

COMPRESIÓN DE VIDEO MPEG-ITU

Análisis de los métodos de codificación de vídeo mediante las técnicas ADPCM, MPEG y las normas del ITU-T.

1- DIGITALIZACIÓN BÁSICA DE VIDEO

En este ítem se analiza la digitalización de vídeo y la compresión a alta velocidad. Este tema se completa más adelante con los mecanismos de compresión avanzada de baja velocidad.

1.1- DIGITALIZACIÓN

El proceso de digitalización de la señal de vídeo sigue los mismos pasos que en el caso de una señal de telefonía: muestreo, cuantificación y codificación. En **ITU-R Rc.601-1** se determina el método a seguir en estos casos. El método desarrollado **CMTT/2** del CCITT y CCIR permite la codificación de alta calidad en 34 y 140 Mb/s.

Se disponen de 3 componentes a digitalizar: una de luminancia (o brillo) y dos de crominancia (o color). La recomendación tiene en cuenta una frecuencia de muestreo de 13,5 MHz para la componente de luminancia y de 6,75 MHz para las componentes de crominancia. El formato de la señal se dice entonces que es del tipo 4:2:2 que corresponde al factor que multiplica a la frecuencia de referencia de 3,375 MHz.

Con estas frecuencias de muestreo el número de muestras (**pixel**) por cada línea de la imagen es de 864 para la componente de luminancia y de 432 para la de crominancia. Sin embargo, como cada línea y cuadro tiene un tiempo de retroceso horizontal y vertical respectivamente, que no aporta información de imagen, el número efectivo de muestras por línea es de 720 para la luminancia y de 360 para la crominancia.

De esta forma se produce la primer etapa de reducción de velocidad mediante la eliminación de los pixel sin información de imagen. Cada muestra o pixel de la imagen se codifica en 8 bits. Por lo tanto, se obtiene una velocidad digital de $216 \text{ Mb/s} = 8 \text{ bit} \times (13,5 + 2 \times 6,75) \text{ MHz}$. De los cuales solo 166 Mb/s corresponden a la parte visible de la imagen.

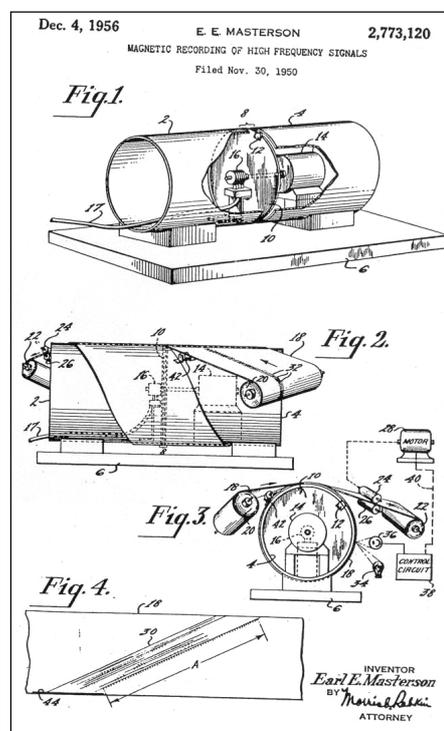
En la fotografía anexa se muestra el primer diseño del registrador de señal de vídeo en cinta magnética de RCA-1950.

1.2- REDUCCIÓN DE VELOCIDAD

La velocidad obtenida es relativamente alta y desde la década de los años 70 se han intentado diversas soluciones para reducir la velocidad manteniendo la calidad. Los métodos genéricos para eliminar la correlación entre porciones de imágenes y redundancias temporales, y de esta forma producir una reducción de la velocidad **CDV** (*Compressed Digital Video*) son:

- Supresión de muestras en borrado de líneas y cuadros.
- Interpolación de muestras, líneas o cuadros. Uso de Transformada Compensación de Movimiento **MCT**.
- Codificación Adaptativa Diferencial PCM **ADPCM** en forma aislada o en conjunto con la MCT.
- Codificación mediante la Transformada Discreta Coseno **DCT**.
- Codificación de Longitud Variable **VLC** (Huffman).

La supresión de aquellas muestras que corresponden al retrasado horizontal y vertical de la imagen se realiza siempre en todos los equipos codificadores digitales. Cuando se trata de obtener una reducción sustancial de la velocidad se recurre a la interpolación de muestras, líneas o cuadros (generalmente más en crominancia que en luminancia). Se recomendó en **ETSI 300-174** (antes CEPT) los algoritmos de compresión para trabajar a 34 o 45 Mb/s. Se trata de una codificación que combina ADPCM con DCT para eliminar la redundancia temporal y espacial asociada a la codificación de longitud variable adaptativa.



COMPRESIÓN DE VIDEO MPEG-ITU

TRANSFORMADA DISCRETA COSENO DCT. La codificación mediante la **DCT** (*Discrete Cosine Transform*) ha sido posible en los últimos años de la década de los 80 gracias a los circuitos integrados de muy alta densidad. En las investigaciones teóricas llevadas a término en la década de los años 70 se determinó la factibilidad de este método de compresión. Se investigaron las transformadas **Fourier DFT** y **Slant DST**.

La transformada de **Walsh-Hadamard DWHT** (funciones Walsh) es muy simple y solo requiere operaciones de suma y resta de funciones ortogonales. En cambio la transformada discreta coseno requiere un tiempo de cálculo y un número de compuertas inferior y proporciona una compresión de bits mayor.

DCT permite reducir la tasa de codificación y la velocidad de transmisión. Se divide la imagen en sectores de 8x8 pixels, que denominamos X_{ij} . Los valores de la matriz de 8x8 pixels luego de la DCT se denominan T_{mn} y se obtienen mediante el cálculo:

$$T_{mn} = (E_m \cdot E_n / 4) \cdot \sum_i \frac{[\cos(2i+1) \cdot n \cdot \pi]}{16} \cdot \sum_j X_{ij} \cdot \frac{[\cos(2j+1) \cdot m \cdot \pi]}{16}$$

Donde $E_n, E_m = 1/\sqrt{2}$ para $n, m = 0$
 y $E_n, E_m = 1$ para $n, m > 0$.
 Los valores de (i, j) y (m, n) varían entre 0 y 7.

