5 Codificación de Video

MPEG

5.1. Introducción

Las siglas MPEG significan Moving Picture Experts Group y es el grupo

de trabajo del subcomité del ISO/IEC (International Organization for Standarization

/ International Electrotechnical Commission) encargado del desarrollo

de las normas internacionales para la compresión, descompresión, procesado y

codificación de imágenes animadas, audio o la combinación de ambas. El MPEG

define la sintaxis de las señales digitales correspondientes a audio y video - tanto

de origen natural como sintetizado, describe su estructura y contenido y regula

el funcionamiento de decodificadores estandarizados. El MPEG no define los

algoritmos de codificación. Esto permite una mejora continuada de los codificadores

y su adaptación a aplicaciones especificas dentro de la norma. Además de

la codificación de audio y video, el MPEG también define sistemas para multiplexor

la información de audio y video en una ´única señal digital, describe los

métodos para verificar que las señales y los decodificadores se ajustan a la norma

y publica informes técnicos con ejemplos de funcionamiento de codificadores y

decodificadores. Los estándares MPEG fueron desarrollados para ser independientes

de la red especifica para proporcionar un punto de interoperabilidad en

entornos de red heterogéneos.

5.1.1. Fases, Secciones y Capas

Hay que aclarar dos cosas. En primer lugar, el MPEG trabaja por fases.

Las fases se identifican con números (MPEG-1, MPEG-2, MPEG-4, MPEG-7).

Estas fases no describen diversas versiones de una ´única norma, sino que son

normas completamente distintas que se encargan de aspectos diferentes de la

comunicación multimedia. Asi, las ´ultimas fases NO reemplazan a las anteriores

sino que las complementan. En segundo lugar hay que explicar el concepto de

capa. Tanto en MPEG-1 como en MPEG-2 se han definido tres capas diferentes.

A veces son llamadas incorrectamente ”niveles”. Cada capa representa un conjunto

de algoritmos de codificación. Las capas se suelen identificar con números

romanos. Por ejemplo: Capa I, Capa II y Capa III (´esta ´ultima es el origen del

popular MP3). El concepto de versión sólo se usa en MPEG-4. La Versión 1 de

MPEG-4 proporciona una serie de herramientas para la codificación del audio.

Con la Versión 2, se añaden nuevas herramientas que incrementan la utilidad

pero que no reemplazan a las herramientas de la Versión 1. La Versión 2 es por

lo tanto completamente compatible con la Versión 1.

5.2. MPEG-1

El primer estándar que el MPEG ( Moving Picture Experts Group) introdujo

fue MPEG-1, usado para compresión de video en CDs. MPEG-1 usa una baja

tasa de bit, similar a la resultante de una cinta de video VHS. MPEG-1 se

considera como un video solamente progresivo (no entrelazado), que alcanza

una tasa de 1.5 Mbps.

5.2.1. Descripción General

El algoritmo de compresión de video de MPEG utiliza dos técnicas fundamentales:

Compensación del movimiento basada en bloques para la reducción

de la redundancia temporal, y Codificación DCT (Discrete cousine Transform)

para la reducción de la redundancia espacial. La técnica de compensación de

movimiento se aplica en ambas direcciones: hacia adelante o causal (forward) y

hacia atrás o no causal (backward). La señal restante es codificada utilizando

las técnicas basadas en transformaciones. Los predictores de movimiento, denominados

vectores de movimiento, son transmitidos junto con la información

espacial.

Reducción de la redundancia temporal

Para soportar el acceso aleatorio al video almacenado, se definen tres tipos

fundamentales de imágenes o cuadros: codificados internamente (I), predictivos

(P) e interpolados bidireccionalmente (B).

La predicción para la compensación del movimiento supone que la imagen

actual puede ser modelada como una traslación de las imágenes precedentes.

En es estándar MPEG, cada imagen es dividida en bloques de 16x16 pixeles

denominados macro bloques. Cada macro bloques es predecido a partir del frame

anterior o del siguiente estimando la cantidad de movimiento en el macro bloque

durante el intervalo entre frames. La sintaxis de MPEG especifica como

representar la información de movimiento para cada macro bloque, utilizando

para ello vectores de movimiento. Sin embargo no especifica como van a ser

calculados estos vectores.

Reducción de la redundancia espacial

Para la reducción de la redundancia espacial en cada frame I o en la predicción

de errores en frames P o B, el estándar MPEG utiliza técnicas de codificación

basadas en DCT.

La operación básica del DCT es transformar una serie de puntos del dominio

espacial a una representación idéntica en el dominio de la frecuencia. La DCT se

aplica sobre una matriz generalmente de 8x8 cuya salida es otra matriz de iguales

dimensiones que contiene los coeficientes DCT que están ordenados de forma

que los que contienen información ´útil están en la esquina superior izquierda.

El coeficiente DC es la posición (0,0) y representa la media de los 63 valores.

Codificación

El ´ultimo paso el la codificación de las imágenes a través de dos pasos:

1. Codificación del elemento DC como un valor relativo respecto a valores

anteriores, ya que este valor tiene altos grados de correlación.

2. Reordenación de los valores DCT en zig-zag ya que hay tantos de estos

elementos cuyo valor es cero que deben ser codificados de forma diferente

que los que no son cero. Se usa RLE (run-length encoding) que no hace

sino contar el número de ceros en la imagen.

