

MPEG-2: UN SISTEMA DE COMPRESIÓN DE VÍDEO PARA LA TELEVISIÓN DIGITAL

Pedro Cuenca Castillo
Teresa Olivares Montes

Pedro Cuenca Castillo y Teresa Olivares Montes están en el Departamento de Informática – Albacete. Universidad de Castilla la Mancha.

RESUMEN

Hoy en día no puede hablarse de la televisión digital sin referirse de algún modo al acrónimo MPEG, iniciales de *Moving Pictures Expert Group* (Grupo de Expertos de Imágenes en Movimiento). Durante muchos años las aplicaciones de la televisión digital se han limitado a los equipos de estudio. La transmisión de las señales digitales que implican flujos de velocidades binarias muy elevadas (216 Mbit/s para señales de acuerdo con la recomendación 601 del CCIR) eran impensables por la anchura de banda requerida. Ahora sin embargo, lo que se consideraba el mayor obstáculo para la transmisión de la televisión digital, gracias a las modernas técnicas de compresión de vídeo, ha pasado a ser una de sus ventajas más importantes. Utilizando estas técnicas de compresión, o reducción de la velocidad del flujo binario, se puede transmitir no uno sino varios programas en un canal de televisión.

1. INTRODUCCIÓN

LA transmisión digital es de importancia primordial para las telecomunicaciones, particularmente en la industria telefónica, pero hay mucho más en el negocio del vídeo digital aparte de la teleconferencia y la videotelefonía. La industria informática, de telecomunicaciones y de electrónica de consumo comparten cada vez con mayor frecuencia la misma tecnología, y esta tecnología incluye compresión digital de vídeo. En vista de la tecnología compartida por la industria de procesamiento de información, el ISO (Organización Internacional de Estandarización) se hizo cargo del esfuerzo de desarrollar un estándar para vídeo y su audio asociado adecuado para su transmisión y almacenamiento en medios digitales. Este esfuerzo fue conocido por el nombre del grupo que lo empezó, MPEG. Las actividades del grupo MPEG cubrieron más que la compresión de vídeo, ya que la compresión de audio aso-

ciado y el problema de la sincronización audiovisual no se puede tratar independientemente de la compresión del vídeo. El sistema de compresión de vídeo MPEG tiene como objetivo fundamental representar una fuente de vídeo con un número mínimo de bits preservando al mismo tiempo el nivel de calidad requerido para una aplicación determinada.

El Instituto Europeo de Normalización (ETSI) ha establecido un norma de compresión de 34 Mbit/s, inspirada fundamentalmente en el proyecto italo-español Eureka 256 que se utiliza para circuitos de contribución, es decir, en transmisiones en las que la señal de televisión tiene que tener suficiente calidad para ser procesada sin deterioro. La velocidad binaria de 34 Mbit/s supone una relación de compresión de 6,35 (216/34). Sin embargo, para aplicaciones de calidad más baja, es decir, con relaciones de compresión más elevadas (velocidades binarias de unos 3 a 15 Mbit/s) han proliferado sistemas de compresión no normalizados que han sido rápidamente adoptados para la distribución por satélite de programas de televisión. Los sistemas de cable han podido satisfacer así la gran demanda de circuitos aprovechando en la mayor medida posible la anchura de banda de los transportadores, a veces, con notable detrimento de la calidad.

En el mercado hay cerca de 20 distintos sistemas patentados de compresión que pueden ser útiles para aplicaciones profesionales de distribución de programas de televisión. No obstante, sin una normalización a nivel internacional, no sería posible la fabricación de chips para la decodificación de señales comprimidas a precios asequibles para el gran público que pronto tendrá acceso a las emisiones de televisión digital.

Afortunadamente, el conjunto de normas MPEG de procesado de las señales de televisión digital, incluyendo métodos de compresión de vídeo, se están imponiendo en todo el mundo. La familia de sistemas MPEG y concretamente el MPEG-2 tiene un gran interés para aplicaciones de radio difusión y ha servido de base para las especificaciones del sistema americano Gran Alianza y para las del sistema europeo de transmisión por satélite. Sobre la base del MPEG-2, que es ya una norma internacional, se podrán ofrecer en el mercado decodificadores por unos miles de pesetas que abrirán al gran público todas las posibilidades de la televisión digital.

No quiere esto decir que todos los sistemas de compresión que se ofrezcan en el mercado tendrán las mismas prestaciones. El chip MPEG puede compararse al procesador 486 o al Pentium que constituyen el elemento fundamental de los modernos ordenadores. Aunque los que disponen del mismo microprocesador pueden considerarse de la misma familia, los distintos elementos (configuraciones de memoria, métodos de procesado y software) que se incorporan a cada uno de los tipos de ordenador que existen en el mercado marcan la diferencia en las prestaciones. Lo mismo ocurrirá con los decodificadores equipa-

dos con el chip MPEG-2. Habrá que agregar a estos chips dispositivos de memoria y circuitos integrados para aplicaciones específicas que también pueden producirse en grandes series. Los equipos dispondrán también de circuitos lógicos programados en fábrica para configurar el decodificador para su aplicación específica. Finalmente, se utilizará software, que puede modificarse con facilidad y rapidez, para que el codificador se adapte a las necesidades particulares.

2. EL GRUPO MPEG Y LA NORMA MPEG-2

MPEG es un grupo internacional creado bajo los auspicios de la organización Internacional (ISO), con participación de los sectores de telecomunicación, informática y radiodifusión. Al comienzo el objetivo de este grupo era proporcionar un método genérico de codificación de imágenes en movimiento y del audio asociado para dispositivos de almacenamiento digital a velocidades binarias de 10 Mbit/s. El desarrollo de la norma MPEG-1, el primer trabajo del grupo, iniciado en 1988 y ya finalizado, se concentró en la compresión de vídeo con velocidades binarias bajas adecuadas para dispositivos de almacenamiento tales como CD-ROM y transmisión en 1,554 y 2,048 Mbit/s. En 1990 comenzó el estudio de la norma MPEG-2 cuyo objetivo era definir un método de codificación de las señales audiovisuales para proporcionar calidad de radiodifusión con velocidades binarias con un límite de 15 Mbit/s sobre la base de la norma de televisión digital definida en la Recomendación 601 del CCIR. Posteriormente se ha extendido el límite de velocidades binarias para tener en cuenta la televisión de alta definición.