Para una matriz de 4x4 pixels (por simplicidad) se tiene el siguiente ejemplo:

24	23	25	31	DCT	132	-10	-0,2	8,1
25	34	23	25		-6,1	0,1	0,4	7,2
31	24	30	20		0,4	0,2	-0,1	-0,2
34	24	25	31		3,2	-0,1	0,1	0,3

OBSERVACIONES:

- El método de codificación DCT convierte una matriz 8x8 puntos X_{ij} en otra T_{mn} .
- Cada punto de la matriz T_{mn} se calcula en base a toda la matriz de puntos X_{ij} .
- En la matriz transformada T_{mn} se encuentran los valores mucho más correlacionados que en X_{ij} .
- Esto permite una codificación final con menor número de bits.
- La matriz 8x8 resulta ser el valor óptimo como velocidad de cálculo y des-correlación de coeficientes.

- De esta forma, un error en la transmisión de los datos T_{mn} se difunde en toda la matriz X_{ij} de recepción.
- En la codificación ADPCM los errores se observan como un granulado de puntos en la pantalla.
- En la DCT los errores se distribuyen en la matriz de 8x8 pixel y son prácticamente imperceptibles a la visión humana.
- DCT requiere una capacidad de cálculo superior a ADPCM y produce un retardo también mayor.
- En transmisiones de baja calidad (exceso de reducción de velocidad o mucha BER) se observan distorsiones por rectángulos (el rectángulo es el resultado de la matriz 8x8, o múltiplos de ella, distribuida con la diferencia de definición entre puntos horizontales y líneas verticales).

- DCT se fundamenta en una representación similar a la frecuencia.
- Se denominan frecuencias normalizadas o coeficientes de transformada.
- El ojo humano es sensible en distinta forma a los coeficientes y por ello pueden ser tratados de distinta manera.
- Los coeficientes se encuentran des-correlacionados y solo unos pocos (superior-izquierdo) son importantes.
- El coeficiente superior-izquierdo (132) mide la energía del término de frecuencia cero (corriente continua **Dc**).
- Por ejemplo, si el bloque de 4x4 tiene un valor constante, solo el primer coeficiente DCT tiene un valor distinto de cero.
- Los otros coeficientes representan las frecuencias (corriente alterna **Ac**), horizontalmente y verticalmente.
- DCT transforma los coeficientes de amplitud de intensidad de pixel en coeficientes de transformadas de frecuencias.

- Los coeficientes de la matriz se leen en zig-zag oblicuo desde el ángulo superior-izquierdo (132/-10/-6,1/ 0,4/0,1/-0,2/...).
- Solo se transmiten los coeficientes que superan cierto umbral.
- El buffer de salida selecciona el proceso de cuantificación para mantener una velocidad constante y reducir el retardo.

CODIFICADOR VLC: La codificación de longitud variable **VLC** (código de Huffman) consiste en asignar a las muestras una longitud de bits de acuerdo con la probabilidad de ocurrencia. En otras palabras, se logra reducir la tasa de bits si a las muestras o pixel de mayor probabilidad de ocurrencia se le asignan un número de bits inferiores que a las muestras de menor probabilidad.

Por ejemplo, se supone que se deben codificar 8 niveles de cuantificación (a1...a8). Cada uno de éstos tiene una probabilidad de ocurrencia (Pa1...Pa8). La codificación de longitud constante requiere 3 bits para codificar los 8 niveles de cuantificación;

COMPRESIÓN DE VIDEO MPEG-ITU

mientras que la codificación de longitud variable requiere menor cantidad en promedio. Se puede obtener la eficiencia del código expresada en términos del número promedio de bits.

En el caso de la codificación de longitud constante el número promedio de bits es 3; mientras que en el caso de longitud variable es 2,65 bits por código. Para obtener este valor se debe realizar la sumatoria de los productos entre el número de bits y la probabilidad de ocurrencia de cada evento.

Para el caso de la señal de vídeo el codificador de longitud variable permite convertir la longitud constante de 4 bits en una longitud variable de 1 a 12 bits. Entonces permite una reducción de velocidad promedio de 4 a 2,4 bit/pixel. Muchos de estas formas para reducir la tasa de bits transmitida se suelen utilizar en conjunto. En los ítems siguientes se mencionan tanto el funcionamiento de cada método como ejemplos de aplicación en conjunto. Como ejemplo se toman equipos comerciales que reducen la velocidad a 34 Mb/s. Existen otros que trabajan a 140 Mb/s con mejor calidad a costa de una mayor velocidad de transmisión. Por otro lado se mencionan los formatos de alta definición **HDTV** y los sistema de baja velocidad **JPEG**, **MPEG** y **H.261**.

COMPRESIÓN DE VIDEO MPEG-ITU

2- RED DE TRANSPORTE A 34 Mb/s

Los equipos para transporte de vídeo primero a 140 Mb/s (primera parte de la década de los 80) y luego a 34 Mb/s (finales de los 80 e inicio de los 90) utilizaban métodos de compresión no normalizados (propietarios de los productores). Se refiere entonces a equipos que dominaron la escena en la década de los años '80.

Hacia mediados de la década de los 90 el proceso MPEG comienza a dominar el mercado de transporte digital de vídeo. Reduce sustancialmente la velocidad aplicando nuevas técnicas de compresión normalizadas. Este ítem se refiere a equipos anteriores a las normas MPEG; las que son detalladas más adelante.

2.1- VERSIÓN CON CODIFICADOR ADPCM

La codificación Diferencial PCM Adaptativa **ADPCM** es una técnica utilizada inicialmente por su simplicidad. En la **Fig 01** se muestra un diagrama a bloques de un codificador digital, para señal de vídeo de calidad comercial a 34 Mb/s (marca *Siemens* según fotografía), que utiliza esta técnica. Los pasos sucesivos del procesamiento son:

1- INTERFAZ DE ENTRADA: Se ingresa la señal compuesta de vídeo con un ancho de banda de 0 a 6 MHz. En la frecuencia de 4,43 MHz para el sistema PAL se encuentra la subportadora de crominancia. Esta interfaz separa la señal de luminancia Y y las componentes de crominancia (denominadas aquí CB y CR). Si se trata de otro sistema (SECAM o NTSC) el procedimiento es similar respetando la forma de demodulación para separar las componentes de vídeo. La misma interfaz entrega la frecuencia de línea que permite sincronizar a una fuente de temporización (base de tiempo). Esta base de tiempo organiza el procesamiento hasta la última etapa de multiplexación.

