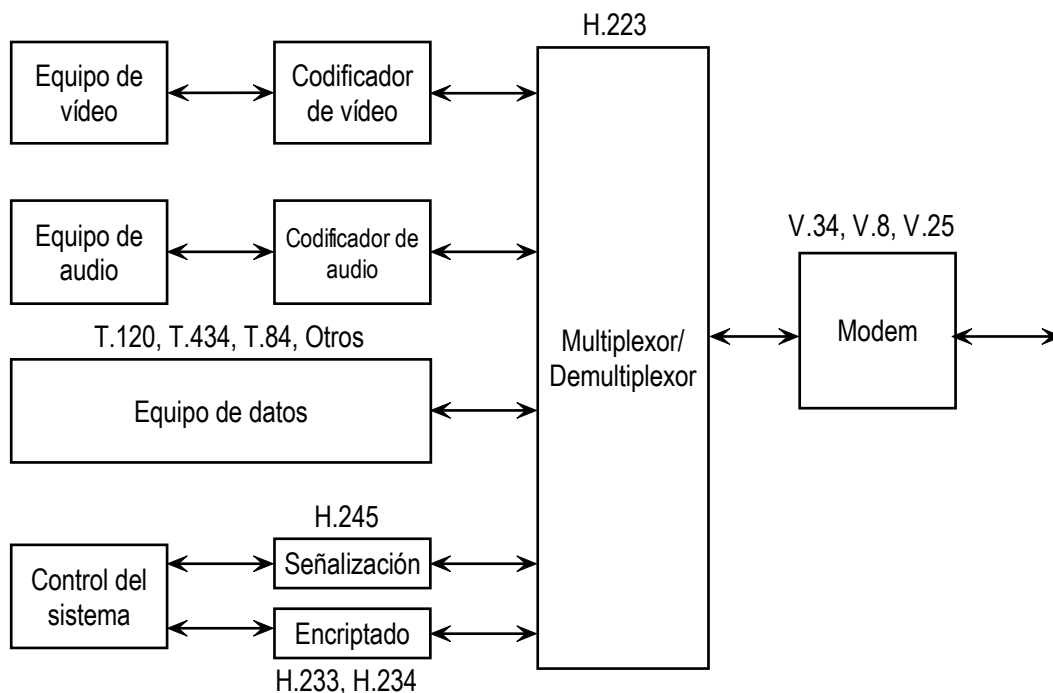


Fig. 6.18. Diagrama genérico de bloques del estándar H.324.
(Basado en *Image and Video Compression Standards*. 2nd Ed.
V. Bhaskaran and K. Konstantinides. Kluwer Academic Pub.1997)

Para comunicaciones mediante modem, H.324 especifica el uso del procedimiento V.8 (V8-bis) para el inicio y fin de la transmisión de datos, así como el modem V.34. El estándar V.34 soporta tasas binarias hasta de 28.8 Kbit/s. En aplicaciones de datos, H.324 especifica el uso del protocolo T.120 como un posible medio para el intercambio de datos. Para transmisión de imágenes fijas puede usarse la Recomendación T.84 (JPEG).

6.20.9 Estándar de codificación de vídeo H.263

El algoritmo de codificación de vídeo H.263 utiliza muchas características de H.261 y de MPEG-1. Así, H.261 y H.263 comparten las mismas estructuras de codificación y decodificación ya descritas. En esta sección solo se comentarán las principales diferencias.

Tasa binaria. La Recomendación H.263 no restringe la tasa binaria. Sin embargo su objetivo es hasta velocidades de 64 Kb/s. En aplicaciones sobre redes telefónicas conmutadas, mediante modems de 28.8 Kbit/s, la máxima velocidad de transmisión de vídeo puede ser alrededor de 20 Kbit/s, ya que el ancho de banda restante debe reservarse para voz, datos y señales de control. En H.261 la tasa binaria objetivo es de $p \times 64$ Kbit/s, donde p es un número entero comprendido entre 1 y 30.

Formato de Imagen. En la tabla siguiente se muestran los formatos de imagen soportados por el estándar H.263.

Formato	Luminancia	Crominancia
Sub-QCIF	128 × 96	64 × 48
QCIF	176 × 144	88 × 72
CIF	352 × 288	176 × 144
4CIF	704 × 576	352 × 288
16CIF	1408 × 1152	704 × 576

Tabla 6.5 Formatos de imagen soportados por el estándar H.263.

Todos los decodificadores H.263 deben ser capaces de funcionar con flujo de datos en formatos sub-QCIF y CIF. El codificador debe ser capaz de funcionar con, por lo menos uno de los formatos sub-QCIF o QCIF, pero no está obligado a funcionar con ambos. Para la visualización, los decodificadores pueden sobremuestrear las imágenes sub-QCIF o submuestrear las QCIF. Estas restricciones aseguran que los codecs H.263, aunque puedan implementarse a bajo costo, permiten la interconexión en red con equipos que utilicen el estándar H.261.

Estructura de los grupos de bloques (GOB). Tanto H.261 como H.263 utilizan una sintaxis jerárquica, en la que secuencia de vídeo se descompone en imágenes, grupos de bloques, macrobloques (MB) y bloques. En el estándar H.263, a fin de proporcionar robustez contra errores, cada grupo de bloques contiene solamente una fila de macrobloques. Así, para QCIF, cada grupo de bloques tiene 11 macrobloques en lugar de $11 \times 3 = 33$ bloques en H.261. Esta característica facilita la inserción de señales adicionales de sincronismo para mejorar el comportamiento frente a errores.

Corrección de errores. En el estándar H.263 no se especifica ningún procedimiento para detección o corrección de errores, si bien se describe un esquema opcional en el Anexo H de la Recomendación.

