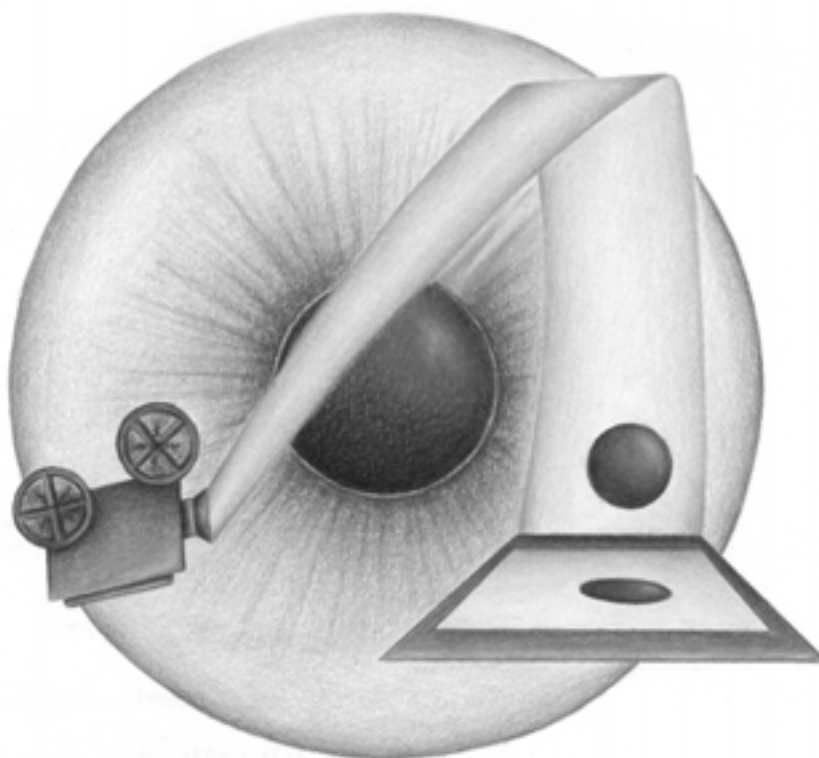


Estándares de compresión de imágenes en movimiento



Índice

Etapa 1: MPEG (Moving Picture Expert Group)	5
Introducción	5
Imágenes en MPEG	5
Imágenes I	6
Imágenes P	6
Imágenes B	6
Group of pictures (GOP)	7
Funcionamiento de un codificador MPEG	9
Esquema de un codificador de MPEG	10
Estimación de movimiento	10
Compensación de movimiento	10
Sumador comparador	11
DCT	11
Cuantificador	12
VLC	13
Funcionamiento de un decodificador de MPEG	13
Distintas posibles formas de trabajo de un decodificador	14
Etapa 2: MPEG-1 (ISO/IEC 11172)	16
Introducción	16
MPEG-1 video (ISO/IEC 11172-2)	16
MPEG audio (ISO/IEC 11172-3)	18
Etapa 3: MPEG-2 (ISO/IEC 13818)	19
Introducción	19
MPEG-2 video (ISO/IEC 13818-2)	19
Perfiles y niveles en MPEG (profiles & levels)	21
Problemas del MP@ML en aplicaciones profesionales	23
MPEG-2 4:2:2 MP@ML	23
MPEG-2 audio (ISO/IEC 13818-3)	25

Etapa 1: MPEG (*Moving Picture Expert Group*)

Introducción

El *Moving Picture Expert Group* (MPEG) es una entidad formada por un grupo de expertos especialistas en el análisis, tratamiento, diseño y desarrollo de sistemas de imagen digital en movimiento. Este grupo se encuentra bajo los auspicios de la Organización Internacional de Estandarización (ISO) y la Comisión Electrotécnica Internacional (IEC), que se creó con el propósito de elaborar un conjunto de recomendaciones y especificaciones técnicas necesarias para la compresión y multiplexado de contenidos audiovisuales.

Este grupo empezó por crear el MPEG-1. En este estándar se definen los parámetros de codificación y compresión de vídeo para aplicaciones domésticas con una calidad de imagen similar a la del VHS y de sonido similar a la del CD, con un ancho de banda de hasta 1,5 Mb/s. Con el paso del tiempo y la consiguiente evolución tecnológica se planteó el MPEG-2, en el que las técnicas utilizadas para la compresión son bastante más complejas y exigen un cómputo más complejo de los procesadores. Con este estándar se consigue vídeo con calidad desde MPEG-1 hasta HDTV (televisión de alta definición) y audio de calidad multicanal. Cabe destacar que este conjunto de normas admite diferentes niveles de combinaciones de trabajo. De este modo, podemos encontrar gran cantidad de combinaciones MPEG con diferentes grados de calidad y anchos de banda, para poder cubrir el máximo de necesidades que demande el mercado audiovisual.

Actualmente ya se han definido y se están definiendo nuevos estándares MPEG, como el MPEG-4 y el MPEG-7. Estos estándares utilizan complejos algoritmos de compresión, los cuales requieren potentes procesadores o *softwares* para su tratamiento. Con éstos se consiguen elevados niveles de calidad de imagen para aplicaciones multimedia (no *broadcast*) con anchos de banda muy reducidos. Éste sería el caso del conocido DivX (MPEG-4) con el que se consigue almacenar una película entera en un CD de 650 Mb.

Imágenes en MPEG

Como se indica en el ejemplo, en MPEG se puede encontrar distintos niveles de imágenes en función de la técnica utilizada para su compresión. Éstas se pueden clasificar en tres grupos: imágenes I, imágenes P e imágenes B.

En el siguiente esquema se puede ver un proceso habitual de trabajo con MPEG. Éste podría simbolizar la emisión de un programa, en el que las imágenes originales (O,

O, O...) procedentes del *player* (Betacam, DVC-pro, DVcam, etc.) del centro de emisión se dirigen al codificador de MPEG, que las comprimirá y las convertirá en imágenes I, P o B según la técnica de compresión aplicada sobre cada una de ellas, con lo que se conseguiría una señal de reducido ancho de banda. La cadena de imágenes I, P y B se transmite por vía terrestre, cable o satélite hasta llegar a su destino, en el que cada imagen será descodificada y reconvertida a su formato original (O, O, O...) para poder ser grabada.



Imágenes I

Imágenes I (Intra): son aquel conjunto de imágenes comprimidas a partir de la información de ellas mismas. De este modo, la imagen I del fotograma 1 se ha deducido sólo a partir del fotograma 1, la imagen I del fotograma 2 se ha deducido sólo a partir del fotograma 2, etc. Por lo tanto, una imagen I contiene todos los elementos necesarios para su reconstrucción en el descodificador.

El proceso de compresión que utiliza este tipo de imágenes consiste en descomponer en bloques de 8×8 píxeles la imagen, a los que posteriormente se aplica la DCT. Los coeficientes de la DCT son cuantificados mediante una tabla de cuantificación predefinida y posteriormente se codifican mediante un código de longitud variable VLC, por lo que se trata de imágenes totalmente autónomas, ya que no dependen de otras. Algunos sistemas de edición en vídeo utilizan este tipo de imágenes como formato de trabajo, ya que se trata de información comprimida que a su vez permite precisión de fotograma al realizar la posproducción.

Las imágenes I pueden crear otras imágenes P o B.