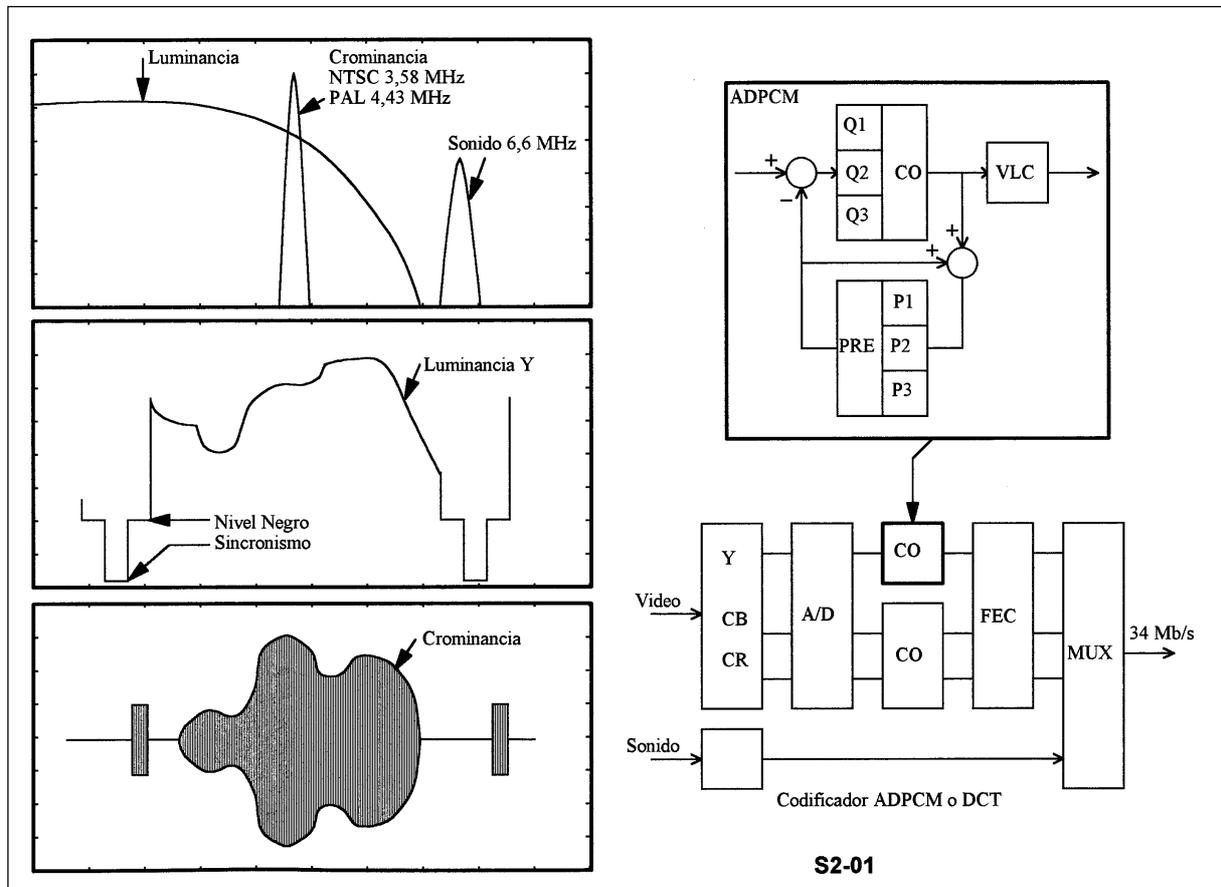
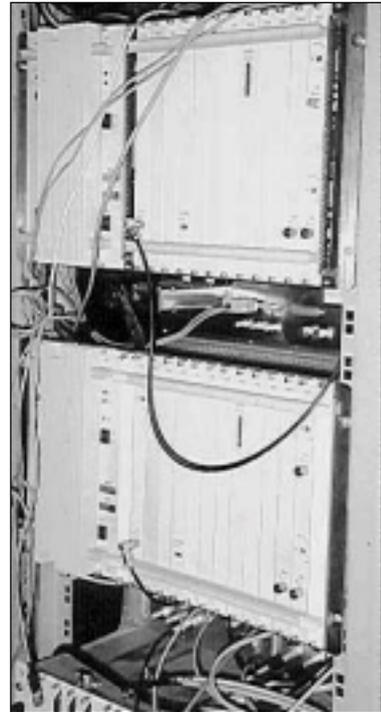


Fig 01. Codificación de vídeo ADPCM/DCT a 34 Mb/s.

COMPRESIÓN DE VIDEO MPEG-ITU

2- CONVERSOR ANALÓGICO-DIGITAL A/D: Se muestrean las componentes de luminancia y crominancia. De acuerdo con **ITU-R Rc.601-1** las frecuencias de muestreo son 13,5 y 6,75 MHz, respectivamente. En esta etapa también se procede a aplicar el primer estadio de reducción de velocidad. Se trata de un muestreo *sample & hold* con codificación en 8 bits lineal sin compresión por pixel.

3- INTERPOLACIÓN: Permite transmitir la mitad de muestras o pixel de crominancia respecto de las de luminancia (lo cual queda en evidencia a partir de las frecuencias de muestreo). Además, solo se transmiten las muestras de CB y CR en cuadros alternados. Existen diversas formas de interpolar, las que, al no encontrarse uniformadas internacionalmente, responden solo a la iniciativa de los diseñadores.

4- CODIFICADOR ADPCM: Codifica la diferencia entre el valor de muestra y una predicción calculada en base a datos anteriores. **J.Max-1960** determinó que el valor diferencia tiene una distribución Poisson. A partir de ella determinó los puntos de cuantificación óptimos desde el punto de vista teórico. Sin embargo, en la práctica no resultaron ser los mejores. **D.Sharma y A.Netravalli-1977** determinan el método subjetivo desde el punto de vista del observador por el cual se selecciona la posición de los umbrales de codificación. El método de determinación de los niveles de cuantificación de acuerdo con la impresión subjetiva del observador consiste:

- Se determina el nivel de ruido de cuantificación N_q que comienza a ser visible para cada valor de diferencia.
- La diferencia es entre la muestra y la predicción.
- Se distribuyen los niveles de cuantificación de forma que la separación entre ellos sea inferior al nivel de N_q .
- En otras palabras, en el entorno de dos niveles el N_q es inferior al subjetivamente visible.