Conviene señalar que las llamadas normas MPEG no se pueden considerar, ni son en realidad, normas concretas de implementación de equipos sino más bien descripciones genéricas del modo en que el conjunto de señales comprimidas de vídeo, audio y datos tienen que ser combinadas o multiplexadas en un flujo de paquetes digitales para su transmisión. Es, pues, muy posible que los fabricantes puedan adoptar distintos codificadores MPEG que darán lugar a diferentes calidades de imagen. El MPEG-2 se caracteriza por la concentración de la complejidad, y por tanto del costo, en el codificador —extremo emisor— simplificando al máximo el decodificador cuyo grado de complejidad, y costo, dependerá de las aplicaciones.

Las características de la norma MPEG-2 a tener en cuenta son:

- a) Las imágenes pueden ser con exploración entrelazada o progresiva.
- b) Las relaciones de frecuencias de muestreo pueden ser: 4:2:0, 4:2:2 y 4:4:4

- c) Distintos valores de resolución de la imagen (teóricamente hasta 16.000×16.000 píxeles) incluyendo todas las frecuencias de campo y trama utilizadas en televisión.
- d) Codificación jerarquizada, lo que permite que un decodificador para definición normal pueda extraer la información que necesita de un flujo de datos de nivel más bajo (por ejemplo de alta definición). Con una sola transmisión puede emplearse decodificadores para diferentes definiciones.

La norma MPEG-2, conocida también como ISO/IEC CD 13818, se compone de tres partes. La primera se refiere a la combinación de uno o más flujos elementales de vídeo, audio y otros datos en flujos simples o múltiples adecuados para almacenamiento o transmisión. Las otras dos partes de la norma describen los métodos de codificación para reducir la velocidad binaria del vídeo y del audio, proceso también conocido como **compresión**.

2.1. Fundamentos de la compresión

Todos los sistemas de televisión contienen la información redundante del sincronismo que en gran parte puede eliminarse antes de la transmisión reduciendo así la velocidad binaria sin pérdida de la calidad de la imagen. Sin embargo para obtener un mayor grado de compresión hay que recurrir a técnicas que de algún modo afectan a la calidad de la imagen. El objetivo fundamental de estas técnicas consiste en representar una fuente de vídeo con un número mínimo de bits preservando al mismo tiempo el nivel de calidad requerido para una aplicación determinada. La compresión de vídeo se realiza en tres operaciones básicas como ilustra la *figura 1*. En primer lugar, se somete la señal de vídeo a un proceso de análisis que da lugar a una forma eficaz de representación que facilita la compresión. Para describir la señal, la representación puede contener más información que la señal original, pero en todo caso la mayor parte de la información importante estará concentrada en solo una pequeña parte de esta descripción. Si la representación es adecuada, únicamente se necesita transmitir esta pequeña fracción de los datos para reconstruir fielmente en el receptor la señal de vídeo.

La segunda operación es la cuantificación, o sea, la conversión a valores discretos de los datos de la representación. La tercera y última operación consiste en la asignación de palabras de código, series de bits utilizados para representar los niveles de cuantificación. En cada una de estas operaciones se intenta aprovechar la redundancia presente en la fuente de vídeo y las limitaciones de la visión humana.

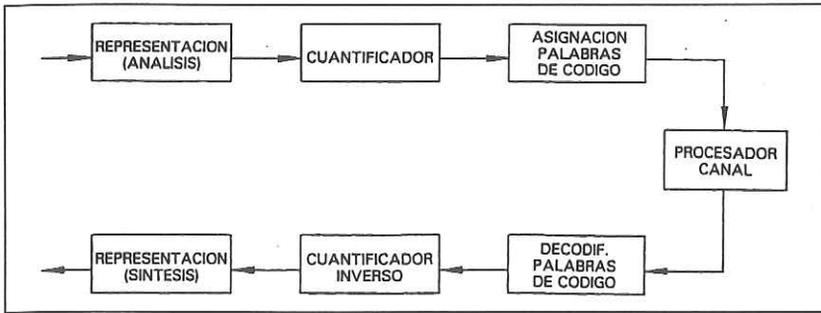


FIGURA 1.
Operaciones básicas de la compresión.

2.2. Compensación de movimiento

Una secuencia de vídeo es una serie de imágenes fijas que se suceden rápidamente para dar la sensación de movimiento continuo. Aunque cada una de estas imágenes es distinta de las adyacentes, la mayor parte de la información que contienen es muy similar. Esto quiere decir que hay una gran redundancia temporal entre las imágenes adyacentes. La compensación de movimiento es un método con el que se pretende eliminar esta redundancia temporal. La diferencia entre una imagen, o trama, y la siguiente se debe fundamentalmente al movimiento de objetos. Con el método de compensación de movimiento, se crea una trama de predicción con la trama actual, o nueva, y la trama precedente (ya codificada y en memoria) mediante el cálculo del movimiento entre las dos tramas y una compensación para el movimiento. La diferencia entre la trama actual y la predicción que se ha hecho para esta trama actual es lo que se llama *residuo de movimiento compensado*. En una secuencia típica de vídeo, este residuo, codificado, tiene una energía mucho menor que la del vídeo original debido a la eliminación de la redundancia temporal. Para el cálculo del movimiento se supone que en las tramas de vídeo consecutivas aparecen los mismos componentes de la escena aunque posiblemente en diferentes posiciones. El movimiento puede ser global o local dentro de la trama y para optimizar el procedimiento se realiza el cálculo del movimiento en cada región local de una trama. El modelo más normal para el movimiento local es la traslación, que es mas bien restrictivo ya que no puede representar otros posibles movimientos como el de rotación o de cambio de escala. Sin embargo, suponiendo que estos movimientos se verifican sólo localmente y si se identifican y procesan esas regiones separadamente cuando falla el modelo, se pueden conseguir excelentes resultados. Un procedimiento muy popular para la estimación del movimiento y que ha sido adoptado por el MPEG-2, es la comparación de

macrobloques. De acuerdo con el MPEG-2, un bloque esta constituido por un conjunto de 8×8 pixels o por valores correspondientes a estos pixeles. Un bloque puede representar la información de luminancia o de crominancia. Un macrobloque consiste en cuatro bloques de información de luminancia (o 16 pixeles \times 16 líneas) y un número variable de bloques de crominancia (señales de diferencia de color C_R y C_B). Cuando el número de bloques de crominancia es dos (un bloque C_R y otro C_B) el formato sería 4:2:0 y si el número es cuatro (dos bloques C_R y otros dos C_B), 4:2:2. Por último, si el numero de bloques de crominancia fuera ocho (cuatro bloques C_R y otros cuatro C_B) el formato sería 4:4:4. Se pueden agrupar macrobloques contiguos para formar franjas (slices). El orden de los macrobloques dentro de una franja es el mismo que en la exploración normal de televisión, de izquierda a derecha y de arriba abajo. La unidad primaria de codificación de una secuencia de vídeo es la trama individual o imagen. Una imagen de vídeo consiste en el grupo de franjas que constituyen el área activa de la imagen. Combinando una o más imágenes (tramas) en secuencia se forma un grupo de imágenes que permite el acceso aleatorio a la secuencia y proporciona márgenes para la codificación entre imágenes como se explicará mas adelante. En el flujo de datos, una secuencia de vídeo se representa por un encabezamiento de secuencia, uno o mas grupos de imágenes, y un código final de secuencia.