Imágenes P

Imágenes P (Preditas): este tipo de imágenes utiliza compresión temporal, es decir, la compensación de movimiento. Se trata de imágenes comprimidas a partir de la información de otras imágenes I o P anteriores. Debido a que la predicción no es perfecta nunca podremos poner gran cantidad de imágenes P entre dos imágenes I, ya que el error de predicción se iría acumulando y se alejaría mucho de la realidad.

Las imágenes P se pueden utilizar para crear otras imágenes P o B.

Imágenes B

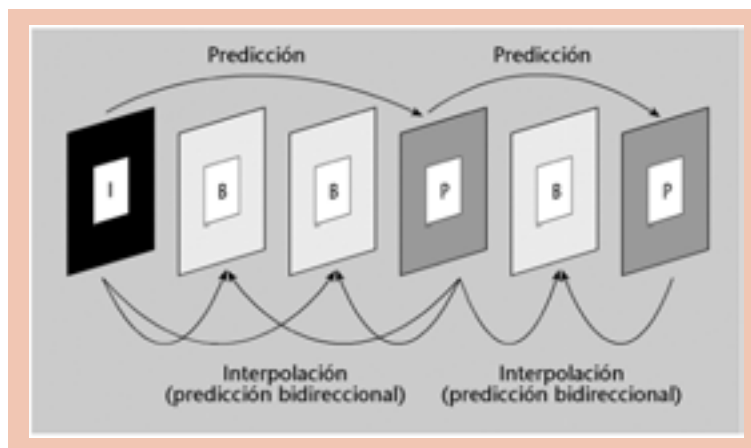
Imágenes B (Bidireccionales): se codifican a partir de las imágenes I o P que las preceden y delimitan, por lo que utilizan compensación de movimiento. Estas imágenes

son las que ofrecen una tasa de compresión más elevada, o sea, son las que ofrecen peor calidad y, a su vez, menor tamaño de datos.

Si se contempla que estas imágenes necesitan para su generación imágenes posteriores a ellas, se puede deducir rápidamente que para descodificar una secuencia de imágenes B será necesario un cierto tiempo de espera, hasta recibir la siguiente imagen I o P.

En el siguiente gráfico se puede apreciar que una imagen I crea las imágenes P y B mediante predicción, que las imágenes P pueden crear otras imágenes P y B y que a partir de una imagen B no se puede crear ninguna imagen.

En la norma MPEG se contemplan tres tipos de imágenes en función de la técnica de compresión aplicada para su creación: imágenes I, imágenes P e imágenes B.



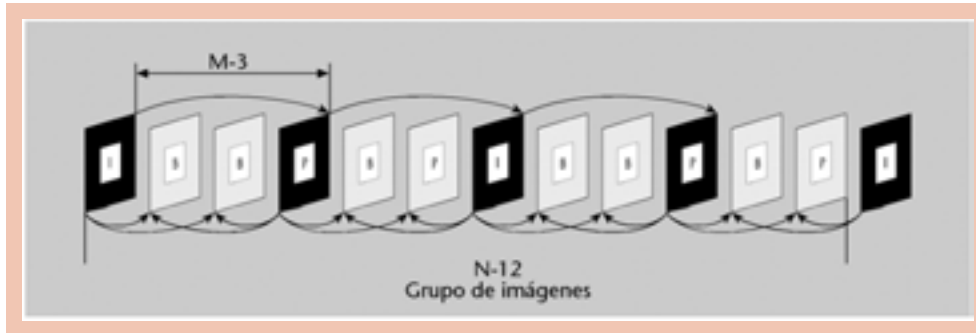
Group of pictures (GOP)

Una secuencia de imágenes combinadas I, P y B se denomina GOP (*group of pictures*). Un GOP puede tener varias estructuras en función del orden de las imágenes. El GOP más simple que se puede encontrar es un grupo de imágenes I, y el GOP más complejo sería una combinación de los tres tipos de imágenes I, P y B.

El tamaño del GOP es variable; existen GOP formados por pocas imágenes y GOP formados por muchas. De todas formas, el tipo de combinaciones entre imágenes I, B y P y el tamaño del GOP lo decide el codificador de MPEG.

Habitualmente se trabaja con GOP de doce imágenes, siguiendo el orden establecido en la gráfica.

En la siguiente secuencia de imágenes interviene una imagen del tipo I, dos del tipo P y seis del tipo B. Dada esta estructura, se deduce que la imagen P1 ha sido creada a partir de una predicción de movimiento hacia adelante (*forward*) de la imagen I1. Igualmente, P2 se crea a partir de aplicar la compensación de movimiento a la imagen P1 y la P3 toma como referencia la imagen P2.



Es importante tener en cuenta que si por alguna razón una imagen I o P, que se encontrará en la memoria del descodificador, se ve afectada por algún error, las imágenes que se reconstruyan a partir de ésta acarrearán dicho error aunque los errores de predicción estén correctamente calculados. Por esta razón, es importante destinar una cantidad importante de bits a la protección de las imágenes que se utilizan como referencia, con la finalidad de minimizar la posible propagación de errores en el proceso de reconstrucción de otras imágenes en el descodificador.

Existen dos parámetros que nos servirán para definir la estructura de un GOP. Éstos son el parámetro “N”, que nos indicará la cantidad de imágenes que tiene el GOP, y el parámetro “M”, que nos indica la distancia existente entre imágenes “referencia” I o P.

En el anterior ejemplo se puede observar una estructura con la que se consigue una calidad satisfactoria, y a su vez una buena resolución de acceso aleatorio (tiempo de acceso < 0,5 segundos, en un sistema de 25 imágenes/segundo), donde los parámetros más comunes son $M = 3$ y $N = 12$, es decir, grupos de doce imágenes I, P y B, en los que la distancia entre imágenes I y P es de tres imágenes.

En el GOP se puede apreciar que la mayor cantidad de imágenes son del tipo B. Estas imágenes son bidireccionales, lo que implica que han sido creadas a partir de las imágenes I y P más próximas, mediante predicción hacia delante (*forward*) y predicción hacia atrás (*backward*). Esto implica que si se quiere descodificar la imagen B5 del GOP previamente se necesita haber descodificado la imagen I1; a partir de ésta se descodifica la imagen P1, y de ésta, la P2. Cuando ya se tiene la imagen P2 podemos recuperar la imagen B5. Por esta razón las únicas imágenes que realmente nos

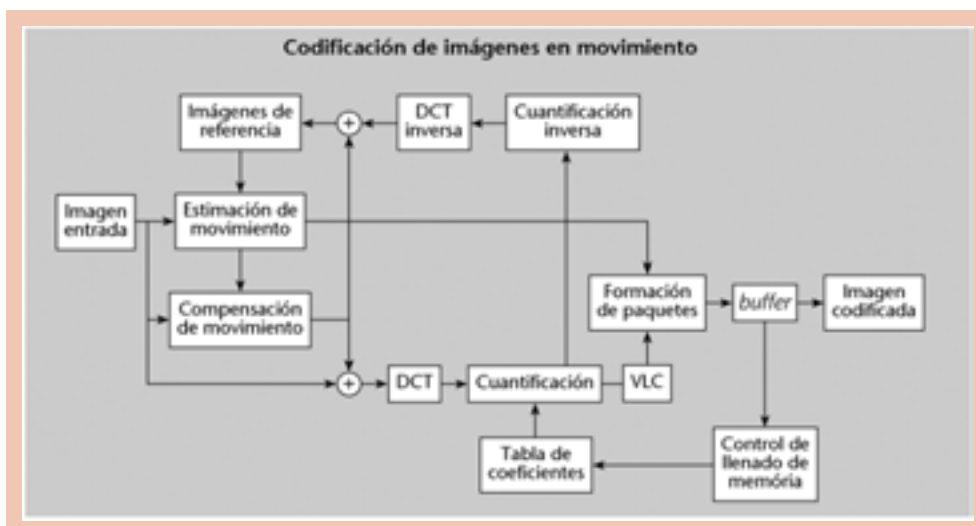
permiten un acceso aleatorio serán las imágenes I, ya que son las únicas que no requieren de otras imágenes para su reconstrucción, son totalmente autónomas.