En la **Fig 01/02** se muestra un ejemplo de este tipo de codificación. La predicción puede ser del tipo **unidimensional** (toma en cuenta solo los puntos sobre la misma línea) o **bidimensional**. En este último caso se puede tener una **predicción intratrama** (teniendo en cuenta puntos de un mismo campo; predicción espacial) o **intertrama** (teniendo en cuenta algunos puntos del cuadro anterior; predicción temporal).

5- PREDICTORES: Se cambian en forma automática de acuerdo con la **dinámica de la imagen**. Por ejemplo, podemos tener los siguientes predictores para el punto de imagen X:

$$\begin{aligned} P1 &= A && \text{unidimensional} \\ P2 &= a.A + b.B + c.C + d.D && \text{bidimensional intratrama} \\ P3 &= X' && \text{bidimensional intertrama} \end{aligned}$$

El predictor P1 se utiliza como referencia y se hace uso de él cada un número determinado de líneas (cada 16) y cuadros (cada 16) para evitar la propagación de errores debida a errores del predictor. Los predictores P2 y P3 se intercambian de acuerdo con el tipo de imagen.

Para ello, se define la diferencia máxima entre puntos M1 (parámetro de actividad espacial intratrama) y M2 (parámetro de actividad temporal intertrama):

$$\begin{aligned} M1 &= \text{máximo (A-B, A-C, A-D, B-C, B-D, C-D)} \\ M2 &= \text{máximo (A-A', B-B', C-C', D-D')} \end{aligned}$$

Se selecciona P2 si $M1 > M2$ y P3 si $M1 < M2$. Es decir, si la actividad espacial en el mismo cuadro supera a la actividad temporal entre cuadros sucesivos se utiliza el predictor P2.

6- CUANTIFICADOR: El codec para el ADPCM es del tipo adaptativo. Se disponen de 4 cuantificadores con salida de 4 bits. La distribución de los niveles de cuantificación depende de las características de la señal. Para seleccionarlos se define un valor M como el mínimo entre M1 y M2. Se determinan 2 umbrales S0 y S1 los cuales señalan el cuantificador a utilizar:

$$\begin{aligned} M < S0 & \text{ se usa Q1 (para pequeños cambios de imagen)} \\ S0 < M < S1 & \text{ se usa Q2 (intermedio entre Q1 y Q3)} \\ S1 < M & \text{ se usa Q3 (para contornos o contrastes)} \end{aligned}$$

Un cuarto cuantificador Q4 se utiliza para las líneas y cuadros de referencia (cada 16) junto con el predictor P1.

7- CODIFICADOR VLC: La codificación de longitud variable **VLC** (código de Huffman) consiste en asignar a las muestras una longitud de bits de acuerdo con la probabilidad de ocurrencia. En otras palabras, se logra reducir la tasa de bits si a las muestras o pixel de imagen de mayor probabilidad de ocurrencia se le asignan un número de bits inferiores que a las muestras de menor probabilidad. Para el caso de la señal de vídeo el codificador de longitud variable permite convertir la longitud constante de 4 bits en una longitud variable de 1 a 12 bits. Entonces permite una reducción de velocidad promedio de 4 a 2,4 bit/pixel.

8- CODIFICADOR FEC: Por ejemplo, BCH(248,234). Permite la corrección de errores en el receptor cuando el número de ellos no supera una ráfaga de algunos pocos bits errados (4 en este caso).

9- CODIFICADOR DE SONIDO: Responde a la recomendación **ITU-T J.41** para el conversor A/D. Se tienen canales de sonido con una banda desde 40 Hz a 15 kHz muestreados a una frecuencia de 32 kHz. La conversión analógica-digital se

COMPRESIÓN DE VIDEO MPEG-ITU

reduce de 14 bits lineales a 11 bit comprimidos (ley A) más un bit de paridad. La velocidad final es de 384 kb/s para cada canal de sonido. Para permitir la operación en estéreo se debe asegurar el muestreo sincrónico. El bit de paridad permite la repetición de la muestra anterior cuando existen errores aislados y la supresión de la señal si se detectan 3 de 4 muestras con error seguidas.

10- MULTIPLEXOR DE 33 Mb/s: Permite reunir las distintas señales en una trama. En este caso la trama contiene 2112 bits y una velocidad final de 33.792 kb/s. La trama está distribuida de la siguiente forma:

- 10 bits para el alineamiento de trama y 1 bit final para paridad en la trama.
- 1984 bits para la señal de vídeo (luminancia y crominancia).
- 96 bits para 4 canales de sonido y 3 bits para sincronismo de línea de vídeo;
- 14 bits para datos (teletex) a 224 kb/s y 4 bits para datos a 64 kb/s;