Otras diferencias adicionales entre H.261 y H.263 son:

- H.263 tiene mayor precisión (medio pixel) en la estimación de movimiento que H.261 (un pixel).
- H.263 no utiliza el filtro de lazo de H.261 para reducir los efectos de bloqueo en la estimación de movimiento por bloques. La precisión adicional en la estimación y compensación de movimiento de H.263 empleando interpolación lineal que, como efecto colateral actúa como filtro de paso bajo.
- En H.261, la matriz de cuantificación puede cambiarse en base a macrobloques. En H.263, dentro de un grupo de bloques, las transiciones de la matriz de cuantificación entre macrobloques se restringen a dos.
- Los coeficientes de la DCT se codifican diferente. En H.263 se realiza una codificación de longitud variable (VLC) según recorrido, nivel y último coeficiente, es decir 3-D. En H.261 la codificación VLC es 2-D: recorrido y nivel. La codificación VLC 3-D permite alcanzar mayores niveles de compresión.

- En H.261 la dirección del macrobloque se utiliza para indicar cuantos macrobloques se han omitido. En H.263 se transmite un único bit por cada macrobloque omitido.
- A diferencia de H.261, no hay un modo de funcionamiento definido en H.263 para imágenes fijas. En su lugar pueden usarse los estándares JPEG (T.81, T.84) para transmitir este tipo de imágenes.

El estándar H.263 especifica técnicas opcionales de codificación que pueden mejorar aún más el rendimiento, entre ellas, el empleo de vectores de movimiento no restringidos, un modo de predicción avanzada, un modo de codificación aritmética basada en sintaxis y un modo de cuadros PB. En la UIT se continua trabajando en extensiones de H.263, que han dado lugar al H.263+. Estos trabajos, en cierta medida se alinean con los de MPEG-4.

Bibliografía adicional

- [1] - Mitchel, J.L., Pennebaker, W.B., Fogg, C.E. and LeGall, D. J. *MPEG Video Compression Standard*. Chapman & Hall. International Thompson Publishing. 1996. *Es un texto que ofrece una descripción bastante amplia de MPEG, orientada particularmente a MPEG-1, incluyendo una parte interesante de la codificación de los algoritmos.*
- [2] - ATSC Digital Television Standard. US Advanced Television Systems Committee. 16 Sep. 1995.
- [3] - Guide to the use of the ATSC Digital Television Standard. US Advanced Television Systems Committee. Oct. 1995. *Este documento, igual que el anterior se ha utilizado extensamente en la preparación de este capítulo sobre compresión de vídeo. De la información disponible, son quizá los más completos desde el punto de vista de aplicaciones, sobre todo porque tratan de un sistema ya probado y adoptado en los Estados Unidos. Pueden obtenerse vía Internet en <http://www.atsc.org>.*
- [4] - Stojancic, M. M. and Ngai, C. "Architecture and VLSI Implementation of the MPEG-2: MP@ML Video Decoding Process. SMPTE Jour. Vol. 104, Nº 2, pp. 62-72. Feb. 1995.
- [5] - ITU/SMPTE Tutorial. Digital Terrestrial Television Broadcasting. Society of Motion Picture and Television Engineers (SMPTE). White Plains, NY. 1994.
- [6] - Blumenfeld, S. "Video Compression". Broadcast Eng. vol. 37, Nº 2, pp. 24-32. Feb. 1995
- [7] - Epstein, S. "Video Compression 101". Broadcast. Eng. Vol. 38, Nº 2. pp. 70-74. Feb. 1996.
- [8] - Freed, K. "Video Compression". Broadcast Eng. Vol. 39, Nº 1, pp. 46-52. Jan. 1997.
- [9] - Nelson, L. J. "Video Compression". Broadcast. Eng. Vol. 37, Nº 11. pp. 42-46. Oct. 1995.
- [10] - Nelson, L. J. "Compression Techniques for Narrowband Channels". Broadcast Eng. Vol. 38, Nº 8, pp. 68-70. Aug. 1996.
- [11] - Bhatt, B., Birks, D. and Hermreck, D. "Digital Television: Making It Work. IEEE Spectrum. Vol. 34, Nº 10, pp. 19-28. Oct. 1997.
- [12] - Whitaker, J. "MPEG: What's it all about?". Broadcast Eng. Vol. 39, Nº 3, pp. 14-16. March 1997.
- [13] - Proceedings of the SMPTE Advanced Television and Imaging Conference. "Unveiling New Technologies and Applications". The Society of Motion Picture and Television Engineers. White Plains, NY. 1994.

-
- [14] - *Digital Video. Concepts and applications across industries*. Theodore S. Rzeszewski, Editor. IEEE Press. Piscataway, NJ. 1995. Es una recopilación de artículos, muchos de ellos clásicos, relativos a temas de televisión digital.
- [15] - *Television Technology Today*. Theodore S. Rzeszewski, Editor. IEEE Press. Piscataway, NJ. 1985. Es una recopilación de artículos que, en muchos casos ya sólo tienen interés histórico, pero que ofrecen conceptos útiles e interesantes sobre codificación digital de vídeo, en particular la parte IV del libro.
- [16] - <http://mpeg.telecomitalialab.com/>
- [17] - <http://www.diffuse.org/fora.html>
- [18] - <http://streamingmediaworld.com/video/docs/MPEG4/>
- [19] - <http://www.atsc.org> .
- [20] - <http://www.iso.ch>
- [21] - <http://www1.ietf.org/IESG/LIAISON/ITU-1832.htm>
- [22] - <http://www.cselt.org/ufv/leonardo/mpeg/standards.htm>
- [23] - <http://www.itu.int/rec/>
- [24] - <http://www-mobile.ecs.soton.ac.uk/>