5.2.2. Tipos de Imagen MPEG

MPEG define tres tipos de imágenes que se encadenan según el esquema de

la figura. Los cuales son el soporte de la codificación diferencial y bidireccional,

minimizando la propagación de errores.

Imágenes I (intra)

Son imágenes que no requieren información adicional para su decodificación.

Son codificadas sin ninguna referencia a otras imágenes, como en JPEG, es

decir, que contiene todos los elementos necesarios para su reconstrucción por el

decodificador y son, por ello, el punto de entrada obligatorio para el acceso a

una secuencia. La tasa de compresión de imágenes I es relativamente pequeña,

comparable con la de JPEG.

Las imágenes I se codifican como una imagen ´única usando solo la información

de la imagen. Los bloques tienen una gran redundancia espacial por lo que

MPEG trata de disminuir esta cantidad de datos.

Imágenes P (Previstas)

Se codifican con respecto a las imágenes de tipo I o P anteriores, gracias a

las técnicas de predicción con compensación de movimiento. Como la compensación

de movimiento no es perfecta, no se podrá multiplicar indefinidamente el

número de imágenes I, ya que utilizan para decodificar otras imágenes P o B, se

propagan amplificando cualquier error de codificación. Su tasa de compresión

es claramente mayor que la de las imágenes I.

Las imagines P requieren aproximadamente la mitad de los datos de las

imágenes I. Normalmente imagen a codificar es muy similar a la anterior (que

puede ser I o P) ya que se trata de una imagen desplazada respecto a la anterior.

La mayoria de los cambios entre la imagen de referencia y la imagen a codificar

se pueden representar como traslaciones de pequeñas regiones de imágenes de

forma que la técnica a utilizar es la predicción por compensación de movimiento.

Cada macro bloque en una imagen P se puede codificar tanto como un macro bloque

I como P. Un macro bloque codificado como P se codifica como un

´área de 16x16 de la imagen tomada como referencia mas un termino de error

que contiene la diferencia entre los dos macro bloques. Para expresar el ´área de

16x16 de la imagen de referencia se usa un vector de movimiento. Por ejemplo

un vector de movimiento (0,0) significa que el ´área de referencia está en la

misma posición que el macro bloque a codificar. El término de error se codifica

nuevamente usando DCT, cuantificación y RLE (run-length encoding).

Imágenes B (Bidireccionales)

Se codifican por interpolación entre dos imágenes de tipo I o P precedentes y

siguiente que las enmarcan. Como no se utilizan para describir otras imágenes,

las imágenes B no propagan los posibles errores de codificación.

Este tipo de imágenes es el que ofrece el factor de compresión más alto,

que generalmente es de una cuarta parte de los datos de las imágenes I. La

codificación de la imágenes B es similar a la de las P salvo porque los vectores

de movimiento se refieren a ´aéreas de imágenes futuras. En macro bloques que

usan referencias a imágenes tanto previas como futuras, los dos ´áreas de 16x16

son promediados.

Dependiendo de la complejidad del codificador utilizado, se podrán codificar

solo las imágenes I, las imágenes I y P o las imágenes I, P y B; sin duda, con

resultados absolutamente diferentes a nivel del factor de compresión y en cuanto

a las posibilidades de acceso aleatorio, asi como del tiempo de codificación y de

la calidad percibida.

Los dos parámetros M y N definen la manera en que las imágenes I, P y B

se encadenan:

M es la distancia (en número de imágenes) entre dos imágenes P (previstas)

sucesivas.

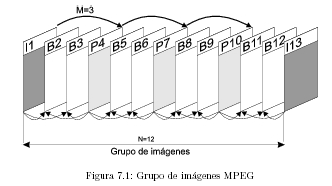
N es la distancia entre dos imágenes I (intra) sucesivas.

Para alcanzar un flujo de video de 1.15 Mbits/s con una calidad satisfactoria,

al tiempo que se mantiene una resolución de acceso aleatorio aceptable (¡0.5

segundos), los parámetros comúnmente utilizados son M=3 y N= 12 como se

muestra en la Figura 7.1;

****

en este caso, una secuencia de video se compone de 1/12 (8.33 %) de imágenes

I, 1/4 (25 %) de imágenes P y de 2/3 (66.66%) de imágenes B. El factor

de compresión global se ve favorecida por el hecho de que son las imágenes más

frecuentes las que tienen un factor de compresión más alto. En la visualización,

tras la codificación y decodificación, es evidente que las imágenes de la secuencia

de video deben ser reproducidas en el mismo orden en que se captaron. Con los

parámetros definidos anteriormente (M=3, N=12), el modo de codificación de

imágenes sucesivas se traduce por la correspondencia número-tipo de imagen

Siguiente:

1(I) 2(B) 3(B) 4(P) 5(B) 6(B) 7(P) 8(B) 9(B) 10(P) 11(B) 12(B) 13(I) 14(B) 15(B) 16(P).....