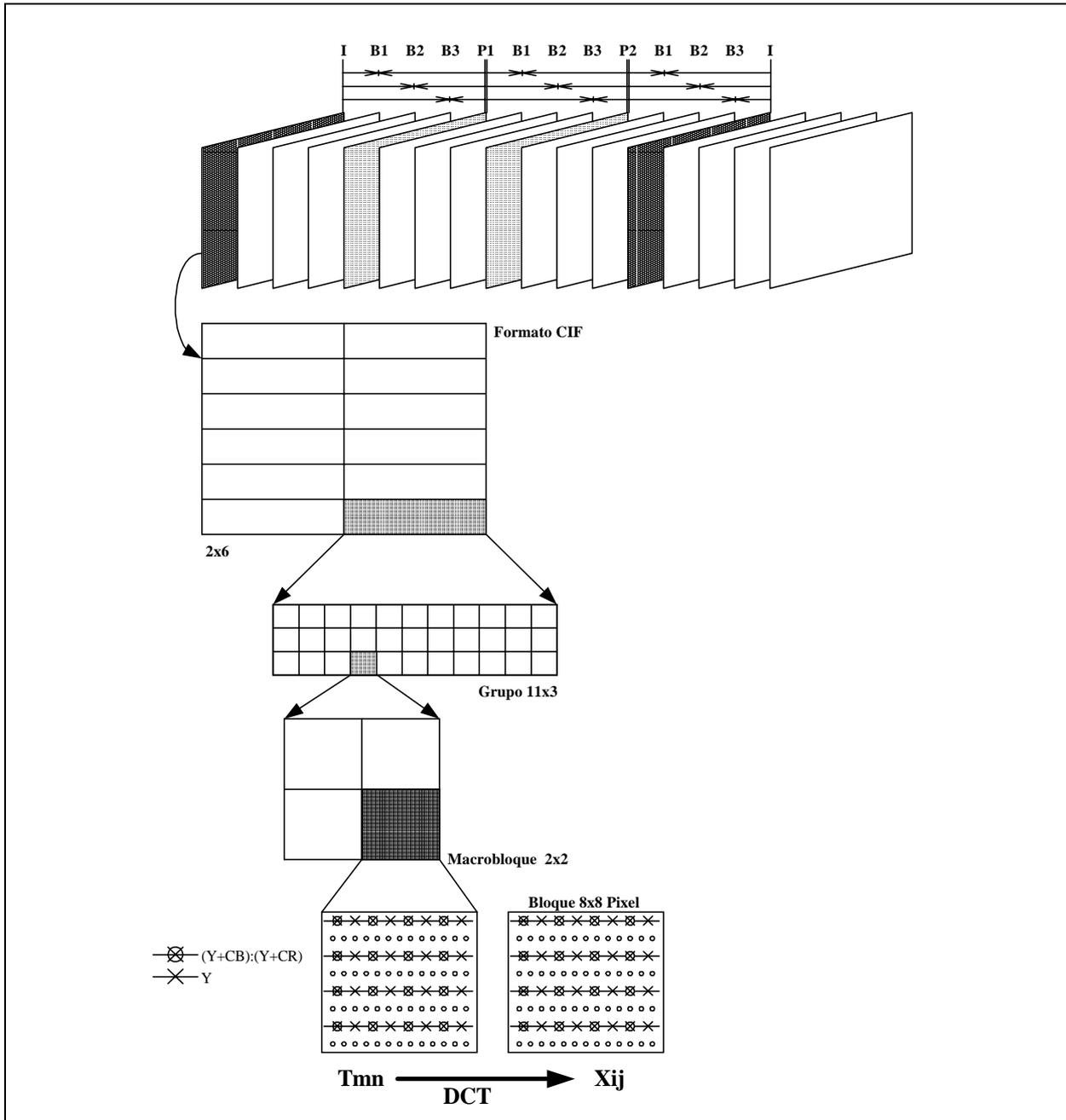


Fig 02. Interpolación, formato GIF y transformada DCT en MPEG.

COMPRESIÓN DE VIDEO MPEG-ITU

11- MULTIPLEXOR DE 34 Mb/s: Téngase en cuenta que la base de tiempo del equipo se sincroniza con la frecuencia de línea de la señal de vídeo. La salida de 34 Mb/s en cambio, se sincroniza con la red de transmisión. Permite la operación plesiócrona del codificador de 33 Mb/s con respecto a la velocidad de 34.368 kb/s de la red de transmisión. Para ello se recurre al proceso de justificación mediante una trama de 1536 bits con la siguiente distribución:

- 10 bits para el alineamiento de trama y 1512 bits para la trama de 33 Mb/s;
- 1 bit para alarma remota; 3 bits reservados.
- 10 bits para control de justificación y 2 bits para justificación positiva.

2.2- VERSIÓN CON CODIFICADOR DCT

En la **Fig 01** se observa un diagrama general donde el CO-DEC es del tipo DCT en lugar de ADPCM. Se trata de un equipo comercial de alta calidad que funciona a 34 Mb/s (marca *Alcatel-Telettra*). Los pasos a desarrollar son:

1- INTERFAZ Y CONVERTOR. Con características similares al equipo anterior.

2- INTERPOLACIÓN: Este método permite una reducción a la mitad de los pixels a codificar. Supongamos tener 4 muestras contenidas en una línea y denominadas S1.S2.S3.S4. En base a ellas se obtienen los siguientes pixel de imagen de luminancia y crominancia:

$$Y1 = S1 + S3 ; \quad Y2 = S2 + S4;$$

$$C1 = S1 - S3 ; \quad C2 = S2 - S4$$

3- TRANSFORMADA DISCRETA COSENO DCT: Permite reducir la tasa de codificación y la velocidad de transmisión. Se divide la imagen en sectores de 8x8 pixels, que denominamos Xij. Los valores de la matriz de 8x8 pixels luego de la DCT se denominan Tmn. El método de codificación DCT convierte una matriz 8x8 puntos Xij en otra Tmn. Como cada punto de la matriz Tmn se calcula en base a toda la matriz de puntos Xij, en Tmn se encuentran los valores mucho más correlacionados que en Xij. Esto permite una codificación final con menor número de bits (reducción de 8 a 4 bits por muestra).

4- CODIFICADOR VLC (HUFFMAN): Permite una reducción adicional de la velocidad promedio de transmisión, siguiendo los lineamientos comentados anteriormente.

5- CODIFICACION FEC: Permite corregir errores. En particular se usa el código del tipo BCH (511/493), que permite corregir 2 errores por bloque con una relación de tasa de error corregida $BER_c = 500 \cdot BER_{nc}^3$ (BER_{nc} es la BER antes de la corrección).

6- SONIDO Y MULTIPLEXOR. Con características similares al ítem anterior.