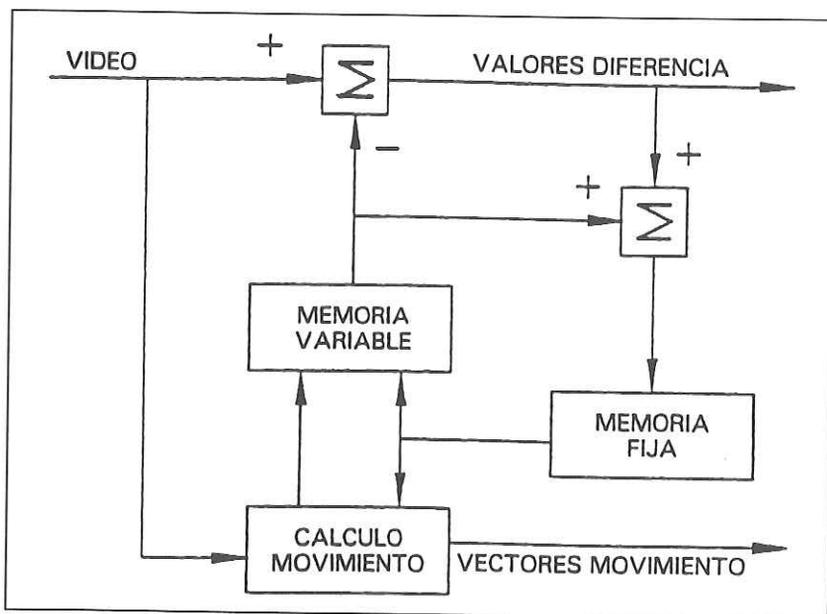


FIGURA 2.
Codificación predictiva inter-trama.

En el diagrama de bloques de la *figura 2* se representan los elementos esenciales que se necesitan para la codificación predictiva. La trama de vídeo de entrada, imagen actual o nueva, se compara con una imagen transmitida previamente que se ha almacenado en memoria. Utilizando la memoria variable, la comparación se hace mediante los macrobloques de la imagen previa que se examinan para determinar si en la nueva imagen hay alguno que presente la máxima semejanza. Si ocurre así, se genera un *vector de movimiento* que describe la dirección y la distancia a la que el bloque se ha movido. Esta comparación de bloques, a pesar de su aparente complicación, puede hacerse mediante circuitos VLSI que no ofrecen dificultades de implementación. Combinando todos los macrobloques en los que se ha encontrado una mayor igualdad se forma una imagen predecible que comparándola con la nueva imagen pixel por pixel proporciona una imagen diferencia.

Para tramas de vídeo típicas, la predicción mediante compensación de movimiento es muy satisfactoria. Sin embargo, en algunas tramas como las de cambio de escena, la predicción no da buenos resultados. Incluso en una trama determinada, algunas regiones pueden predecirse bien con compensación de movimiento, pero no así otras en las que no resulta adecuado el modelo de movimiento de traslación. El modelo de compensación de movimiento anteriormente descrito se conoce como *codificación inter-trama*. Para solventar los problemas en los casos en los que no es conveniente aplicar codificación predictiva, hay que recurrir a la codificación de una trama sin tener en cuenta otras. A este proceso se le conoce como *codificación intra-trama*. Se utiliza esta codificación para inicializar el receptor y acceder a un canal (cuando se enciende el receptor o cuando se cambia de canal). Con predicción de movimiento compensado es necesario que el decodificador disponga de una trama inicial de referencia para arrancar el proceso de predicción.

En base a lo dicho anteriormente, la norma MPEG define tres tipos de tramas o imágenes: *intra-trama (I)*, *de predicción (P)* y *bidireccional (B)*. El primer tipo, imágenes **I**, corresponde a la codificación intra-trama, en la que se utiliza únicamente la información contenida en la propia imagen y no depende de la de otras imágenes. Las imágenes **I**, proporcionan un mecanismo para el acceso aleatorio a los datos de vídeo comprimido. El segundo tipo, las imágenes **P**, como anteriormente se ha explicado, utilizan para la codificación la información contenida en la imagen previa, **I** o **P**, más próxima. A esta técnica se le conoce como *predicción hacia adelante*. Por último, las imágenes **B**, utilizan como referencia imágenes futuras y previas. Se denomina a esta técnica *predicción bidireccional*. Las imágenes **P** pueden propagar errores de codificación cuando las imágenes **P** (o las **B**) se han predicho mediante imágenes **P** con defectos de predicción. Por el contrario, las imágenes **B** no propagan errores ya que nunca se usan como refe-

rencia. Como veremos más adelante, las imágenes I proporcionan poca compresión, en las imágenes P la compresión es mayor, pero la compresión es aún mayor en las imágenes B. En la *figura 3* podemos ver ilustrado lo que se ha descrito anteriormente.

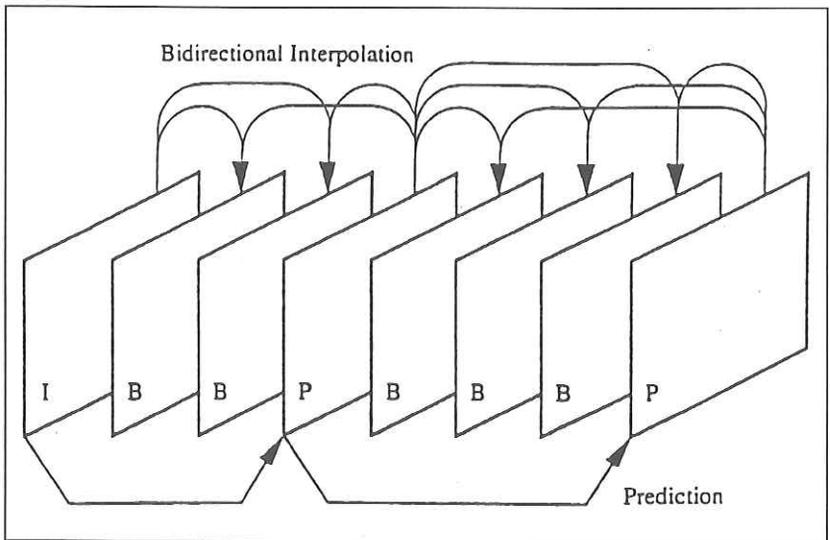


FIGURA 3.
Relación temporal de los tipos de cuadro MPEG.