Sin embargo, para codificar o decodificar una imagen B (bidireccional), el codificador y el decodificador necesitaran la imagen I o P que la precede y la imagen

P o I que la sigue. El orden de las imágenes ser’ a, por tanto, modificado antes de

la codificación, de forma que el codificador y el decodificador dispongan, antes

que las imágenes B, de las imágenes I y/o P necesarias para su tratamiento, o sea:

1(I) 4(P) 2(B) 3(B) 7(P) 5(B) 6(B) 10(P) 8(B) 9(B) 13(I) 11(B) 12(B) 16(P) 14(B) 15(B)...

El aumento del factor de compresión facilitado por las imágenes B se paga,

pues, con un tiempo de codificación/decodificación más largo (duración de dos

imágenes) y un aumento en el tamaño de la memoria necesaria tanto en el codificador como en el decodificador (hay que almacenar una imagen suplementaria).

5.2.3. Transformada Directa del Coseno

La señal de video tiene la caracteristica de ser señal .espacial”, es decir, de

ocupar un espacio de dos dimensiones: una en el sentido horizontal y otra en el

sentido vertical. El concepto de la DCT se basa sintéticamente en tomar cada

pixel de un bloque de 8 por 8 pixels que es una muestra de una señal variable

en el tiempo, proporcional a la luminancia y de otra señal variable en el tiempo,

proporcional a la crominancia. Estas dos señales son las que se pasarán, separadamente,

al dominio de las frecuencias. Imaginemos las dimensiones de un

pixel (visible con una lupa en una pantalla de plasma de 60”): ese pixel nace

como una muestra en un CCD que es excitado por la luz incidente desde el

lente y que provee una muestra de luminancia y otra de crominancia, ambas

funciones del tiempo. Esas muestras, de acuerdo a Fourier, tienen componentes

en diversas frecuencias que tendremos que transmitir fielmente para poder recomponer

la señal en el display. Ahora bien, figurémonos si podremos distinguir

dichas frecuencias hasta el enésimo armónico cuando apenas podemos ver un

pixel con una lupa. El enésimo armónico se confundirá con el ruido. Esta es

debido a una de las propiedades, que determinó Fourier, que la amplitud de

los coeficientes va disminuyendo a medida que aumenta la frecuencia. Esto, nos

permitirá transmitir los coeficientes de alta frecuencia comprimiéndolos fuertemente.

La tecnología actual de computación nos permite calcular con mucha

precisión y velocidad los coeficientes de Fourier para los distintos pixels.

En el proceso de cuantizacion de la DCT, los coeficientes de alta frecuencia

son divididos por un valor mal y el resultado se redondea al entero más cercano.

El valor de n varía con la posición del coeficiente en el bloque, con frecuencias

más altas se asignan valores más altos. Como resultado, coeficientes que representan

frecuencias espaciales bajas son cuantificados con pasos relativamente

bajos y tienen una S/N alta. Los coeficientes que representan las frecuencias

espaciales más altas son cuantificados con pasos grandes y padecen distorsión y

S/N baja. El proceso de cuantificación es controlado por tablas especificas y al

decodificador se le proporciona información acerca del modelo de cuantificación

lo que permite la decodificación correcta.

5.2.4. Descomposición en capas de una secuencia de video

MPEG

Una secuencia de video MPEG es básicamente la salida del material en

bruto de un codificador y solo contiene lo necesario para que un decodificador

restablezca la imagen original. La sintaxis de la señal comprimida es definida

de manera rigurosa por MPEG. La figura 7.2 muestra la construcción de una

secuencia de video MPEG constituida por capas bien definidas. una secuencia

de video es la combinación de con un código de inicio, seguido por un encabezamiento,

y terminando con un código final. Códigos de soporte adicional

pueden ser situados al inicio de la secuencia. La secuencia de soporte especifica

el tamaño horizontal y vertical de la imagen, norma de barrido, la tasa de imágenes,

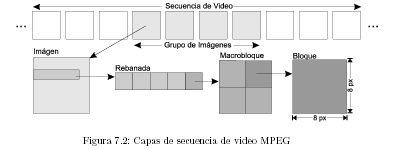
si se usa un barrido progresivo o entrelazado, el perfil, nivel, velocidad de

transferencia de bits, y cuales matrices de cuantificación se usan para codificar

imágenes espaciales y temporales. Sin la secuencia de soporte de datos, un decodificador

no puede comprender el flujo de bits y por lo tanto no puede comenzar

la operación de decodificación correcta. Esto ocurre generalmente cuando un televidente está cambiando canales de un lugar a otro en su televisor.

****

Grupo de Imágenes (GOP)

Son grupos de imágenes. Las imágenes se combinan para producir un GOP

(grupo de imágenes) que comienza con una imagen I. El GOP es la unidad

fundamental de codificación temporal. En el estándar MPEG, el uso de GOP es

opcional, pero esta en la practica es necesaria. Entre imágenes I se colocan un

número variable de imágenes P y/o B como ya se ha descrito. Un GOP puede

ser abierto o cerrado. En un GOP cerrado, las ultimas imágenes B requieren de

una imagen I para el siguiente GOP por decodificar y la secuencia de bits puede

ser cortada al final de la GOP.