Estas imágenes tienen una propiedad muy importante: no se utilizan nunca como imagen de referencia, entre otras razones a causa de su baja calidad. Por esta razón nunca acarrearán errores en el decodificador, por lo que se les puede aplicar un elevado factor de compresión, ya que una vez utilizadas no se van a necesitar más.

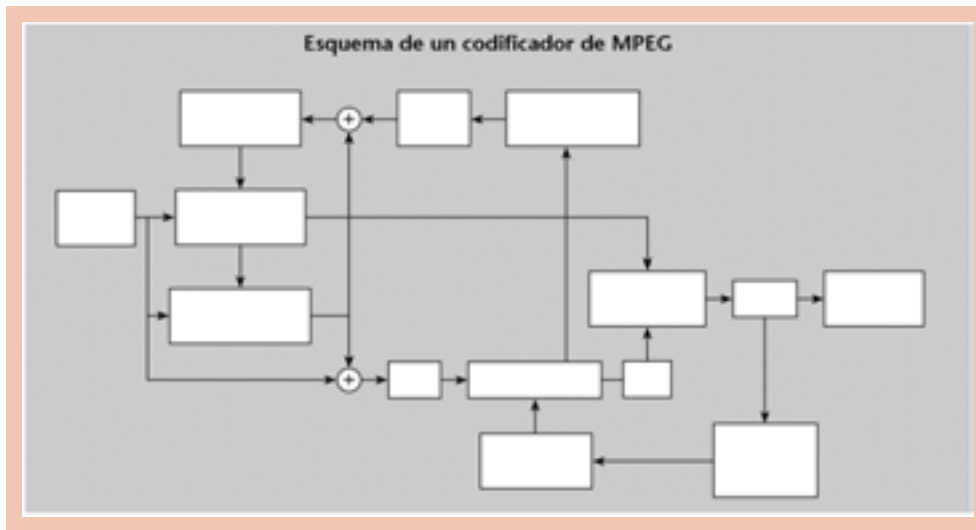
Usar imágenes B en MPEG implica trabajar con predicción bidireccional (*forward* y *backward*), lo que supone que el orden de transmisión de imágenes sea distinto del orden con el que se van a visualizar. Si el decodificador recibe un grupo de imágenes con un orden distinto del que va a tener que reproducir, necesita una memoria *buffer* que sea capaz de almacenar todas las imágenes de referencia necesarias para la correcta decodificación de la totalidad del GOP.

Funcionamiento de un codificador MPEG

La principal función de un codificador de MPEG es la de determinar qué grado de compresión aplica a cada imagen en función de la calidad y ancho de banda que se desea. Para realizar este proceso necesita analizar la secuencia de imágenes de vídeo digital que le entran, según la norma utilizada (4:2:2, 4:2:0, 4:1:1, etc.) y determinar la cantidad de imágenes de cada tipo I, P y B que va a utilizar para conseguir el ancho de banda deseado con la máxima calidad de imagen. Posteriormente tendrá que reordenar las imágenes en la salida en función del tipo de predicción I, P o B utilizado, de esta forma al decodificador le llegarán las imágenes en el orden adecuado para poder ser correctamente recuperadas.



Esquema de un codificador de MPEG



Si se analiza el proceso de un codificador de MPEG, se observa que una imagen que entra en el codificador se dirige al circuito de estimación de movimiento, al circuito de compensación de movimiento y, a su vez, al sumador.

Estimación de movimiento

En el circuito de estimación de movimiento se calcularán los vectores que se aplicarán a partir de la imagen de referencia para deducir una imagen predicha, sea P o B. Lógicamente, si la imagen que se predice es P, sólo se necesitará una imagen de referencia I u otra P anteriores a la que se está prediciendo.

Si la imagen que se deduce es una imagen B será necesaria una imagen I o P anterior a la imagen B que se está calculando en el caso de que la predicción sea *forward*, y una imagen I o P posterior a la imagen B que se está calculando, en el caso de que la predicción sea *backward*. En este caso se requerirá de una memoria que almacene la futura imagen B hasta que no llegue la siguiente imagen I o P de referencia que se utilizará para calcular los vectores de desplazamiento.

El circuito de estimación de movimiento es el "cerebro" principal del codificador, ya que en éste se realizan las predicciones de desplazamiento. Cuanto más exacta sea una predicción, menor será el error de predicción que se tendrá que codificar posteriormente, con lo que se reducirá el ancho de banda.

Compensación de movimiento

En este circuito llega la imagen original y los vectores de movimiento, y ahí se aplicarán los vectores de movimiento a la imagen original, con lo que se obtendrá la imagen predicha P o B.

Sumador comparador

Esta etapa es la encargada de comparar la imagen real de la entrada con la imagen que se ha predicho. En el caso de que la predicción que se ha realizado sea perfecta, la imagen real y la imagen predicha serán idénticas y, por consiguiente, el error de predicción será nulo, de modo que sólo se transmitirán los vectores de desplazamiento.

En el caso de que la predicción sea incorrecta, la imagen real y la predicha serán muy distintas; todas estas diferencias se tendrán que codificar. El error que se codifique se analizará mediante la DCT y se le aplicará un factor de compresión en la cuantificación.

DCT

En esta etapa del proceso, el circuito DCT analiza el contenido espectral de cada bloque de la imagen mediante la transformada discreta del coseno. Determinando los coeficientes de la matriz de la DCT se puede visualizar la cantidad de componente continua (DC) y de alta frecuencia (AC) que tiene el bloque.



Es importante recordar que cuando se realiza el proceso de codificación de vídeo, los coeficientes de los bloques representan los distintos niveles de cuantificación. Habitualmente se trabaja a 8 bits, o sea, existen 256 posibles niveles de cuantificación (del 0 al 255), en los que un píxel negro estaría codificado con el coeficiente 0 y un píxel blanco, con el 255.

A su vez, cuando se calculan los coeficientes de la matriz de la DCT, éstos son recuantificados a 11 bits, es decir, 2.048 niveles de cuantificación, los cuales se sitúan entre el 0 y el 2.047 para el coeficiente de continua DC, y entre el -1.024 y el 1.023 para los coeficientes de alterna AC.

Es determinante la capacidad que tiene la DCT para concentrar en pocos coeficientes toda la información correspondiente al bloque de imagen; dado que se trabajará puntualmente con imágenes de extremada resolución, lo habitual será que gran cantidad de coeficientes asociados a alta frecuencia tengan un valor nulo.

Cuantificador

En el proceso de cuantificación se aplica un factor a cada coeficiente de la matriz de la DCT para determinar el grado de compresión que se aplica a cada frecuencia. El conjunto de todos los factores de compresión definen la matriz de cuantificación.

Ésta se utiliza también para asegurar un ancho de banda (*bit/rate*) lo más constante posible y, sobre todo, para controlar el nivel de llenado de la memoria (*buffer*) en la salida del conversor.

En este punto, se realiza la codificación por umbral asignando el valor 0 a cualquier coeficiente de la matriz de la DCT que se encuentre por debajo del umbral definido.