COMPRESIÓN DE VIDEO MPEG-ITU

3- COMPRESION DE IMÁGENES: MPEG

Este ítem complementa la descripción de los métodos de compresión de imágenes. Los mismos son aplicados en variados servicios, desde la TV digital hasta la videoconferencia por Internet. La compresión de imágenes puede ser del tipo sin pérdida de información (*STAC* o *Double Space* usados para expandir la capacidad del *Hard Disk* o *PKZip* para datos en el *Floppy Driver*). O con pérdida de información (JPEG y MPEG)

3.1- CLASIFICACION DE LOS STANDARD

Para aplicaciones de transmisión por enlaces terrestres (radioenlace y fibra óptica) se utilizan codificadores de 34-140 Mb/s. Para enlaces satelitales se recurre a velocidades substancialmente menores (típico entre 1,5 y 6,6 Mb/s), aplicando con mayor énfasis las mismas técnicas ya mencionadas. Las velocidades inferiores admiten varias técnicas alternativas. Desde 1988, el entonces CCITT y la ISO, iniciaron los esfuerzos para obtener normas para compresión de imágenes. El elenco de normas se dispone en la **Tabla 01**.

Tabla 01. Formatos de compresión de imágenes fijas y en movimiento.

JPEG	(<i>Joint Photographic Experts Group</i>) para imágenes fijas (fotos, gráficos, rayos X) según ISO-10918. Este tipo de formato (.jpg) permite una sustancial reducción de datos respecto a otros tipos de formatos. Una fotografía típica de este libro ocupa 200 kb/s en formato BMP (<i>Bit MaP</i> con extensión .bmp) o TIF (<i>Target Image Format</i> con extensión .tif) y solo 15 kb/s en JPEG (.jpg) o en GIF (<i>Graphic Interchange Format</i>). JPEG es similar a MPEG pero no utiliza la correlación entre imágenes (compensación de movimiento); por ello es posible su aplicación para imágenes en movimiento de muy baja velocidad. La calidad de codificación es mejor que MPEG.
MPEG	(<i>Motion Picture EG</i>) se trata de imágenes en movimiento para aplicaciones en multimedia y vídeo comercial; utiliza la compresión por correlación entre tramas (compensación de movimiento). Se disponen de las siguientes variantes: MPEG-1. Nace en 1988 para estandarizar videoconferencia a 1,5 y 2 Mb/s, según ISO-11172 del año 1992. Se aplica a grabaciones en CD-ROM con formato (.mpg). Un tipo de formato distinto es el (.avi) de Microsoft. El MPEG-1 se aplica en ITU-T H.261 y ITU-R Rec.723. MPEG-2. Nace en 1990 para codificar la norma ITU-R Rec.601. Se aplica en vídeo digital comprimido en broadcasting (estándar actual desde 3 a 10 Mb/s). Responde a la norma ISO-13818. MPEG-3. Iniciada en 1992 para aplicaciones de HDTV; luego fue abandonada. MPEG-4. Se inicia en 1994. Originalmente para imágenes de muy baja velocidad, en la actualidad funciona de modelo para cualquier tipo de aplicación futura. Es de interés en sistemas de baja calidad como en los sistemas móviles. Se estima el uso a velocidades hasta 64 kb/s para sistemas móviles o la red PSTN y hasta 2 Mb/s para TV/film. MPEG-7. Normalizado a partir de 1996 se aplicará para funciones de Interfaz para Descripción de Contenidos de Multimedia (catálogos de imágenes, páginas amarillas, radios y canales de TV, etc).
DTV-ATSC	(<i>Digital Television-Advanced TSC</i>). Este es el formato normalizado por la FCC para la distribución broadcasting de TV que reemplaza a NTSC en el período 1998 hasta 2006.
DVB	(<i>Digital Video Broadcasting</i>). El DVB es la norma de Europa para aplicaciones similares a ATSC. El proyecto se inició en 1993 y reemplazó al Eureka. Se dispone de formatos para DVB-S(atélite), DVB-C(able) y DVB-T(terrestre). Utiliza anchos de banda de 6 y 8 MHz. Como se ha analizado en detalle la norma ATSC en este caso la descripción se limita a señalar las similitudes y diferencias.
H.320	Norma ITU-T para circuitos de vídeo teléfono y teleconferencia a Px64 kb/s (interfaz abierta).
H.32x	Norma del ITU-T para medios LAN y ATM interactivos, similar a MPEG-2. La codificación de vídeo es H.262.
H.324 POTS.	La videoconferencia para las operadoras POTS utiliza la norma H.324 a 28,8 kb/s con módem V.34; el audio se codifica mediante G.723 y el vídeo mediante H.263/261. El multiplexor es H.223. La norma G.723 trabaja desde 5,3 a 6,3 kb/s con un ajuste dinámico dependiendo de la actividad vocal. Se fundamenta en un algoritmo LPC con tramas de 30 mseg (4x7,5 mseg) utilizando el codec ACELP (2 codebook fijo y adaptativo). La norma H.263 trabaja con sub-QCIF y QCIF.
AVI	(<i>Audio-Video Interleaver</i>) diseñado por Microsoft para CD-ROM con pocos cuadros por segundo y baja calidad.
DVI	(<i>Digital Video Interactive</i>), similar a AVI pero diseñado por Intel.

COMPRESIÓN DE VIDEO MPEG-ITU

NORMA JPEG.

Desde 1967 la ISO ha buscado un algoritmo común para compresión de imágenes fijas. En la actualidad el algoritmo JPEG lleva el número ISO-10918. El concepto es dividir la imagen en bloques de 8x8 pixel y comprimir los datos para reducir la capacidad de memoria de almacenamiento necesaria.

En JPEG se aplican las siguientes técnicas generales:

-La digitalización se realiza mediante un scanner en bloques de imágenes de 8x8 pixel. La reducción de redundancia utiliza la descorrelación de datos mediante el uso de una transformada reversible DCT. Se aprovecha también la reducción y codificación de la entropía para minimizar el número promedio de bits de codificación (codificación Huffman).

-La imagen que se divide en bloques de 8x8 pixel se digitaliza en una componente de luminancia de 16 niveles por pixel. Se codifica la matriz de datos en **DCT** obteniendo coeficientes de frecuencia. Esta matriz genera 2 secuencias de coeficientes **Ac** y **Dc**. La matriz se lee en zig-zag para reordenarla en forma lineal desde el extremo superior izquierdo. Se determinan entonces coeficientes **Ac** decrecientes. Permite ordenarlos en forma decreciente de importancia. La lectura en zig-zag de los coeficientes DCT se aplica también en MPEG y las normas derivadas.

-Se codifica a continuación en forma diferencial (coeficientes **Dc** respecto a la matriz anterior). Los coeficientes se dividen por 10 y se aproxima al entero más cercano. Por último se efectúa la codificación VLC (Huffman) y se obtiene de esta forma la señal definitiva.