Imagen de tipo I, P o B

Cuando un número de rebanadas se combinan, construyen una imagen, la

cual es la parte activa de un campo o un cuadro. La imagen de soporte inicial

define que imágenes I, P o B codifica e incluye una referencia temporal para

que la imagen pueda ser representada en el momento adecuado. En el caso de

tomas panorámicas e i inclinaciones, los vectores en cada macro bloque serán los

mismos. Un vector global puede ser enviado para toda la imagen, y luego se

pueden enviar vectores individuales que lleguen a crear la diferencia en el vector

global.

Rebanada (Slice)

horizontal que está ordenada de izquierda a derecha. En MPEG, las rebanadas

pueden comenzar en cualquier sentido y ser de tamaño arbitrario, pero

el DVB establece que deben comenzar en el borde izquierdo de la imagen. Las

rebanadas son la unidad fundamental de sincronización para la codificación de

la longitud variable y diferencial, los vectores iníciales en una rebanada son

enviados completamente, mientras que los demás vectores son transmitidos diferencialmente.

Pág. 16/39 En imágenes I, los primeros coeficientes DC de las rebanadas son enviados completamente y los demás coeficientes DC son transmitidos

en forma diferencial. En imágenes de diferencia, esta técnica no se utiliza.

Macrobloque (Macroblock)

Es la unidad fundamental de la imagen que además está compensada en movimiento.

Cada macro bloque es un vector de desplazamiento en dos dimensiones

situado en la parte superior de la secuencia. En una imagen B, el vector puede

ser hacia adelante o hacia atrás. La compensación de movimiento puede ser en

modo de cuadro o en modo de campo, el cual es indicado. La escala utilizada

para la recuantificacion de los coeficientes también es indicada. Usando los

vectores, el decodificador obtiene información acerca de las imágenes anteriores

y las posteriores, produciendo así una predicción de imágenes. Los bloques

son transformados inversamente para producir una imagen de rectificación que

es adicionada a la imagen prevista que ha sido producida a la salida del decodificador.

En un formato de codificación 4:2:0, cada macro bloque tendrá 4

bloques Y, y dos bloques de color diferente. Para hacer posible la identificación

de cada bloque y sus componentes, estos se envian en un orden especifico. Cada

macro bloque tiene un tamaño de 16x16 pixels.

Bloque(Block)

Es la unidad fundamental de la información de la imagen y esta representada

por un bloque de coeficientes DCT, que tienen un tamaño de 8x8 pixeles, los

cuales representan datos Y, Cr o Cb.

5.3. MPEG-2

Los distribuidores querian economizar la transmisión digital, pero dado que

MPEG-1 no era adecuado para el satélite y MPEG-2 estaba todavia en desarrollo,

se creó una modalidad ”intermedia” de MPEG llamada MPEG-1.5, que no

siendo un estándar oficial, se usa todavia en algunas redes de satélite). MPEG-

1.5 usa un gran ancho de banda multiplexando varios streams MPEG-1, lo cual

permite cubrir la deficiencia de MPEG-1 de no poder transmitir varios programas

de video a la vez sobre el enlace satélite. La segunda fase de MPEG,

llamada MPEG-2 se acabó convirtiendo en el estándar de facto en el mundo

de la televisión digital ya que arregla muchos de los problemas inherentes a

MPEG-1, tales como la resolución, escalabilidad y manejo de video entrelazado.

MPEG-2 permite imágenes de mucha más calidad (hasta niveles de HDTV) y

permite que muchos canales de diferentes tasas de bit se multiplexen dentro de

un mismo flujo de datos. MPEG-2 también consta de tres capas (o estándares),

cubiertas por la: ISO/IEC 13818-1 Sistemas MPEG-2 (ITU-T Rec. H.222.0),

ISO/IEC 13818-2 Video MPEG-2 (ITU-T Rec. H.262) y ISO/IEC 13818-3 Audio

MPEG-2, aprobadas finalmente como estándar la ISO/IEC en Noviembre de

1994. Con una calidad superior al MPEG-1, MPEG-2 fue universalmente aceptado

para transmitir video digital comprimido con velocidades mayores. Con

MPEG-2 pueden conseguirse elevados ratios de hasta 100:1, dependiendo de las

caracteristicas del propio video. MPEG-2 normalmente define dos sistemas de

capas, el flujo de programa y el flujo de transporte. Se usa uno u otro pero

no los dos a la vez. El flujo de programa funcionalmente es similar al sistema

MPEG-1. La técnica de encapsulamiento y multiplexacion de la capa de compresion

produce paquetes grandes y de varios tamaños. Los paquetes grandes

producen errores aislados e incrementan los requerimientos de bu\_ering en el

receptor/decodificador para demultiplexar los flujos de bits. En contraposición

el flujo de transporte consiste en paquetes fijos de 188 bytes lo que decremento

el nivel de errores ocultos y los requerimientos del bu\_ering receptor. El registro

ITU-T H.262 trata con codificación de video de alta calidad con posible video

entrelazado de NTSC, PAL o Televisión de Alta Definición (HDTV). Es un intento

para operar en un rango de 2 a 15 Mbit/s. Sin embargo puede funcionar a

velocidades superiores de 100 Mbit/s. Un amplio rango de aplicaciones, velocidades,

resolución calidades de las señales y servicios son direccionados, incluyendo

todas las formas de medios de almacenamiento digital, televisión (incluyendo

HDTV), broadcasting y comunicaciones. Entre las varias mejoras o extensiones

introducidas en los codificadores MPEG-2, tenemos:

Nuevos modos de predicción de campos y tramas para scanning entrelazado.