La sintaxis del MPEG permite al codificador elegir la frecuencia con que se presentan imágenes I para tener en cuenta la necesidad de acceso aleatorio y la situación de los cambios de escena en la secuencia de vídeo. En el codificador también puede determinarse el número de imágenes bidireccionales intercaladas entre cualquier par de imágenes de referencia (I o P). El resto de las imágenes dentro de un determinado grupo de imágenes serían del tipo P. La organización de los tres tipos de imágenes en una secuencia, seleccionada por el decodificador, depende de las características de la aplicación. En la *figura 4* se representa una posible disposición de las imágenes Y, P y B. Sin embargo, en el flujo de vídeo es conveniente cambiar el orden para que las imágenes se presenten en el decodificador en una secuencia más apropiada. Las imágenes de referencia necesarias para construir una imagen B tienen que enviarse antes de la información que describe la imagen B tal como se indica en la *figura 4*.

Cada trama de la señal de vídeo con exploración entrelazada consta de dos campos separados por un período de campo. El MPEG permite codificar la trama como una imagen, o bien codificar los dos campos como dos imágenes. La codificación por tramas es el método preferido

Orden de presentación

1I 2B 3B 4P 5B 6B 7P 8B 9B 10I 11B 12B 13P

Orden en el flujo de vídeo

1I 4P 2B 3B 7P 5B 6B 10I 8B 9B 13P 11B 12B

FIGURA 4.
Organización del flujo de datos de vídeo.

normalmente cuando la escena contiene mucho detalle y poco movimiento. La codificación por campos, en donde el segundo campo puede predecirse mediante el primero, es más adecuada cuando existe movimiento rápido.

2.3. Transformada de coseno discreta

La compensación de movimiento reduce la redundancia temporal de la señal de vídeo, pero todavía queda redundancia espacial en el residuo de movimiento compensado y con más razón si no se ha aplicado este proceso y se ha codificado la trama original. Es decir, existe redundancia espacial en los tres tipos de trama que hemos considerado y para reducirla el codificador MPEG emplea, como la mayoría de los sistemas de compresión, el método de la transformada de coseno discreta (DCT). En general, la DCT transforma un bloque con un número determinado de pixels en otro bloque con un número igual de coeficientes de la transformada. La información de pixeles o la residual (cuando se ha aplicado compensación de movimiento) se convierte mediante la DCT al dominio de frecuencias, concentrando la mayor parte de la energía en una pequeña fracción de los coeficientes de la transformada. Aunque en teoría la transformada podría aplicarse directamente a una trama completa, por muchas razones es más conveniente subdividir la imagen en bloques de 8×8 pixels y realizar la transformación sucesivamente en cada uno de los bloques hasta que toda la imagen se ha transformado. En el decodificador se aplica la transformación inversa para recuperar la imagen original. El primer coeficiente obtenido, tiene un valor mucho más elevado que los restantes. Este coeficiente en particular, es por definición el doble de la media de los 64 pixeles del bloque y representa la energía de la componente continua de todo el bloque. Los demás coeficientes representan niveles de energía a frecuencias horizontales crecientes de izquierda a derecha y a frecuencias verticales crecientes desde arriba hacia abajo. Si la imagen tiene poco detalle, los valores de frecuencias altas son muy pequeños.

Por ejemplo, si un bloque representa una escala de grises, al aplicar la transformada se obtiene el coeficiente de componente continua, otros tres distintos de cero, y 60 coeficientes iguales a cero.

2.4. Cuantificación de los coeficientes

Todos los procesos descritos hasta ahora han dado lugar a una representación de la imagen en forma de vectores de movimiento, coeficientes de la transformada y componentes de luminancia y crominancia, pero no se ha conseguido ninguna compresión. Concretamente, el proceso de transformación no supone ninguna reducción de bits ya que el número de bits por coeficiente necesario para el bloque transformado es igual al del bloque original. Lo que ocurre es que los coeficientes de frecuencia de la transformada se adaptan mucho mejor a las técnicas de compresión que posteriormente se aplican y ésta es su gran ventaja. Estas técnicas aprovechan, entre otras cosas, el hecho de que los bloques transformados contienen un gran número de coeficientes de valor 0 o muy próximo a cero, especialmente cuando se trata de tramas P o B. El **proceso de cuantificación** asigna por aproximación un valor a cada coeficiente de frecuencia dentro de una limitada gama de valores admitidos. El codificador selecciona una «matriz de cuantificación» que determina el modo en que será cuantificado cada uno de los coeficientes del bloque transformado. Como se sabe, el ojo humano tiene poca sensibilidad para los detalles y por tanto no es necesario cuantificar con precisión los coeficientes de alta frecuencia de la transformada, es decir, habrá menos valores admitidos para éstos que para los coeficientes de frecuencias bajas. Esta operación se realiza dividiendo los coeficientes por un valor n mayor que uno y redondeando el resultado al entero más próximo (en el campo digital). El factor de ponderación n varía con la posición del coeficiente en el bloque, correspondiendo a los coeficientes de frecuencias más altas, mayores valores de n . En consecuencia, gran parte de los coeficientes de la transformada se cuantifican al valor cero, habrá muy pocos de alta frecuencia distintos de cero y algunos de baja frecuencia distintos de cero. Un caso especial es el coeficiente que representa el valor de la componente continua del bloque que normalmente se cuantifica con la máxima precisión.

La matriz de cuantificación del MPEG que contiene los valores de n tiene también en cuenta lo siguiente:

- a) Si esta procesando información de luminancia o de crominancia, lo que supone una distinta respuesta del ojo humano.

- b) Si el bloque proviene de una imagen Y o bien de una imagen de predicción, ya que la distribución de las amplitudes de los coeficientes es distinta.
- c) La situación del bloque dentro de la imagen y el contenido de la imagen. Algunos bloques tienen que codificarse con más precisión que otros como, por ejemplo, en el caso de bloques correspondientes a gradientes muy suaves en donde las pequeñas imprecisiones son muy apreciables.