Existen varias aplicaciones que requieren un flujo de bits constante en la salida del codificador. Por esta razón, se utiliza una memoria de almacenamiento que se vaya llenando en instantes de elevada entropía y vaciando cuando exista mucha redundancia.

Rara vez se dispondrá de un contenido visual en el que la relación entre la redundancia y la entropía sea constante, ya que por lo general estos parámetros irán cambiando constantemente en función del grado de variación de la imagen en el tiempo y la definición. Por esta razón, los coeficientes de cada bloque de imagen serán distintos, al igual que los coeficientes de la DCT y, por lo tanto, el ancho de banda que requiere la imagen. Cuanto más cambiante y con mayor resolución sea una imagen, mayores valores de ancho de banda se obtendrán. Así pues, la calidad de la imagen irá variando según la cantidad de redundancia y entropía que contenga.

Las matrices de cuantificación utilizadas para la compresión de imagen también se tendrán que transmitir para que el decodificador pueda aplicar la matriz inversa de cuantificación y recuperar, así, la totalidad de la información transmitida.

172	889	221	45	13	3	0	0	16	11	10	16	24	40	51	51
930	738	372	28	11	0	32	0	12	12	14	19	26	58	60	55
85	237	129	6	50	10	7	12	14	13	16	24	40	57	69	56
32	45	18	33	14	5	0	0	14	17	22	29	51	87	80	62
0	13	0	2	10	0	2	0	18	22	37	56	68	109	103	77
0	9	11	6	0	0	0	0	24	35	55	64	81	104	113	92
3	5	0	0	8	0	7	0	49	64	78	87	103	121	120	101
0	0	0	7	0	0	0	0	72	92	95	98	112	100	103	99
Matriz de la DCT del bloque de 8x8 píxeles								Matriz de cuantificación							
172	889	221	45	13	3	0	0								
930	738	372	28	11	0	32	0								
85	237	129	6	50	10	7	12								
32	45	18	33	14	5	0	0								
0	13	0	2	10	0	2	0								
0	9	11	6	0	0	0	0								
3	5	0	0	8	0	7	0								
0	0	0	7	0	0	0	0								
Matriz de la DCT cuantificada															

Observad que los coeficientes de la matriz de cuantificación aumentan con los patrones de alta frecuencia, lo que indica que los pasos de cuantificación a alta frecuencia serán mayores que a baja frecuencia (con lo que se destinará menos precisión a las componentes de mayor resolución). Observad también que los coeficientes verticales son distintos de los horizontales, lo que indica que la matriz de cuantificación no trata de la misma forma la resolución vertical y la horizontal. Esto se ha diseñado así debido a la naturaleza de la visión humana.

Una vez aplicada la cuantificación se procederá a realizar la lectura en zigzag, empezando la lectura por el coeficiente de mayor peso DC para acabar por el de menor peso de AC (el de mayor frecuencia).

VLC

Antes de transmitir los errores de predicción, con los correspondientes vectores de desplazamiento y las matrices de cuantificación utilizadas para comprimir el error de predicción o bloques intra, se tendrán que “empaquetar” todos estos datos mediante una jerarquía y orden establecido en la normativa MPEG. Toda esta información será recomprimida de nuevo mediante un código reversible VLC.

Al circuito de estimación de movimiento le entran las imágenes de referencia que recibirá el decodificador, ya que éstas también han sido creadas a partir de la suma de la imagen compensada en movimiento con el error de predicción comprimido.

Funcionamiento de un decodificador de MPEG

El proceso de decodificación de MPEG no conlleva demasiada complejidad. Dicha sencillez resulta imprescindible si se contempla que cada receptor deberá tener un decodificador, ya que cuanto más sencillo sea el proceso que se le exija, menos costoso será.

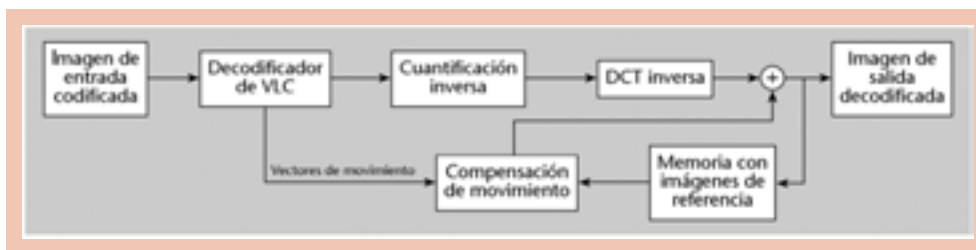
Si se codifica una secuencia de imágenes con un buen codificador, éste realizará una correcta estimación de movimiento creando un error de predicción muy reducido. Si esto sucede, la cantidad de información que hay que codificar es muy poca, por lo que el ancho de banda que hay que transmitir será reducido y el decodificador no deberá hacer frente a elevados flujos de información. Por otro lado, si el codificador es sencillo, realizará predicciones de movimiento incorrectas, lo que generará elevados errores de predicción. Cuando esto suceda, será necesario codificar y transmitir todo este error de predicción para que el decodificador pueda sumarlo a la imagen de referencia después de haberle aplicado la compensación de movimiento. En este caso, el ancho de banda en la salida del codificador será muy elevado.

Por esta razón, cuanto mejor sea el codificador, más sencillo podrá ser el decodificador.

Los pasos que realiza un decodificador son prácticamente el proceso inverso del codificador. Partiendo de una imagen de entrada, se aplica sobre ésta un proceso de descodificación VLC inverso, y se recuperan las matrices de cuantificación, los vectores de desplazamiento y la imagen I o errores de predicción codificados.

La imagen I o errores de predicción son leídos mediante una lectura en zigzag complementaria a la aplicada en la codificación, con lo que se recupera la matriz DCT cuantificada de 8×8 píxeles. A continuación, a partir de las matrices de cuantificación transmitidas, se recupera la matriz DCT original en la etapa de cuantificación inversa. Si sobre esta información se realiza la DCT inversa se obtiene el bloque original, al que se le puede sumar la imagen anterior de referencia compensada en movimiento a partir de los vectores de desplazamiento transmitidos. Con estas operaciones se obtendría el bloque original de 8×8 píxeles.

En el caso de una imagen I, los vectores de desplazamiento tendrán valor (0,0) y el circuito de DCT inverso entregará el bloque original de 8×8 píxeles directamente.



El MPEG define que no se necesite almacenar más de dos imágenes de referencia en el decodificador, más una memoria adicional para la imagen que se esté descodificando en cada momento.

Distintas posibles formas de trabajo de un decodificador

Por lo general existirá una gran cantidad de posibles estructuras visuales, tales como secuencias con un montaje trepidante, secuencias con un montaje lento, *zooms*, *travellings*, panorámicas, planos generales (con elevado contenido visual), primeros planos, etc. Cada una de estas estructuras tendrá un cierto valor de redundancia y entropía, de tal forma que cada una deberá codificarse de forma distinta.



Ejemplo 1

Supongamos que a un decodificador le llega una secuencia de vídeo en la que aparece una imagen fija. El conjunto de imágenes de esta secuencia tendrá mucha redundancia, ya que todas las imágenes son iguales. Este hecho implica que todas las imágenes I tendrán la misma información. Por esta razón, las imágenes P y B estarán formadas en su mayoría

por vectores de movimiento nulos, ya que si, por ejemplo, en un GOP de $M = 3$ y $N = 12$, la primera imagen (I1) y la segunda imagen (B1) son iguales, en la predicción de movimiento se generarán unos vectores de movimiento nulos, por lo que la predicción de movimiento será perfecta (ya que no existe movimiento), con lo que desaparecerá la posibilidad de errores en la predicción. Obtendremos una imagen de calidad con un elevado factor de compresión.