Cuantizacion mejorada.

Nuevos códigos intra-trama de longitud variable (VLC).

Extensión escalada de resoluciones para compatibilidad, servicios jerárquicos

y robustos.

Dos nuevas capas de sistema para multiplexaje y transporte que provee

celdas/paquetes de video de alta o baja prioridad, cuando son llevados a

través de una red conmutada.

Incrementos soportados por accesos aleatorios.

Soporte resistente para incremento de errores.

Múltiples programas con un multiplexor (MPEG-1 no puede hacer esto,

y esto fue un driver principal para el MPEG-2).

7.3.1. Descripción General

Codificación MPEG-2

Al igual que MPEG-1, la norma no define explicitamente el método de codificacion, sino ´únicamente la sintaxis que controla el tren binario a la salida del

codificador, lo cual deja gran libertad a su diseñador. Se describe ahora la codificacion

fijación MPEG-2 bastante similar a la MPEG-1 ya comentada. A partir de la

imagen digitalizada en formato 4:2:0 (caso del main profile), el codificador elige

para cada imagen su tipo (I, P o B) y si esta debe ser codificada en modo frame

(imagen) o field (campo).

El codificador a continuación debe estimar los vectores de movimiento para

cada macro bloque de 16x16 pixeles. El número de vectores depende del tipo de

imagen y del modo de codificación escogido para cada bloque. En el caso más

general, donde el codificador es capaz de generar imágenes B (bidireccionales),

deberá reordenar las imágenes antes de la codificación y la transmisión. La

unidad básica de codificación es el macro bloque, compuesto por 4 bloques de

luminancia de 8x8 pixeles y (en el caso del formato 4:2:0) de 2 bloques de

crominancia (un Cr y un Cb) de 8x8 pixeles que abarcan la misma zona de

la imagen. Todos los macro bloques de la imagen se codifican secuencialmente

de izquierda a derecha y de arriba abajo, eligiéndose un modo de codificación

independiente para cada uno de ellos.

Una vez que se ha elegido el modo de codificación, la predicción con compensación

de movimiento del contenido del bloque se hace a partir de la imagen de

referencia (I o P) pasada (caso de las imágenes P) y eventualmente futura (caso

de las imágenes B). La predicción se elimina de los datos reales del macro bloque,

lo que da la señal de error de predicción.

Al igual que MPEG-1, la norma no define explícitamente el método de codificacion

, sino ´únicamente la sintaxis que controla el tren binario a la salida del

codificador, lo cual deja gran libertad a su diseñador. Se describe ahora la codificacion

MPEG-2 bastante similar a la MPEG-1 ya comentada. A partir de la

imagen digitalizada en formato 4:2:0 (caso del main profile), el codificador elige

para cada imagen su tipo (I, P o B) y si esta debe ser codificada en modo frame

(imagen) o field (campo).

El codificador a continuación debe estimar los vectores de movimiento para

cada macro bloque de 16x16 pixeles. El número de vectores depende del tipo de

imagen y del modo de codificación escogido para cada bloque. En el caso más

general, donde el codificador es capaz de generar imágenes B (bidireccionales),

deberá reordenar las imágenes antes de la codificación y la transmisión.

La unidad básica de codificación es el macro bloque, compuesto por 4 bloques

de luminancia de 8x8 pixeles y (en el caso del formato 4:2:0) de 2 bloques de

crominancia (un Cr y un Cb) de 8x8 pixeles que abarcan la misma zona de

la imagen. Todos los macro bloques de la imagen se codifican secuencialmente

de izquierda a derecha y de arriba abajo, eligiéndose un modo de codificación

independiente para cada uno de ellos. Una vez que se ha elegido el modo de

codificación, la predicción con compensación de movimiento del contenido del

bloque se hace a partir de la imagen de referencia (I o P) pasada (caso de las

imágenes P) y eventualmente futura (caso de las imágenes B). La predicción

se elimina de los datos reales del macro bloque, lo que da la señal de error de

predicción.

Decodificación MPEG-2

La decodificación es más sencilla que la codificación, ya que no tiene que efectuar

alguna estimación de movimiento, que es una de las partes más complejas

del codificador. La memoria intermedia (bu\_er) de entrada recibe los datos del

canal de transmisión, y el decodificador lee el tren binario hasta encontrar el

principio de una imagen, su tipo (I, P o B) y su estructura (frame o field). Empieza

la decodificación con la primera imagen I, almacenándola en su memoria,

así como la imagen P siguiente, para servir de referencia a las imágenes P o B

que dependen de ella. Para las imágenes I, la decodificación propiamente dicha

consiste en aplicar a cada bloque la decodificación VLC, la decuantificacion de

los coeficientes y la transformación DTC inversa.

Para las imágenes P o B, este proceso consiste en construir la predicción de

cada macro bloque a partir de su tipo, de los vectores de movimiento y de las

imágenes de referencia memorizadas. El decodificador lee, decodifica y decuantifica

los coeficientes DTC del error de predicción transmitido para cada bloque

de 8x8 pixeles, y, después de la transformada DTC inversa, añade el resultado

a la predicción. Pág. 20/39 La reconstrucción de la imagen se efectúa cuando

todos los macro bloques han sido tratados.