Además de esta cuantificación dependiente de la frecuencia, también es posible reducir el número de niveles de cuantificación necesario para describir los valores de los coeficientes utilizando una ley de cuantificación no lineal, es decir, dependiente de la amplitud. Vemos en la *figura 5* que los valores altos de los coeficientes se codifican con menos precisión que los pequeños. La longitud de la palabra de código a la salida del cuantificador se reduce con respecto a la entrada. Además, todos los valores de la zona muerta se llevan a cero. El MPEG permite cambiar los valores de la matriz de cuantificación para cada bloque DCT cuando se codifican imágenes de gran complejidad. Naturalmente todos los cambios de la matriz tienen que transmitirse al decodificador.

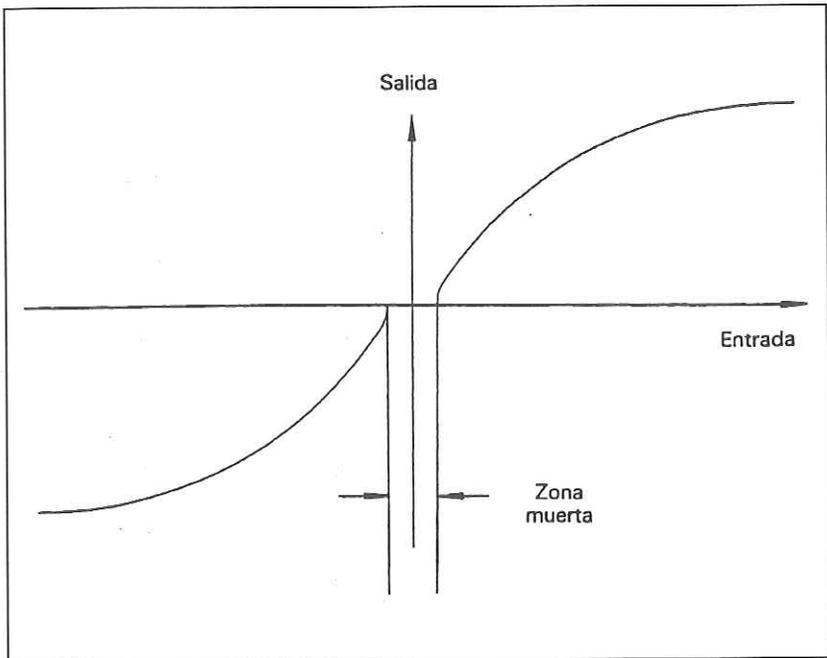


FIGURA 5.
Característica de la cuantificación no lineal.

2.5. Codificación de longitud variable y de series

Los códigos de longitud variable (VLC) o de entropía aumentan la eficacia de la codificación asignando palabras de código más cortas a las secuencias que ocurren con más frecuencia y de mayor longitud a las de menor frecuencia. Como ya se ha dicho, la combinación del proceso DCT y la cuantificación da lugar a que, después del redondeo, muchos coeficientes sean cero, especialmente en las frecuencias de orden elevado. En la práctica, el flujo de coeficientes DCT procesados a la salida de la memoria puede esperarse que contenga largas series de ceros y la probabilidad de que esto ocurra puede aumentarse explorando la memoria en zig-zag tal como se representa en la *figura 6*. Al explorar el bloque en el sentido de coeficientes de frecuencia creciente, se agrupan los coeficientes de baja frecuencia lo que da lugar a que al principio de la secuencia se encuentre la mayoría de los coeficientes de un valor distinto de cero más importantes (en lo que respecta a energía y percepción visual). Seguirán a estos coeficientes largas series de ceros que pueden ser eficazmente representados mediante codificación de longitud de series (RLC). Con esta técnica se codifica el número de coeficientes consecutivos de valor cero que aparecen antes de que se codifique un coeficiente de valor distinto de cero, añadiendo a conti-

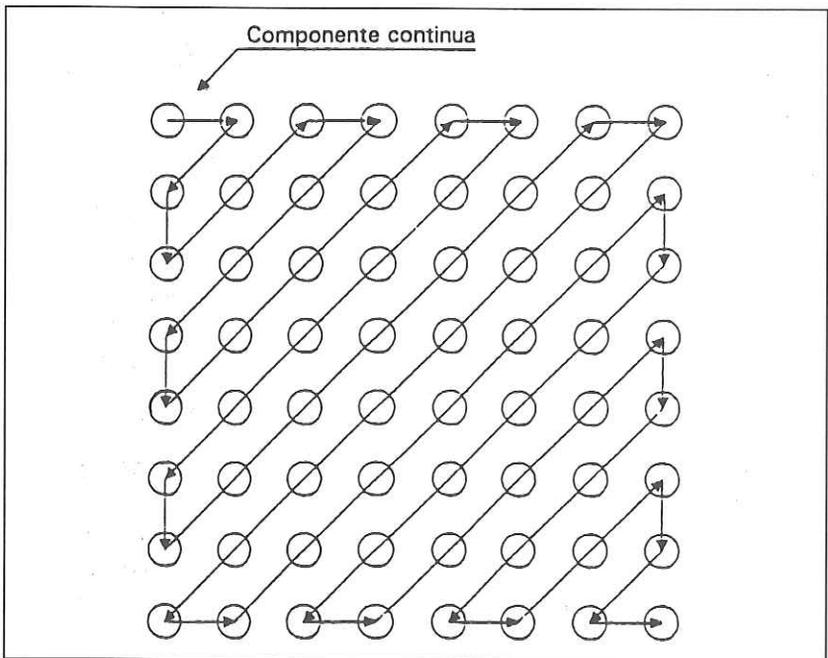


FIGURA 6.
Exploración en zig-zag de un bloque de 8 x 8 píxeles.

nuación el valor de este coeficiente. El resultado es, pues, un conjunto de parejas de valores, en el que cada pareja representa el número de coeficientes consecutivos de valor cero y el valor de un coeficiente distinto de cero. Finalmente estas parejas se someten a una codificación de longitud variable (VLC) que asignará códigos distintos a las parejas de acuerdo con la frecuencia con que aparecen.