Ejemplo 2

En este caso se realiza un desplazamiento uniforme, de modo que el contenido de la escena se desplaza también en la misma dirección y sentido opuesto al de la cámara. Todos los bloques de la cámara experimentarán el desplazamiento con la misma dirección, sentido y magnitud. Por esta razón, podemos asignar un único vector de desplazamiento para la mayoría de los bloques afectados de la imagen, con lo que nos ahorramos tener que codificar todos de los vectores de cada bloque. De esta forma, los bloques de una imagen P referenciada a partir de una imagen I serán idénticos a los de ésta, pero estarán desplazados todos en la misma magnitud.

A su vez, aquel conjunto de bloques asociados a la nueva información que aparezca en pantalla deberían utilizar una codificación (intra), ya que al tratarse de información nueva no la vamos a poder recuperar a partir de una imagen de referencia anterior. Por

norma general, en un codificador real este conjunto de bloques serán los que corresponderán al error de codificación, ya que rara vez se dará el caso de que una imagen utilice codificación intra e inter simultáneamente.

Ejemplo 3

Supongamos que nos encontramos en una secuencia en la que se produce un cambio de plano. En este punto, prácticamente el contenido visual de la narración va a ser distinto por completo, lo que conlleva que si el cambio de plano se produce, por ejemplo, sobre una imagen B, ésta sólo debería ser predicha a partir de una imagen I o P posterior a ella.

En el caso de que la predicción se realice a partir de una imagen de referencia I o P anterior a la B, el error de predicción será muy grande y los vectores de desplazamiento sólo se aplicarán sobre aquellos bloques en los que el contenido visual sea igual en las dos imágenes.

Etapa 2: MPEG-1 (ISO/IEC 11172)

Introducción

El *Moving Picture Experts Group* (MPEG) definió en 1988 el estándar ISO/IEC 11172. Con este estándar se pretendía lograr un procedimiento de codificación capaz de ubicar imágenes en movimiento en soportes de tipo CD-ROM, con un flujo máximo de datos de 1,5 Mb/s. Este ancho de banda es suficiente para trabajar con aplicaciones de audio digital sin compresión, pero el problema surge cuando se plantea un formato diseñado para aplicaciones de audio como formato de almacenamiento de vídeo. Lógicamente, esto supone trabajar con elevados factores de compresión de la imagen, de modo que no se puede considerar un estándar adecuado para aplicaciones de difusión.

Entonces, se puede deducir que MPEG-1 es un estándar diseñado básicamente para aplicaciones multimedia que no requieran elevados niveles de calidad de imagen, como por ejemplo ver un vídeo a tamaño reducido en la pantalla del ordenador.

La norma MPEG-1 se compone de tres partes diferenciadas:

- MPEG system (ISO/IEC 11172-1), que define la estructura del múltiplex MPEG-1.
- MPEG video (ISO/IEC 11172-2), que define la codificación de vídeo MPEG-1.
- MPEG audio (ISO/IEC 11172-3), que define la codificación de audio MPEG-1.

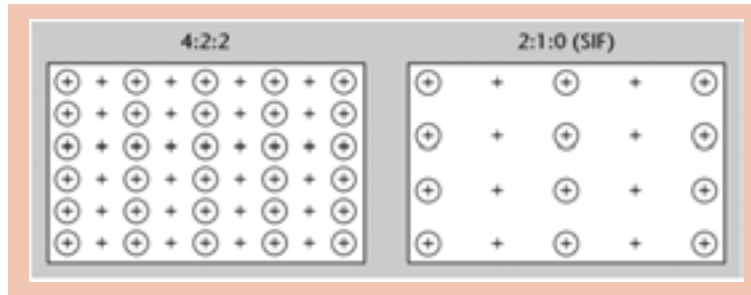
MPEG-1 video (ISO/IEC 11172-2)

Este sistema utiliza codificación intercuadro, lo que supone analizar la secuencia en el dominio temporal y realizar un proceso de predicción y compensación de movimiento, con lo que se generan imágenes I, P y B.

Si se parte de la recomendación ITU-601, en la que se definen los parámetros para la digitalización de una señal de vídeo con calidad profesional “contribución”, se puede observar que el ancho de banda trabajando a 8 bits es de 216 Mb/s, lo que supone un volumen de información muy elevado para muchas aplicaciones multimedia. Por esta razón se utilizará esta recomendación como punto de partida para la compresión. Ya se ha comentado que el ancho de banda del MPEG-1 no debe superar los 1,5 Mb/s, ya que para convertir imágenes de 216 Mb/s a 1,5 Mb/s es necesario utilizar técnicas de compresión con pérdidas (*lossy*) eliminando tanto redundancia como entropía.

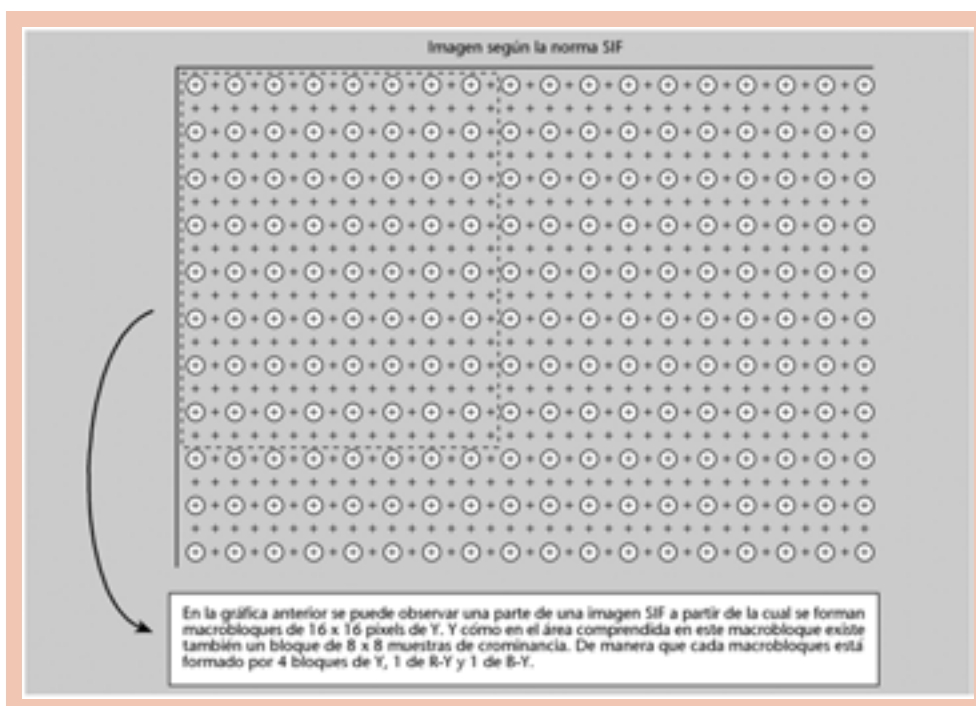
La entropía de entrada se reduce submuestreando en las tres dimensiones (horizontal, vertical y temporal). Si la señal de la entrada es del tipo 4:2:2, se empieza por descartar uno de cada dos campos, que a efectos prácticos se traduce en eliminar todas las líneas pares o impares.