La última etapa de la decodificación es poner las imágenes en el orden inicial

de visualización. Como se vio anteriormente, la necesidad de memoria para el

decodificador es de unas 3 imágenes (dos imágenes de referencia más la imagen

en via de reconstrucción), siendo para una imagen 4:2:0, de aproximadamente

16 Mbits.

Modos de predicción especificos MPEG-2 (Imágenes Entrelazadas)

Una diferencia que hay que destacar en MPEG-2 respecto a MPEG-1 es

que los slices no necesariamente abarcan toda la imagen, y además deben estar

compuestas ´únicamente de macro bloques contiguos situados en la misma linea

horizontal.

La principal novedad con respecto a MPEG-1, además de los perfiles y niveles,

provienen del tratamiento de las imágenes entrelazadas. Hay básicamente

dos modos de visualizar video: escaneo progresivo o escaneo entrelazado. El escaneo

progresivo visualiza todas las lineas horizontales de una sola vez como

si fuese un ´único cuadro. El escaneo entrelazado visualiza sólo la mitad de las

líneas horizontales de una vez/pasada (el primer campo que contiene todas las

líneas de número impar es visualizado y a continuación se visualiza el segundo

campo que contiene todas las líneas de número par). El entrelazado se basa en

una característica de nuestros ojos que es la persistencia de la visión (la cual

podria ´únicamente ser psicológica, no fisica), así como también en la persistencia

del fosforo en el tubo de imagen de la televisión para mezclar los campos resultando

en una ´única imagen. La ventaja del video entrelazado es que se pueden

obtener altas velocidades de refresco (50 o 60 Hz) con ´únicamente la mitad de

datos.

La desventaja es que la resolución horizontal es esencialmente dividida por

dos debido a que tiene que ser filtrada para evitar parpadeos y otros artefactos.

El tratamiento es diferente según la importancia de los movimientos entre los

dos campos de una misma imagen (los casos extremos son, por un lado, cuando

se transmiten peliculas cinematográficas por televisión (telecine) donde no hay

movimiento entre los dos campos de TV, puesto que proceden de la exploración

del mismo fotograma de la pelicula, y por otro lado, las imágenes de video de

acontecimientos deportivos, donde puede haber importantes movimientos entre

los dos campos de una imagen). La codificación progresiva, es apropiada para

los casos donde hay poco movimiento entre dos campos sucesivos. Los bloques

y macro bloques se dividen en la imagen completa, y la DCT se efectúa sobre

puntos verticales que distan 20 ms en el tiempo, lo que no plantea problemas si

los dos campos difieren poco. La entrelazada es preferible cuando el movimiento

de un campo a otro es importante. En este caso, a fin de evitar un contenido

en frecuencias verticales elevadas que reduciria la eficacia de la compresión tras

efectuar la DTC, la división de los macro bloques se hace considerando cada uno

de los campos como una imagen independiente en el interior del cual se toman

los bloques.

5.3.2. Niveles y perfiles MPEG-2

La información de brillo y color son tratadas de forma diferente por el sistema

visual humano, ya que es más sensible al brillo que al color. Con lo que se usa un

componente especial para representar la información del brillo, la luminancia,

una para el color y la saturación, la crominancia. Cada muestra de color se

codifica en señal Y-U-V (Iluminancia, U y V crominancia) partiendo de los

valores del sistema RGB. Con este sistema las diferencias de color pueden ser

muestreadas sin resultados visibles, lo que permite que la misma información

sea codificada con menos ancho de banda.

El esquema de compresión MPEG sirve para una amplia gama de aplicaciones.

La primer aproximación, MPEG1, se restringió al concepto de muestreado

4:2:0, explorado progresivo (no-entrelazado) y resolución SIF (352x240). El

MPEG-2 ofreció una gama más amplia de opciones de parámetros que producen

millones de posibles combinaciones. El concepto de los ”perfiles ”niveles “fue

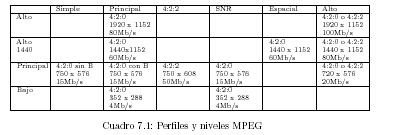
introducido luego para restringir la opción de parámetros con propósitos prácticos.

Las opciones ofrecidas permiten definición estándar (720x576 o 720x480)

así como formato HDTV norteamericano (1280x720P o 1920x1080I) o europeo

(1920x1152I o 1440x1152I). La tabla siguiente resume las combinaciones de niveles

y perfiles definidas:

****

Un perfil es básicamente el grado de complejidad esperada en la codificación,

mientras que un nivel describe el tamaño de la imagen, la resolución de esta o

la velocidad de transferencia de bit usada en ese perfil. En principio, hay 24

combinaciones posibles, pero no todas están definidas. Un codificador MPEG

cuando entrega un perfil y un nivel determinado, debe además ser capaz de

decodificarlo a perfiles y niveles inferiores. Un perfil simple no soporta una

codificación bidireccional y de este modo solo genera imágenes I y P. Esto reduce

la tasa de compresión simplificando el codificador y el decodificador; permitiendo

un sencillo hardware.