2.6. Codificador MPEG-2

En la *figura 7* se representa el diagrama de bloques simplificado del codificador MPEG-2. La parte correspondiente a la compensación de movimiento aparece también en la figura 2. Se ha complementado ésta con la inclusión, en el bucle de realimentación, del cuantificador y DCT inversos que simulan las funciones del decodificador. Junto con la salida de las unidades VLC y RLC de codificación de entropía, se aplica al multiplexor la información de vectores de compensación de movimiento. Cuando se emplea codificación de entropía, las palabras de código son de longitud variable en función de la estructura del vídeo, o sea, la velocidad binaria no es constante. Es, pues, necesario intercalar un *buffer*, o tampón, para que el flujo de bits se transmita a una velocidad constante. Para evitar que el buffer se sature o se vacíe, mediante un bucle de realimentación se envían datos de control al cuantificador. Este control es una parte esencial del codificador y su objetivo

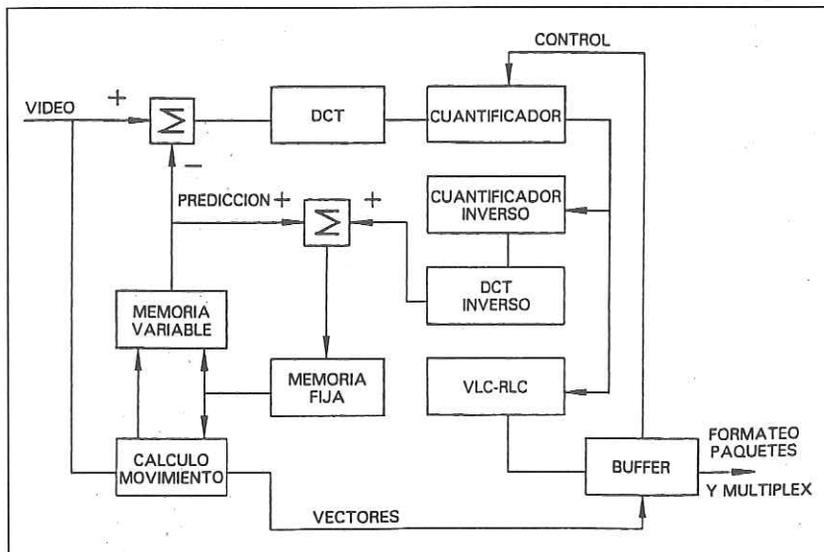


FIGURA 7.
Diagrama de bloques simplificado del decodificador MPEG 2.

es mantener constante la velocidad binaria media a un valor igual al que tiene que utilizarse en el canal. Si la velocidad binaria disminuye apreciablemente, la codificación puede hacerse más precisa para aumentarla. Por el contrario, si el buffer está cerca de su capacidad máxima, se dan instrucciones al cuantificador para que codifique los valores de los coeficientes con menos precisión, reduciendo así el número de bits necesarios para describir una determinada gama de valores. Además de los coeficientes codificados y la información de los vectores de movimiento, llegan al multiplexor el audio, datos auxiliares y de control y la información de sincronismo.

2.7. Multiplex en MPEG-2

Evidentemente sería muy ventajoso que un solo sistema de televisión digital pudiese servir para los distintos medios de transporte, es decir, transmisión por satélite, televisión terrenal, distribución por cable o B-RDSI con ATM. Es, sin embargo, muy difícil conseguir que sean iguales todos los elementos que componen la cadena de transmisión ya que son muy distintas las características físicas de los diferentes medios. Ello impide que puedan aplicarse los mismos métodos de modulación y codificación de canal de un modo general. Lo que verdaderamente es susceptible de aplicación generalizada es el sistema de multiplex, y en América ya se ha adoptado para las futuras emisiones de TVAD un sistema de multiplex basado en la especificación MPEG-2 y lo mismo se ha hecho en Europa para la televisión digital por satélite y cable. No hay duda de que para la televisión terrenal en Europa y para ésta y otras aplicaciones en el resto del mundo, el MPEG-2 será la base de los sistemas de multiplex.

La primera parte de la norma MPEG-2, conocida como parte de sistemas en la terminología del MPEG, se refiere a la combinación de uno o más flujos elementales de vídeo, audio y otros datos en flujos simples o múltiples adecuados para almacenamiento o transmisión. Los flujos elementales de datos de vídeo y audio codificados de acuerdo con las partes 2 y 3 de la norma, se les da un formato de paquetes para constituir los flujos elementales de paquetes, conocidos por las siglas PES. Estos, a su vez, pueden combinarse para formar un *flujo de programa* o un *flujo de transporte*. Cada uno de estos flujos está optimizado para un determinado conjunto de aplicaciones. Se define como sigue:

Flujo de Programa: Combina uno o más flujos elementales con una base de tiempo común para formar un flujo único. Los paquetes de flujo de programa pueden ser de longitud variable y

relativamente grande. Se supone que tienen que utilizarse en condiciones en que no existen errores.

Flujo de Transporte: Combina uno o más flujos elementales con una o más bases de tiempo distintas para formar un flujo único. Los paquetes de flujo de transporte tienen longitud fija y relativamente corta. Se supone que el flujo de transporte se utiliza en condiciones en que se producen errores, como por ejemplo, en redes ATM en las que se pueden producir pérdida de información.

Los dos tipos de flujos, que tienen una estructura de paquetes, proporcionan la sintaxis necesaria para sincronizar la decodificación del vídeo y del audio asegurando a la vez que los buffers de datos de los decodificadores no se saturan ni se vacían. Incluyen además la información de sincronismo para el vídeo y el audio. Tanto en el sistema americano de televisión de alta definición de la Gran Alianza como en el sistema europeo del proyecto DVB se utilizará el flujo de transporte que es el adecuado para medios en los que probablemente se producirán errores.

Cada paquete de transporte, de 188 octetos de longitud, comienza con un encabezamiento de cuatro octetos con el que se pueden identificar el contenido del paquete y la naturaleza de los datos. Los 184 octetos restantes constituyen la carga útil que consiste en paquetes PES individuales con sus encabezamientos. Los paquetes PES de vídeo, audio o datos, que se vuelven a empaquetar para formar el flujo de transporte, no son de longitud fija y son más largos que los paquetes de transporte. Sin embargo, los dos tipos de paquetes se alinean de forma que el comienzo de un paquete PES de vídeo o audio coincida con el de un paquete de transporte, completando con octetos de relleno los paquetes de transporte parcialmente ocupados. Esto quiere decir que cada paquete de transporte contiene únicamente un tipo de datos (vídeo, audio o datos auxiliares). La estructura de los paquetes de transporte no tiene ninguna relación con la estructura lógica de los datos de entrada como, por ejemplo, las líneas o tramas del vídeo. Los cuatro octetos de encabezamiento sirven también para la sincronización del paquete, control de errores y acceso condicional. A veces es necesario disponer de información adicional a la del encabezamiento. Esto se consigue con el llamado encabezamiento de adaptación, un campo de longitud variable situado en la carga útil del paquete de transporte. Con ayuda de este campo se pueden realizar un gran número de funciones, como por ejemplo, acceso aleatorio al flujo de bits comprimido (cambio de canal) o inserción de un canal local.