Seguidamente, la señal 4:2:2 se convierte en otra 2:1:0, con lo que se consigue redimensionar la imagen a 360×288 para la señal de luminancia Y, y 180×144 para la crominancia. Con este redimensionamiento se consigue una relación de compresión de unos 3:16. Esto genera lo que se conoce como señal o norma SIF (*Source Input Format*).



En realidad, en MPEG-1 se utiliza una variante del formato SIF de 352×288 píxeles en el sistema europeo y de 352×240 píxeles en el sistema americano. La razón de esta modificación responde a la necesidad de adaptar el formato de imagen al hecho de que MPEG-1 aplica los vectores de movimiento en regiones formadas por macrobloques de 16×16 píxeles. Esta adaptación nos permite obtener un número exacto de bloques por imagen ($352/16 = 22$ bloques horizontales, $288/16 = 18$ bloques verticales).

Después de esta compresión de 5:1 se forman los macrobloques de 16×16 píxeles. Este proceso genera seis bloques de 8×8 por cada macrobloque, cuatro bloques de Y, un bloque de R-Y y otro bloque de B-Y.



Si se suma la compresión que se ha aplicado al pasar de 4:2:2 a SIF con una compresión temporal mediante el cálculo de vectores de desplazamiento, entonces se conseguirán los niveles de compresión que permitirán comprimir datos de 216 Mb/s a los 1,5 Mb/s que define como flujo máximo la norma MPEG-1.

Así pues, éste es un sistema que utiliza compresión intercuadro con predicción y compensación de movimiento. Al tener que reducir la redundancia al nivel de la entropía, se trata de un sistema de compresión irreversible (*lossy*).

MPEG audio (ISO/IEC 11172-3)

Dentro del estándar de codificación de audio para MPEG-1 (ISO/IEC 11172-3), existen tres niveles distintos de calidad. A medida que evolucionan las técnicas de compresión y sus algoritmos asociados se van estandarizando nuevos niveles de compresión. Por esta razón, el nivel o capa 1 (*layer 1*) utiliza técnicas de compresión más sencillas que el nivel o capa 3 (*layer 3*), comúnmente conocido como MP3.

El MPEG-1 audio sólo contempla la codificación de audio monofónico y estereofónico (en ningún caso se puede trabajar con audio multicanal).

La primera capa, o sea, MPEG-1 Layer-1, es una versión simplificada del MUSICAM, que resulta adecuado para aplicaciones de reducida compresión y bajo coste de codificadores. Se trata de un formato muy parecido al PASC.

La segunda capa, es decir, MPEG-1 Layer-2 es idéntico al MUSICAM y muy probablemente sea el estándar que se usará en el DAB (*Digital Audio Broadcasting*).

La tercera capa, el MPEG-1 Layer-3 o MP3, es una combinación de las mejores características del ASPEC y el MUSICAM; básicamente se aplica en comunicaciones de banda estrecha (Internet).

Etapa 3: MPEG-2 (ISO/IEC 13818)

Introducción

MPEG-2 es una extensión ampliada del formato MPEG-1. Por esta razón, con este estándar se consiguen factores más elevados de compresión sin tener que renunciar a la calidad de imagen. Así, la principal diferencia entre un codificador/decodificador de MPEG-1 y un codificador/decodificador de MPEG-2 se traduce en la capacidad de cómputo de cada uno de ellos y, a su vez, el precio.

Al igual que el MPEG-1, el estándar MPEG-2 define tres subnormas:

- MPEG-2 system ISO/IEC 13818-1, que define la estructura del múltiplex MPEG-2.
- MPEG-2 video ISO/IEC 13818-2, que define la codificación de vídeo MPEG-2.
- MPEG-2 audio ISO/IEC 13818-3, que define la codificación de audio MPEG-2.

MPEG-2 video (ISO/IEC 13818-2)

Con el MPEG-2 video se consigue una gran variedad de calidades de imagen en función del bit/*rate* deseado. Se contemplan desde aplicaciones con las tasas del MPEG-1 de 1,5 Mb/s (calidad de vídeo doméstico) hasta calidades correspondientes a la HDTV (televisión de alta definición) con tasas de 100 Mbits/s. Por esta razón, este formato es el que se adoptará en el momento de transmitir señales de televisión digital.

Al encontrarse todas las variantes ubicadas dentro del estándar MPEG, no existe ningún problema de falta de compatibilidad hacia abajo, entre MPEG-2 y MPEG-1. Esto significa que cualquier información codificada en MPEG-1 puede ser interpretada sin ningún problema por un descodificador de MPEG-2.

MPEG-2 introduce un nuevo concepto (ya utilizado en imágenes fijas codificadas en JPEG): la **escalabilidad**. La escalabilidad consiste en permitir decodificar la imagen de vídeo en distintos grados de calidad, en función de bit/*rate* deseado. Los datos de vídeo se encuentran codificados en una serie de trenes binarios que formarán distintas capas. La primera capa (por ejemplo, donde sólo se encuentran los coeficientes de DC de la DCT) se llama capa base, y es totalmente independiente del resto. Cuando se desee, se pueden descodificar las otras capas (o sea, el resto de los coeficientes de la DCT) para obtener una mayor resolución de imagen. También existe la posibilidad de trabajar con escalabilidad temporal eliminando campos o imágenes de forma periódica, que se podrían recuperar cuando se desee.



La escalabilidad es muy útil en aplicaciones en las que se desee transmitir la información de la capa base por un canal de transmisión más seguro y/o estable y el resto de las capas menos estables o inmunes a posibles errores.

Este tipo de escalabilidad es muy habitual en Internet. Por ejemplo, en el momento en que se accede a una web con imágenes o fotografías es posible ver la imagen a baja resolución y pantalla pequeña; en este caso se está trabajando a partir de la capa base y cuando se desea se puede ampliar la imagen a tamaño real con la resolución completa. En este caso, se están descodificando la totalidad de los coeficientes de la DCT.

El estándar MPEG-2 acepta señales de vídeo digitales en formato progresivo (alguna versión del NTSC) como con entrelazado de campos (PAL y SECAM).

La norma MPEG-2 contempla la posibilidad de generar *bit/rates* fijos y variables, lo que nos permite adaptar la información codificada a las características del formato de transmisión o almacenamiento que se desee. Habitualmente se encontrarán flujos de bits constantes en la salida de un codificador de MPEG cuando la señal se transmita por antena, cable o vía satélite, y flujos de bits variables en los procesos de almacenamiento como el DVD.

Por ejemplo, en el caso de la DVB-T (*Digital Video Broadcasting Terrestre*), es decir, la emisión de señales de televisión digitales utilizando las antenas de televisión convencionales, se requiere un ancho de banda fijo, debido a que el canal de transmisión genera ruido con mucha facilidad. Por esta razón, al transmitir señales digitales en canales ruidosos es indispensable la protección contra errores y la creación de un flujo de datos constante. Para realizar este proceso es indispensable ESTA FRASE ACABA CORTADA, FALTA LA CONTINUACIÓN!!!

En el proceso de grabación sobre un formato como el DVD es muy difícil que se genere ruido y, a su vez, es imprescindible optimizar la capacidad de los datos. En este caso será útil trabajar con un flujo variable de datos en la salida del codificador de MPEG.