Un perfil simple solamente está definido en el nivel main (principal) como

(SP@ML). El perfil simple tiene una tasa máxima de 15 Mbps. El perfil main

(principal) corresponde actualmente al mejor compromiso entre calidad/tasa

de compresión, utilizando los tres tipos de imágenes (I, P y B), a costa de un

codificador y decodificador, más complejos.

El perfil principal soporta imágenes I ,P, B 4:2:0 y ofrece cuatro opciones

que van de SIF (352x488) a todos los modos a HDTV europeos (1920xl152). Se

diseño para una cantidad grande de usos. El más común es el MP@ML.

El perfil 4:2:2 principal es un nivel de calidad SDTV que permite procesamientos

futuros en estudio y completa compatibilidad con equipo de producción

digital (contribución).

Los perfiles escalables están previstos para operaciones posteriores y permitir

transmitir una imagen básica (base layer) en términos de resolución

espacial (spatially scalable profile) o de cuantificación (SNR scalable profile),

así como información suplementaria independiente (enhanced layer) que permite

mejorar sus caracteristicas, por ejemplo para transmitir la misma emisión en

definición estándar (SD) y HD (High Definition), o permitir una recepción con

calidad aceptable en caso de recepción dificil y de calidad ´optima en buenas

condiciones.

El perfil high (alto) soporta tanto el SNR y la escalabilidad espacial como

también la opción de muestreado 4:2:2. La especificación DVB requiere codificacion

4:2:0 lo que simplemente significa que la resolución de la información de

color tiene un cuarto de la resolución de la información de video. Los formatos

de muestreo 4:2:0 (4 muestras Y por cada muestra Cb y Cr) son adecuados

para aplicaciones de distribución de señal como direct-to-home (directo-a-casa)

o como distribución de la post-producción. Los formatos de muestreo 4:2:2 (4

muestras Y por cada 2 muestras Cb y 2 Cr) son adecuados para propósitos de

contribución con posibilidades de post producción. Con este sistema se dobla la

cantidad de información de color vertical que se transmite.

5.4. MPEG-4

Al principio, el estándar MPEG-4 se creó como un intento para mejorar la

calidad del video codificado de bajas velocidades a través de la estandarización

de nuevas técnicas mejoradas de compresión, orientado inicialmente a las videoconferencias

e Internet.. Más adelante, su progresión recondujo este estándar al

mundo de la TV interactiva, la computación y las telecomunicaciones.

El objetivo es crear un contexto audiovisual en el cual existen unas primitivas

llamadas AVO (objetos audiovisuales). Se definen métodos para codificar estas

primitivas que podrian clasificarse en texto y gráficos. Las nuevas características

ofrecidas por este estándar se pueden resumir en:

Las escenas se descomponen en 2 componentes básicas: audio y video.

Estos dos objetos son codificados de forma independiente.

Los objetos pueden ser tanto video natural (p.ej. generado por una cámara)

como imágenes sintéticas (generadas por un ordenador).

Ofrece soporte para manipulación de las imágenes sintéticas (soporte para

animación, utilización de imágenes estáticas 2D-3D como logos etc).

Permite interacción de los usuarios sobre la escena que se está renderizando.

Se ha mejorado la base del algoritmo MPEG para incrementar la robustez

para el trato de errores.

La comunicación con los datos de cada primitiva se realiza mediante uno

o varios .elementary streams.o flujos de datos, cuya característica principal es

la calidad de servicio requerida para la transmisión. Ha sido especialmente dise

˜nado para distribuir videos con elevadas tasas de compresión, sobre redes con

bajo ancho de banda manteniendo una excelente calidad para usuarios con buen

ancho de banda.

5.4.1. Codificación

MPEG-4 guarda muchas similitudes con el MPEG-1 y el MPEG-2, tal como

la compresión basada en la DCT (Discreet Cosine Transformation) con frames

I, P y B, todos dentro del GOPs. También tiene una serie de mejoras, especialmente

para bajos flujos de datos. Esto incluye mejor estimación de movimiento

y filtraje de desbloqueo. Su calidad y flujo de datos (20Kbps hasta 1000Kbps)

es enormemente mejor que en el MPEG-1 y, generalmente, más competitivo que

otras soluciones Web.

En efecto, el MPEG-4 ofrece mejores características a bajos flujos de datos,

tipicos de la web. A diferencia de otros codecs para la web, el MPEG-4 soporta

contenido entrelazado, resoluciones de hasta 4096 x 4096 y un flujo de datos entre

5Kbps y 10Mbps en la versión 1. Teóricamente, el MPEG-4 permite desde un

ancho de banda muy bajo (telefonia móvil) hasta la televisión en alta definición

(HDTV). Por supuesto, los dispositivos actuales no soportan la reproducción

de todo el rango de especificaciones pero, con el tiempo, se presentaran nuevos

equipos en el mercado.

El codec de video MPEG-4 soporta, nativamente, canal alfa, así se pueden

hacer composiciones de video sobre un fondo en tiempo real. Esto puede ser

usado para una segmentación, ya que es posible separar internamente el fondo

de la imagen sobre una escena. Esto es debido a la propia concepción del codec

MPEG, que extrae la imagen en movimiento (principal) de la fija (secundaria)

para realizar la compresión. Para comprender la segmentación, imaginemos

un video donde un señor está leyendo, mientras camina por una sala. Con un

codec convencional, cada vez que el señor va al principio de la sala y regresa, se

está comprimiendo (transmitiendo) toda esa información. Con la segmentación,

el codec puede recordar la ”imagen “de la sala una sola vez (fondo o background),

y comprimir (transmitir) el resto de la información, en este caso, el señor que

se pasea leyendo.