NORMA MPEG.

Este estándar ISO es del año 1990 y descansa en la reducción de la redundancia temporal entre tramas (**MCT**) y la redundancia espacial en la misma trama (**DCT**). En ITU-T se aplica el mismo concepto que en ISO. Se disponen las variantes MPEG-1 (ISO/IEC 11172) y MPEG-2 (ISO/IEC 13118). El MPEG-3 para HDTV ha perdido interés en la actualidad y emerge a partir de 1999 para menores velocidades MPEG-4. MPEG sirve como referencia para las normas I-ITU y DTV de la FCC.

Las principales características son:

-La variante denominada **MPEG-1** con velocidad de 1,15 y 1,5 Mb/s es aplicable para video-game y CD-ROM. La resolución (líneas/cuadros) es de 240/30 para NTSC; 288/25 para PAL y Secam.

-En **MPEG-2** se tiene en cuenta aplicaciones de calidad VHS, PAL y HDTV. Se tiene previsto el funcionamiento en multimedia **ATM** con celdas de 2 prioridades. El funcionamiento es aceptable aun con tasas de pérdida de celdas de 10^{-4} .

-MPEG utiliza un cuadro de tipo **CIF** (H.261 en la **Fig 03**) de 352 x 288 pixel. Para más detalles ver la descripción más adelante. La crominancia tiene la mitad de pixel y de líneas (176 x 144). Se efectúa un sub-muestreo de tipo (4:2:0) con refresco de cuadros de 25 Hz. Se utilizan imágenes predictivas que incrementan en el retardo de codificación (600 mseg).

-MPEG aplica **DCT**, **VLC** y la compensación de movimiento **MCT** (**Fig 03**). Cada macrobloque puede tener predicción: *Intra*, *Forward-Predicted*, *Backward-Predicted*, *Average*. El codificador DCT tiene entrada de niveles (-255,+255) y salida (-2047,+2047). MCT se aplica sobre los macrobloques en tanto que DCT se aplica sobre los bloques. La salida del codificador se aproxima con una tasa 1/10.

EJEMPLO DE APLICACIÓN. En las conexiones de TV mediante satélite se ha difundido la codificación digital de baja velocidad mediante algoritmos derivados de MPEG. Generalmente mediante una codificación propietaria del fabricante se puede encontrar codificadores a 6,6 o 3,3 Mb/s. En las aplicaciones comerciales se encuentran variantes significativas. Una de ellas adopta un cuadro de 480 líneas y 368 pixel/línea. Se aplica la técnica DCT y compensación de movimiento. Adopta una tasa de 4,36 Mb/s y 1,95 para las componentes de vídeo y 200 kb/s para el audio en estéreo. La modulación adoptada en la QPSK y se obtienen anchos de banda de 3 y 6 MHz ocupados en el transponder satelital. El equipo desarrollado por *Oak/C-Cube/Leitch* se basa sobre un *chip* CL450 del tipo **MPEG**. El chip es originalmente usado para CD-ROM interactivo posee una relación de compresión máxima de 30:1. Para aplicaciones satelitales se requiere una banda de 4,8 MHz para 6,2 Mb/s; pudiendo trabajar a la mitad de velocidad.

3.2- COMPRESION DE IMÁGENES EN MPEG

La compresión de imágenes involucra los siguientes procesos (ver **Fig 04**):

- Formato de pantalla: **CIF** para 4:3 o **HDTV** de 16:9.
- Codificación Diferencial mediante predicción y transformada compensación de movimiento: **MCT**.
- Interpolación de imágenes entre las de referencia y las de compensación de movimiento.
- Transformada discreta coseno **DCT** sobre bloques de 8x8 pixel.
- Codificación Huffman de longitud variable: **VLC**.

COMPRESIÓN DE VIDEO MPEG-ITU

A continuación se describen algunos de estos procesos, la transformada DCT y la codificación VLC se pueden encontrar en el ítem 2 de este mismo trabajo.

FORMATO DE CUADRO CIF (*Common Intermediate Format*). El concepto de formato CIF pretende superar las diversas normas analógicas en cuanto a número de líneas y puntos. Téngase en cuenta que en tanto en el formato analógico la separación entre pixel vertical y horizontal es distinto (por ello la matriz de 8x8 pixel se ve como un rectángulo) en el formato digital la separación es la misma. Esta interfaz común se aplica en las distintas variantes MPEG y ITU-T. El formato CIF sigue los siguientes lineamientos (ver la **Fig 02**):

- El CIF se basa en un cuadro de 288 líneas y de 30 cuadros/seg (alternativas 15-10-7,5).
- La frecuencia de muestreo es de 6,75 MHz.
- Se obtienen 352 muestras de luminancia y 176 muestras de crominancia por línea.
- La imagen se divide en **Bloques** de 8 líneas con 8 muestras luminancia (8x8 pixel) y 4 x 4 muestras de crominancia.
- Los bloques se agrupan en **Macrobloque** (4 bloques de luminancia y 2 de crominancia para 8x8 pixel de crominancia).
- Con 11 macrobloques de ancho y 3 de altura se obtiene un **Grupo de Bloque**.
- Con 2 grupos de ancho y 6 de altura se obtiene la **imagen CIF**.
- Solo H.261 tiene previsto el cuadro mitad **QCIF** (mitad de pixel y líneas), se aplica en videoteléfono Px64 kb/s.
- Las características de la señal CIF y QCIF se indican en la **Tabla 02**.

Tabla 02: Características principales de codificación a baja velocidad H.261.

<i>Denominación</i>	<i>CIF</i>	<i>QCIF</i>
Señal	Y - CR,CB	Y - CR,CB
Pixel/Línea	352 - 176	176 - 88
Línea/Imagen	288 - 144	144 - 72
Posibles Imagen/seg	30, 15, 10, 7,5	30, 15, 10, 7,5
Velocidad Mb/s	36, 18, 12, 9	9, 4,5, 3, 2,3

Tabla. Velocidades requeridas para aplicaciones comerciales.