Perfiles y niveles en MPEG (profiles & levels)

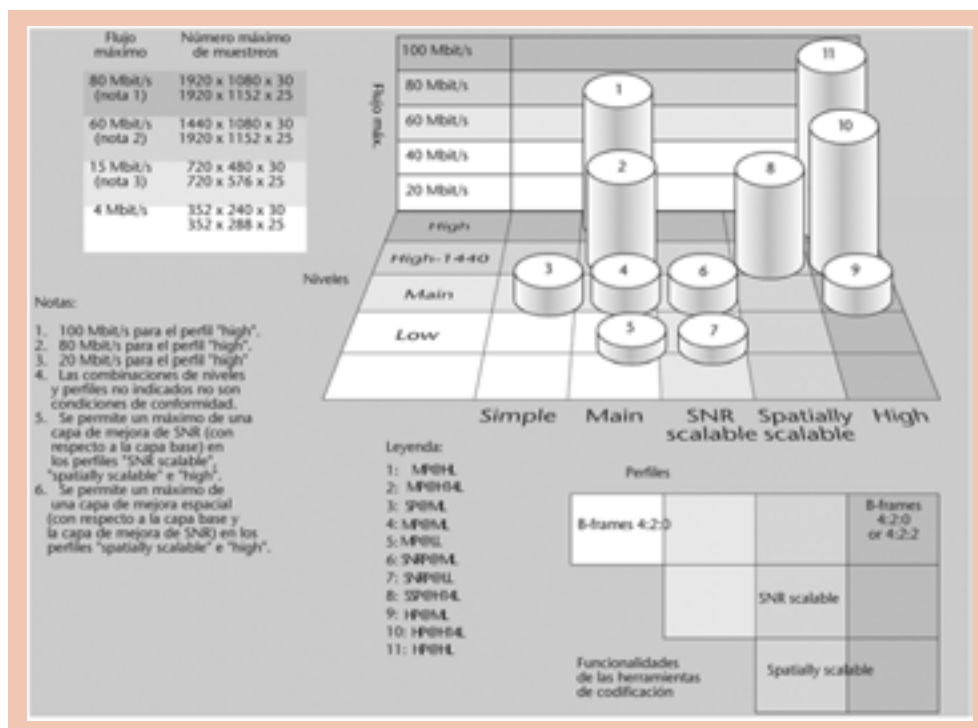
Una de las principales características de este estándar es que cuando se definió se acordó como condición que debía ser una norma abierta y que contemplara diferentes niveles de calidad para cubrir todas las posibles necesidades del mercado audiovisual.

Por esta razón se buscó una forma de clasificación de los distintos niveles de trabajo a partir de dos parámetros: los perfiles y los niveles.

Perfiles (*profiles*): en el perfil se especifican los parámetros con los que se va a comprimir la información. En este parámetro se definirá cuál será la norma de la imagen con la que se trabajará (4:2:2, 4:2:0,...) y el tipo de imágenes que se utilizarán en la predicción temporal, I, I/P, o I/P/B.

Niveles (*levels*): los niveles nos indican la resolución vertical y horizontal de la imagen.

La combinación de perfiles y niveles nos dará como resultado el ancho de banda de la información codificada.



El significado de los perfiles es el siguiente:

- El perfil *simple* está destinado a simplificar el codificador y el descodificador para que sean económicos, en contra de la tasa de compresión que no utiliza predicción con imágenes B (bidireccionales). Observad que con este perfil sólo se ha es-

tandarizado un único nivel formando la combinación *Nivel Main* y *Perfil Simple* (*Simple Profile at Main Level*, se escribe habitualmente SP@ML).

- El perfil *main* corresponde actualmente al mejor compromiso calidad/tasa de compresión. Utilizando los tres tipos de imágenes (I, P y B), en este caso se requiere un codificador y un decodificador más complejos.

Los perfiles escalables están destinados a aplicaciones futuras en las que se pueda desear transmitir señales en distintas calidades simultáneamente; por ejemplo, supongamos que en el futuro hay usuarios con complejos descodificadores que les permitan la descodificación de imágenes de alta definición y que al mismo tiempo hay usuarios que con un descodificador más sencillo sólo pueden ver imágenes en calidad estándar. En este caso, se transmitirá una capa base con los principales coeficientes de la DCT para los usuarios con un descodificador sencillo, mientras que a aquellos que tengan un descodificador más complejo se les enviará otra señal con más información de imagen para que puedan obtener imágenes de alta definición.

Estos perfiles son también de gran utilidad en transmisiones con posibilidad de tener mucho ruido. En este caso, la capa base se protege más que las capas de realce, de esta forma se asegura una calidad mínima.

Existen dos tipos de escalabilidad, la escalabilidad espacial (*spatial scalable*) y la SNR *scalable*. La escalabilidad espacial consiste en enviar unos coeficientes de la DCT como capa base y el resto como capa de realce. La SNR *scalable* afecta a la cantidad de bits utilizados en la cuantificación de los coeficientes de la DCT. En este caso se transmiten los coeficientes a n bits y los de realce, que darán mayor calidad de imagen, se transmiten aparte, con lo que se mejora la relación entre la señal y el ruido.

El perfil *High* está destinado a aplicaciones de teledifusión de alta calidad HDTV; se pueden partir de señales según las normas 4:2:0 y 4:2:2.

Existe compatibilidad descendente entre perfiles de manera que un descodificador de un perfil determinado podrá descodificar todos los perfiles inferiores.

El significado de los niveles es el siguiente:

- *Low* (bajo): corresponde a la resolución SIF utilizada en MPEG-1. La tasa máxima de transmisión de datos es de 4Mbits/s.
- *Main* (principal): corresponde a la resolución del tipo ITU-601, o sea, 720×576 a partir de una norma 4:2:0. En este caso se aceptan flujos de hasta 15 Mbits/s.
- *High 1440* (alto 1.440): está destinado a aplicaciones de HDTV hasta una resolución de 1.440×1.152 . La tasa máxima de bits es de 80 Mbits/s.
- *High* (alto): es un formato optimizado para aplicaciones de HDTV con una resolución de imagen de 1.920×1.152 . Pudiendo trabajar con flujos de datos de hasta 100 Mbits/s con el perfil *High*.

La combinación MP@ML (*Main Profile @ Main Level*) es la que se ha escogido como norma para la DVB (*Digital Video Broadcasting*), debido a sus óptimas prestaciones, que permiten la obtención de imágenes entrelazadas a partir de la norma 4:2:0 y retornan flujos que habitualmente oscilarán entre los 4 y 9 Mbits/s en función de la calidad y redundancia de la secuencia.

Aquellas combinaciones entre perfiles y niveles que no aparecen no se han definido como válidas.

Problemas del MP@ML en aplicaciones profesionales

El ancho de banda del MPEG es muy variable; básicamente depende de los perfiles y niveles utilizados. Además, utilizando predicción temporal es imposible conseguir precisión de fotograma en la edición.

Los puntos más flojos del MPEG son los siguientes:

- El MPEG se optimizó para aplicaciones difusión punto-multipunto, o sea, con una sola codificación se transmite la información a muchos usuarios. No se consideró la multigeneración.
- No se tuvieron en cuenta los procesos de *croma key*.
- Sólo se codifican las líneas activas, no se ha contemplado la codificación de las líneas de borrado de pantalla donde en muchos casos se inserta el VITC.
- El límite superior de la frecuencia binaria del MP@ML es de 15 Mbits/s. Las evaluaciones del grupo MPEG han demostrado que una imagen de calidad subjetiva similar al CCIR-601 se puede conseguir con frecuencias binarias de unos 9 Mbits/s en aplicaciones de primera generación, pero esto sólo se consigue con un GOP de doce imágenes, lo que significaría una precisión en la edición de doce *frames* válida para difusión pero no para su aplicación en estudio. Estas evaluaciones han demostrado también que después de la tercera generación la calidad que se consigue no es la adecuada para aplicaciones de estudio.