2.8. Estructura en capas del flujo de vídeo

La sintaxis, o conjunto de reglas, del MPEG-2 puede dividirse en *dos categorías fundamentales*: la primera es la **sintaxis sin jerarquización** de las señales de vídeo, que se ha estructurado como una extensión del MPEG-1 añadiendo, entre otras cosas, nuevos métodos de compresión para señales de vídeo con exploración entrelazada. La segunda es la **sintaxis jerarquizada** que permite la reconstrucción de una señal de vídeo utilizando partes de un flujo binario. Esto se consigue estructurando el flujo binario en dos o más capas, comenzando por una capa básica, que puede ser única, y añadiendo una o más capas que mejoran la calidad que proporciona la capa básica.

Este tipo de sintaxis del MPEG-2 permite aplicaciones que no serían posibles con una sola capa de vídeo. Entre las áreas de aplicación pueden citarse las telecomunicaciones de vídeo, servicios de vídeo con múltiples niveles de calidad, televisión de alta definición con TV de definición normal incorporada y sobre todo vídeo en redes de transferencia asíncrona (ATM), sobre las cuales se puede implementar técnicas que hacen al medio más robusto frente a errores de transmisión o a pérdidas de información. Esto último, esta siendo un campo de investigación muy importante en la actualidad. Aunque siempre existe la alternativa de transmitir simultáneamente distintos tipos, o reproducciones, de vídeo codificados con independencia, es mucho más eficaz la codificación jerarquizada o por capas en las que el ancho de banda asignado a una determinada reproducción de vídeo puede ser reutilizado parcialmente para codificar otra reproducción de vídeo. Como es lógico, los codificadores de vídeo que utilizan esta técnica serán más complejos que los correspondientes a una sola capa de vídeo. Por otra parte, cuando se transmite un flujo binario codificado por capas, pueden utilizarse decodificadores que permitan la recepción de uno o otro tipo de vídeo de acuerdo con la complejidad del decodificador.

2.9. Perfiles y niveles

La norma MPEG-2, con sus diferentes técnicas de compresión y codificación suele considerarse como un conjunto de instrumentos o herramientas. De este conjunto puede extraerse los instrumentos necesarios para cada aplicación determinada. De la gama completa de instrumentos se han definido cinco subconjuntos a los que se denomina **Perfiles**. Éstos tienen una complejidad progresiva, y cada uno de ellos añade instrumentos adicionales al Perfil precedente. Quiere esto decir que cada Perfil adicional puede ejercer más funciones que el anterior, pero también necesitará más área de silicio y por consiguiente su precio será más elevado para el usuario. La entrada de todos los sistemas

es vídeo en componentes, pero de los cinco perfiles, los cuatro primeros codifican la señal de diferencia de color en líneas secuenciales, como en el SECAM, es decir, R-Y en una línea, B-Y en la siguiente, y así sucesivamente.

El **Perfil Simple** es el que tiene menos instrumentos. Utiliza compensación de movimiento y transformada de coseno discreta.

El **Perfil Principal** tiene todos los instrumentos del Perfil Simple y además la predicción bidireccional. Para la misma velocidad binaria dará una calidad de imagen mejor que el Perfil Simple. Un decodificador de Perfil Principal decodificará imágenes codificadas de acuerdo a los Perfiles Principal y Simple. Este tipo de compatibilidad con Perfiles de menos complejidad se aplica a todos los sucesivos Perfiles.

El **Perfil Jerárquico o de capas** también conocido como «*SNR Scalability Profile*», en donde SNR representa las iniciales de «Signal-to-Noise Ratio» (relación señal/ruido). Le llamaremos *Perfil de Jerarquía en RSR*. Tiene todos los instrumentos del Perfil Principal y además la capacidad de separar los datos en dos partes, que pueden considerarse como una señal de base y una señal de realce o mejora. La primera es una versión de la imagen con una relación señal/ruido reducida, pero que solo necesita una fracción de la velocidad binaria de la señal completa. Cuando se codifican conjuntamente las señales de base y de realce, aumenta la relación señal/ruido de la imagen hasta el máximo posible y se obtiene la misma calidad que con el Perfil Principal (para la misma velocidad binaria). Cada una de las dos señales utiliza una parte de la velocidad binaria total, y ambas tienen la misma resolución espacial. Esta posibilidad de dividir los datos no exige una mayor velocidad binaria, y tiene interesantes aplicaciones. La señal de base puede hacerse más resistente a errores que la señal de realce, de forma que cuando exista la posibilidad de perder información, esta se produzca en la señal de realce y no en la señal de base. Esto se puede conseguir, por ejemplo, en la transmisión de vídeo sobre redes ATM, donde la información puede enviarse con dos niveles diferentes de prioridad. Además, en el caso de que no se pueda transmitir el flujo con toda la velocidad binaria, podría transmitirse una versión de menor velocidad.

El **Perfil de Jerarquía Espacial** (Spatial Scalable Profile, en la terminología del MPEG-2) tiene todos los instrumentos del anterior más un segundo método para dividir los datos. En este Perfil, la imagen codificada puede dividirse de acuerdo con la resolución (número de elementos por línea que pueden distinguirse). Una parte de la señal dividida puede ser decodificada para dar una imagen con menor resolución que la correspondiente a la imagen original. Cuando se decodifican conjuntamente las partes en que se ha dividido la señal se obtiene una imagen que tiene la misma resolución que la original. Aquí también te-

mos una señal de base y una señal de realce, cada una de las cuales utiliza una parte de la velocidad binaria total. Este Perfil, a diferencia del anterior, tiene el inconveniente de que para dar la misma calidad que el Perfil Principal, necesita una velocidad binaria superior en un 10 a 15 %. Como en el Perfil anterior, la capa de base puede hacerse más resistente y difundir una señal de menor resolución que cubra mayor superficie, o en condiciones de recepción más difíciles.

El **Perfil Alto** consta de todos los instrumentos del Perfil anterior más la capacidad para codificar señales de diferencia de color en líneas simultáneas. Aunque a velocidades binarias bajas la calidad se resiente, a velocidades altas, normales en aplicaciones de este Perfil, la posibilidad de codificación de diferencias de color en líneas simultáneas supone una mejora de calidad. Este Perfil es el sistema de más complejidad y mejores prestaciones, diseñado para las aplicaciones más exigentes en las que la velocidad binaria no constituye un problema.

Conjuntamente con los cinco Perfiles, se han definido cuatro **Niveles** que corresponden al formato de imagen utilizado a la entrada. Para el **Nivel Bajo** el formato de entrada es igual a 352 muestras por 288 líneas. Al **Nivel Principal** corresponde el formato de imagen de la Recomendación 601 (720 muestras por 576 líneas). El **Nivel Alto-1440** al que corresponde una entrada de televisión de alta definición con 1440 muestras por 1152 líneas. El cuarto nivel es el **Nivel Alto** al que corresponde una entrada de alta definición con 1920 muestras por 1152 líneas.