5.4.2. Versiones

A diferencia de otros formatos, el MPEG-4 tiene varias versiones. La versión

1, cuya especificación fue aprobada en octubre de 1.998; y la versión 2, aprobada

en diciembre de 1.999. La versión 2 añade una serie de mejoras y ampliaciones

al MPEG-4, pero la mayoria de las implementaciones soportan sólo la versión

1. Las funciones del estándar, están divididas en varios apartados que están soportados

por las distintas plataformas. Las mejoras proporcionan una serie de

apartados para diferentes mercados. Por ejemplo, los sistemas basados en telefonia

móvil no pueden soportar la reproducción de televisión en alta definición

(HDTV). La versión 1 tiene nueve perfiles de video y cuatro de audio, y la version

2 añade otros siete perfiles de video y cuatro de audio. Esto varia desde un

simple rectángulo de video con un simple flujo de audio (Simple Visual Profile)

hasta el total desarrollo de un fichero multimedia con las mejoras de la versión

2 (Advanced Coding E\_ciency). Generalmente solo se usará el modo principal

(Main Visual Profile).

5.4.3. Imágenes Fijas

El MPEG-4 incluye un codificador para imágenes fijas que se pueden usar

como fondos o mapas de texturas para objetos 3D. Comparado con el JPEG,

el compresor Wavelet ofrece mejor calidad al 25% del tamaño del archivo, ideal

para usos en la web. El compresor permite que el servidor reduzca dinámicamente

el tamaño del archivo (y la Pág. 34/39 calidad) para anchos de banda

muy bajos, reduciendo la necesidad de que el autor realice varias versiones para

distintas velocidades de conexión.

5.4.4. Objetos Multimedia

El MPEG-4 gana mucho en objetos multimedia. En lugar de definir el video

como un flujo de muestras de audio y cuadros de video, podemos construir

objetos mediáticos como flujos estándar de audio y video, imágenes fijas, textos,

voz sintetizada, modelos tridimensionales 3D, etc. En lugar de mostrar un video

en un rectángulo, esos objetos multimedia pueden estar mapeados en los objetos

de una escena. Por ejemplo, el audio puede estar mapeado en un objeto y este,

a su vez, ocupar una posición en un espacio 3D. En otras palabras, el MPEG-4

es capaz de mezclar lo mejor de Shockwave, Flash, VRML, y video digital en un

simple archivo, servidor y reproductor. Esto permite la creación de complejas

creaciones o presentaciones de objetos multimedia que pueden ser vistas sobre

conexiones de baja velocidad.

5.4.5. Aplicaciones

La importancia del MPEG-4 se ve en la cantidad de aplicaciones que se han

anunciado a propósito de ´el. La industria del cable ha anunciado un intento de

sustituir la actual implementación del MPEG-2, por el MPEG-4 en un futuro no

muy lejano. Esta es una decisión obvia, ya que el MPEG-4 permitirá a la industria

duplicar o triplicar el número de canales disponibles sobre el ancho de banda

existente, al igual que permitir la televisión interactiva (ITV) y el video bajo

demanda (VOD). Como se puede ver, gracias al MPEG-4 se abre un gran abanico

de posibilidades que irán creciendo a medida que se vayan creando nuevas

herramientas para la creación y distribución de aplicaciones multimedia. Otra

característica del MPEG-4 es el MPEG-J, una libreria de Java para controlar el

MPEG-4. Combinando ambos, es posible crear aplicaciones de la complejidad

de Java dentro del entorno del MPEG-4 y, aún más complejo, el propio applet de

Java puede estar ”incrustado”dentro del flujo MPEG. Esto permitirá sistemas

de comunicación de una via, como el cable, que se podrian implementar en un

receptor de cable, permitiendo sistemas de navegación, publicidad interactiva, la

tienda en casa, resultados deportivos y un sin fin de posibilidades. Otros proponen

la idea de sustituir las emisiones de radio en AM, con flujos de audio de alta

calidad MPEG-4, con datos incluidos (tipo RDS en la radio FM convencional)

al igual que flujos de video para equipos móviles, videoconferencia, sistemas de

seguridad CCTV, estudios de video portátiles, etc. Es dificil predecir el futuro

de cual Pág. 35/39 será ´útil y cual no, pero lo que si está claro es que gracias

al MPEG-4 tendremos muchas soluciones. Aunque RealNetworks, Microsoft y

Apple mantienen sus propias tecnologias de codificación de video para Internet,

seguirán proliferando las alternativas que utilizan MPEG-4, ya que ´este será el

estándar más utilizado para la distribución e intercambio de video bajo demanda

en Internet, según un informe de Gartner. Según el informe, MPEG-4 será el

estándar dominante en compresión de video y distribución sobre Internet para

el año 2005, debido principalmente a su bajo coste, su alta velocidad de reproduccion

, y la paulatina introducción de la banda ancha, que permitirá mayor

calidad en los videos bajo demanda y mayor velocidad de descarga.