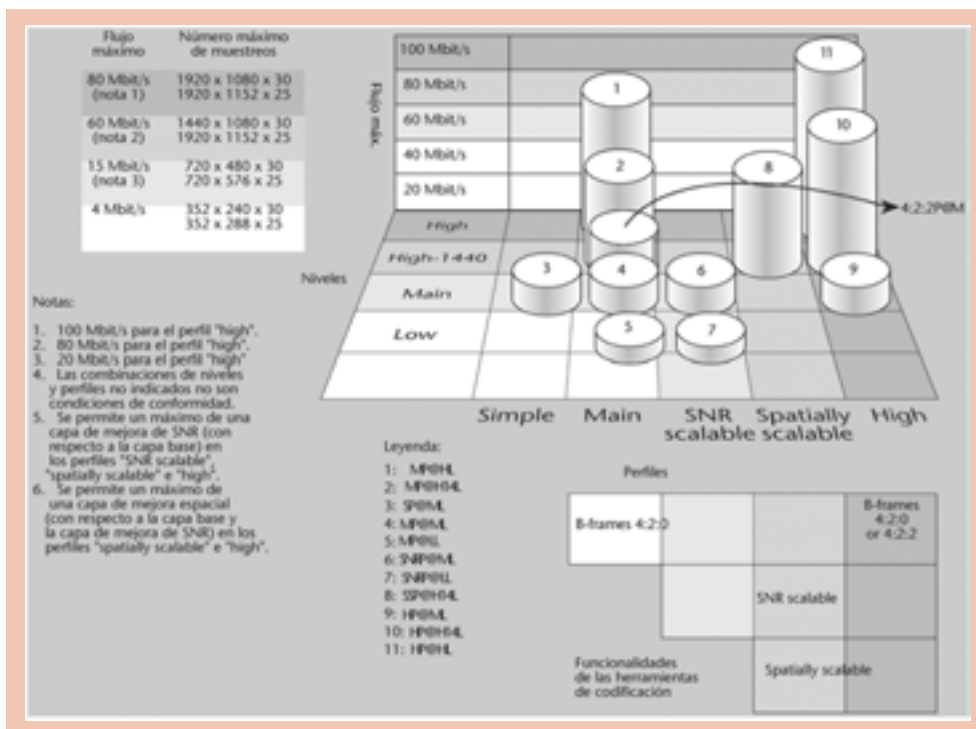
MPEG-2 4:2:2 MP@ML

Contemplando todas las deficiencias del ML@ML, algunos fabricantes han propuesto una nueva subnorma del MPEG-2 para aplicaciones de estudio. Ésta se denomina 4:2:2 Profile@ Main Level, y permite trabajar a partir de un perfil *main* y un nivel *main*, pero con algunas modificaciones. En este caso, la señal que se codifica utiliza como referencia la norma 4:2:2, con lo que se obtienen flujos de datos de hasta 50

Mbits/s; en este caso se requiere un decodificador con un *buffer* de entrada de 32 Mb de memoria RAM.

Trabajando con esta variante se obtienen una serie de ventajas:

- Si se mantiene la norma original 4:2:2 se mejora ligeramente la resolución horizontal y vertical de la crominancia.
- En las aplicaciones profesionales, el código de tiempo vertical (VITC) va grabado en las líneas de borrado de pantalla, Por esta razón, cuando se codifica la señal esta información desaparece (debido a que las líneas de borrado de pantalla no se digitalizan), por lo que la extensión 4:2:2P@ML aumenta el número de líneas codificadas, que pasan de 720×576 a 720×608 en los sistemas de 625/25 y de 720×480 a 720×512 en los sistemas de 525/30. De esta forma, se puede incorporar la información de código de tiempo vertical (VITC) en las 32 nuevas líneas.



El aumento de líneas y la nueva norma obligan al flujo de datos a pasar de 15 Mbits/s a 50 Mbits/s, con lo que se rompe con la estructura típica de GOP de doce imágenes. De esta forma se conseguirá también una mayor precisión en los procesos de edición en vídeo.

Contemplando el elevado ancho de banda con el que se está trabajando, se considera adecuado no comprimir el audio multiplexado con la señal de vídeo, ya que éste no supone ni el 10% del flujo total de datos.

MPEG-2 audio (ISO/IEC 13818-3)

El audio multicanal es el campo de la acústica y la tecnología orientado a la captación, producción y difusión de diferentes fuentes sonoras para la creación de espacios sonoros envolventes.

Las principales aplicaciones de esta tecnología se encuentran orientadas al campo de los audiovisuales, en el que se pretende potenciar la sensación de realismo sonoro ayudando a que el espectador se introduzca dentro de la narración y utilizando en la proyección una serie de altavoces cuidadosamente distribuidos en la sala de forma que el espectador reciba sonido desde la parte frontal, lateral y posterior.

Este tipo de tecnología se comenzó a estudiar hacia los años setenta. En esta época tuvo muy poco éxito a causa de las dificultades técnicas que comportaba y de la poca aceptación por parte del público, que todavía estaba asimilando la introducción del cine en color. Desde entonces su evolución fue lenta, hasta la mitad de los años noventa. Actualmente es una tecnología que se encuentra en plena expansión, en pocos años han aparecido gran variedad de formatos de tratamiento y difusión.

Originariamente, esta tecnología iba orientada a ser aplicada en salas cinematográficas como es el caso del Dolby Estereo. Hacia los años ochenta se comenzaron a comercializar los primeros aparatos para aplicaciones domésticas de esta tecnología: *Dolby Surround* y *Dolby Surround pro-logic*. Desde entonces hasta la actualidad, diferentes empresas y organizaciones han creado sus propios formatos de audio multicanal, tanto para ser difundidos en cine como en el ámbito particular, como es el caso de Sony con el SDDS, el DTS, George Lucas con el THX, Dolby con el Dolby Digital o la Motion Pictures Expert Group con el Musicam Surround o MPEG2.

En los años setenta, Dolby introdujo un nuevo formato óptico denominado Dolby Estereo. Este nuevo formato conseguía codificar cuatro canales de audio con los dos estéreo. Actualmente este formato se encuentra totalmente implementado en casi todas las salas del mundo y es una de las ventajas que ha ofrecido el cine en su combate con la televisión.

Esta revolución en el sonido del cine avanzaba en paralelo con otra revolución: la del vídeo doméstico. Al principio sólo se utilizaban para registrar determinados programas para visionarlos posteriormente, pero al cabo de un tiempo la industria cinematográfica reparó en que conseguía recaudar grandes sumas de dinero gracias a este formato. Por este motivo, se redoblaron los esfuerzos e incluso se llegaron a hacer películas pensadas exclusivamente para el usuario del vídeo doméstico. El público estaba acostumbrado a escuchar audio de alta fidelidad en la cadena de música y Dolby Estereo en el cine; fue entonces cuando empezaron a reclamarlo en televisión y en vídeo.

En 1982 se empezó a comercializar el Dolby Surround para vídeo y televisión, que representa las mismas características que en el cine el Dolby Estereo.

Con este sistema se utilizan tres altavoces frontales y dos posteriores, todos alimentados por un amplificador con decodificador incorporado.



Actualmente, la mayoría de las salas cinematográficas y algunos particulares tienen salas de escucha de audio multicanal sobre plataformas digitales, como el Dolby Digital (AC-3) o el DTS, en las que se puede escuchar simultáneamente desde uno hasta siete canales de audio.