Los decodificadores para cualquier determinado nivel, pueden también utilizarse para los niveles inferiores. Sin embargo, no podrán decodificar los niveles superiores a menos que, como parte de estos, exista una capa de resolución más baja. No todas las combinaciones de Niveles y Perfiles resultan útiles o necesarias y hasta la fecha sólo se ha demostrado interés por once de las 20 posibles combinaciones. Se les llama Puntos de Conformidad del MPEG-2 y en cada uno puede operarse en una gama de velocidades binarias. En la *figura 8* se han representado los Perfiles y Niveles del MPEG-2 con indicación de sus características más sobresalientes. Los once cuadros en los que se ha señalado la velocidad binaria son los puntos de conformidad. Los valores de pixels (o muestras) y líneas se refieren al área activa de la imagen. Los valores de las líneas, indicados anteriormente en el texto y en la *figura 8*, corresponden a los sistemas de 25 Hz de frecuencia de trama. Para el caso de 30 Hz, el número de líneas sería 240 para el Nivel Bajo, 480 para el Principal y 1080 para los Niveles Alto-1440 y Alto. Los sistemas MPEG-2 permiten exploración progresiva o entrelazada a frecuencias de 50 ó 60 Hz. En cada punto de conformidad existe un límite máximo de muestras de luminancia.

	SIMPLE Sin tramas B 4:2:0	PRINCIPAL Tramas B 4:2:0	JERARQUÍA RELACIÓN S/R Tramas B 4:2:0	JERARQUÍA ESPACIAL Tramas B 4:2:0	ALTO Tramas B 4:2:0
ALTO 1920 pixels 1152 líneas		80 Mbit/s			100 Mbit/s
ALTO-1440 1440 pixels 1152 líneas		60 Mbit/s		60 Mbit/s	80 Mbit/s
PRINCIPAL 720 pixels 576 líneas	15 Mbit/s	15 Mbit/s	15 Mbit/s		20 Mbit/s
BAJO 352 pixels 288 líneas		4 Mbit/s	4 Mbit/s		

FIGURA 8.
Perfiles y niveles del MPEG-2.

3. APLICACIONES

Los fabricantes sólo estarán dispuestos a producir receptores para los servicios que ofrezcan garantías de continuidad y aceptación por el público. En Europa sólo se han interesado por el MPEG-2 Nivel Principal/Perfil Principal que será la base de los receptores de primera generación para las emisiones por satélite y los correspondientes servicios de cable. Se han evaluado las prestaciones del sistema Nivel Principal/Perfil Principal mediante métodos de simulación, llegando a las siguientes conclusiones provisionales:

- Para conseguir una calidad de estudio (Recomendación 601) se necesita un sistema con velocidad binaria de unos 9 Mbit/s. Para una calidad equivalente a la del PAL, se necesitarían unos 5 Mbit/s.
- Para la misma velocidad binaria, los sistemas de 50 Hz proporcionan mejor calidad que los de 60 Hz.
- Es más fácil codificar vídeo de películas que vídeo de cámara. En el primer caso, 4 Mbit/s puede considerarse aceptable.

Con excepción del Perfil Simple, los demás perfiles del MPEG-2 tienen parámetros adecuados para la televisión de alta definición. El sistema Gran Alianza propuesto como norma de transmisión de televi-

sión de alta definición en los Estados Unidos probablemente utilizará la combinación Nivel Alto/Perfil Principal.

Volviendo a la situación en Europa, existe un gran interés en las aplicaciones de los Perfiles con capas jerarquizadas para las emisiones terrenales de televisión digital. Por ejemplo, con el Perfil de Jerarquía Espacial que permite dividir los datos de acuerdo con la resolución y también con la relación señal/ruido, se puede crear una señal compuesta por tres elementos que conjuntamente, darían lugar a una señal de alta definición. de las tres partes de la señal, la de base utilizaría jerarquía espacial para proporcionar una señal de 625 líneas. El resto de la señal se puede dividir aplicando el criterio de relación señal/ruido (jerarquía en RSR), creando así una segunda señal que junto con la capa de base proporcionaría una señal de alta definición con una relación señal/ruido reducida. El tercer elemento de la señal, conjuntamente con los otros dos, daría una señal con la calidad total de alta definición.

Existen otras posibilidades, como por ejemplo, un sistema de dos capas utilizando jerarquía espacial entre un servicio de definición normal y otro de alta definición, o bien un sistema de dos capas con jerarquía RSR entre dos servicios con el mismo número de líneas.

4. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha presentado un sistema estandarizado de compresión de vídeo orientado a la televisión digital. Son muchas las aplicaciones futuras que se tendrán con esta norma que ha sido aceptada mundialmente. Se pretende con esta norma establecer un sistema estandarizado para la transmisión de vídeo en particular y en general para un sistema multimedia. Como trabajo futuro, nos planteamos partir de esta norma estandarizada, y dotarla de técnicas de ocultación de errores o pérdidas de información, que se hayan podido producir en la transmisión de la señal codificada en una red ATM, obteniendo por tanto un sistema resistente a errores o pérdidas de información cuando se transmite señal codificada con MPEG-2 sobre una red de transmisión ATM.

5. BIBLIOGRAFÍA

- Generic Coding Of Moving Pictures And Associated Audio. Recommendation H.262. ISO/IEC 13818-2. Draft Internacional Standard. March 1994.
- CUENCA CASTILLO, Pedro: «Redes de Interconexión en Sistemas Distribuidos con Tecnología ATM». Tesis de Licenciatura. Universidad de Valencia. Junio 1996.
- CUENCA CASTILLO, Pedro y OLIVARES MONTES, Teresa: «Mecanismos de Control de Error para Mejorar la Transmisión de Vídeo MPEG sobre Redes ATM». Departamento de Informática. Escuela Universitaria Politécnica de Albacete.

DIDIER LE GALL: «MPEG: A Video Compression Standard for Multimedia Applications». Communications of the ACM, vol. 34, n° 4, pp. 47-58, April 1991.

EDUARDO GAVILÁN ESTELAT: «MPEG-2: Pieza Clave de la Televisión Digital». Instituto Oficial de RadioTelevisión Española.

GREGORY K. WALLACE: «The JPEG Still Picture Compression Standard». Communications of the ACM, vol. 34, n° 4, pp. 31-45, April 1991.