<i>Norma</i>	<i>Pixel</i>	<i>JPEG</i>	<i>MPEG-2</i>	<i>H.261</i>
CIF	352 x 288	3 a 8 Mb/s	1,15 a 3 Mb/s	VHS
QCIF	176 x 144	---	---	Px64 kb/s
CCIR 601	720 x 486	15 a 25 Mb/s	5 a 10 Mb/s	PAL
HDTV	1920 x 1080	60 a 100 Mb/s	20 a 40 Mb/s	HDTV

INTERPOLACION DE IMÁGENES. La interpolación de imágenes permite reducir la cantidad de imágenes que son codificadas. Algunas de las imágenes son usadas de referencia y otras son el resultado de interpolación.

Se adoptan 3 tipos de imágenes (**Fig 02**):

- Intraframe (cuadro **I**). Se trata de cuadros con compresión moderada para usarlas como referencia. No poseen predicción temporal de movimiento MCT como en el caso de las imágenes intermedias. Posee solo compresión espacial mediante codificación DCT.
- Imagen Predictiva (cuadro **P**). Se codifica en base al I anterior (compensación MCT) y sirve para el P siguiente.
- Imagen interpolada (cuadro **B**). Calculada en base a las imágenes I y P anteriores y posteriores. El B no se transmite. Una posible secuencia de imágenes es (I0-B1.B2.B3-P1-B4.B5.B6-P2-B7.B8.B9) (**Fig 02**). Se trata de un cuadro de referencia I cada 12 cuadros. El cuadro P1 se calcula en base a I0 y el P2 en base a P1. Los cuadros B1.B2.B3 se calculan en base a I0 y P1; los B4.B5.B6 en base a P1 y P2 y B7.B8.B9 en base a P2 e I0.
- La interpolación tiene una tasa de 1/2 o 1/3 (número de cuadros B). Se denomina actividad *Backware* y *Foreware*.

COMPENSACION DE MOVIMIENTO MCT (*Motion Compensation Transform*). Este método permite compensar los cambios de posición de las imágenes con un corrimiento (*offset*) pequeño (máximo un macrobloque). Consiste en una comparación con el macrobloque de la imagen anterior en el tiempo (predicción interframe). Se verifica la adaptación con un macrobloque que posee un *Offset*. El *offset* es un corrimiento sobre los ejes x-y y para estimar los corrimientos de la imágenes. La transformada DCT trabaja sobre un bloque mientras que la MCT trabaja sobre el macrobloque.

Las particularidades de MCT son (**Fig 03**):

- MCT se aplica en la predicción diferencial mediante un proceso multi-etapas. Para seleccionar el macrobloque con menor diferencia se efectúan 3 pasos comparativos.

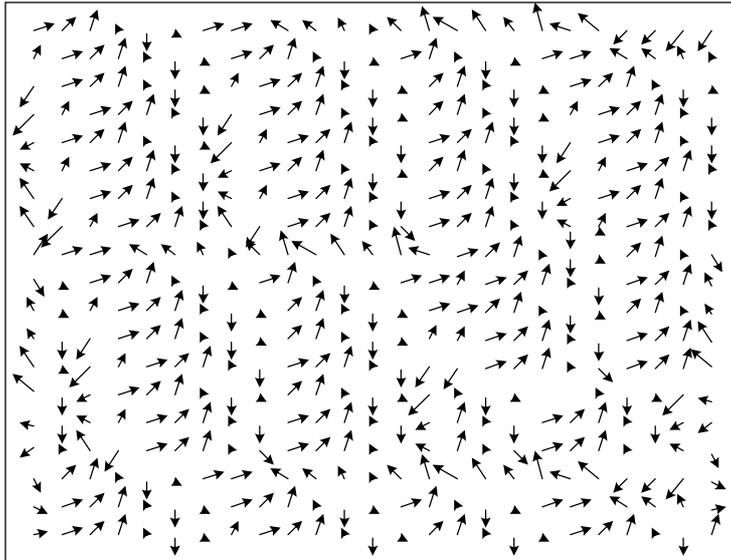
COMPRESIÓN DE VIDEO MPEG-ITU

-Por ejemplo:

Primer paso:	9 posibilidades con offset máximo es	(± 4 pixel en el eje-x) y (± 4 en el eje-y).
Segundo paso:	8 posibilidades con offset máximo es	(± 2 pixel en el eje-x) y (± 1 en el eje-y).
Tercer paso:	2 posibilidades con offset máximo es	(± 1 pixel en el eje-x) y (± 0 en el eje-y).
Resultado	El desplazamiento máximo es de	(± 7 pixel en el eje-x) y (± 5 en el eje-y).

-Son posibles otras diferentes formas de aproximaciones sucesivas al macrobloque más adaptado. La predicción se basa en el análisis de menores diferencias. Se obtiene entonces el vector de movimiento que identifica al macrobloque más adaptado. El vector se codifica y se transmite junto con la matriz 8x8. MPEG dispone de 1 o 2 vectores de acuerdo con el tipo de MCT aplicado.

Fig 03. Transformada MCT de compensación de movimiento.



COMPRESIÓN DE VIDEO MPEG-ITU

3- NORMAS PARA IMÁGENES ITU-T.

En la **Tabla 03** se indican los sistemas definidos para comunicaciones multimediales en el ITU-T. Las mismas son analizadas a continuación.

Tabla 03: Normas para conferencias multimediales del ITU-T

Norma ITU-T	Año	Aplicación	Vídeo	Audio	Múltiplex	Control
H.320	1990	ISDN	H.261	G.711	H.221	H.242
H.324	1995	POST	H.263	G.723	H.223	H.245
H.323	1996	LAN	H.261/263	G.711	H.225	H.245
H.310/321	1996	ATM	H.262	MPEG-1	H.222	H.245

3.1- ITU-T H.320.

Se trata de tecnologías referidas como velocidades **Px64 kb/s** para video-teléfono. El estándar cubre desde 64 a 2048 kb/s con un retardo inferior a 150 mseg. Se señala un protocolo de conectividad internacional que permite la comunicación entre aparatos de distinta producción y compatible con ISDN. La norma **ITU-T H.320** involucra las siguientes funciones (ver la **Fig 04**):

- H.261 para la señal de vídeo;
- G.721/722/728 para sonido;
- H.221 para el entramado de datos;
- H.230 para el control y
- H.242 para la señalización.

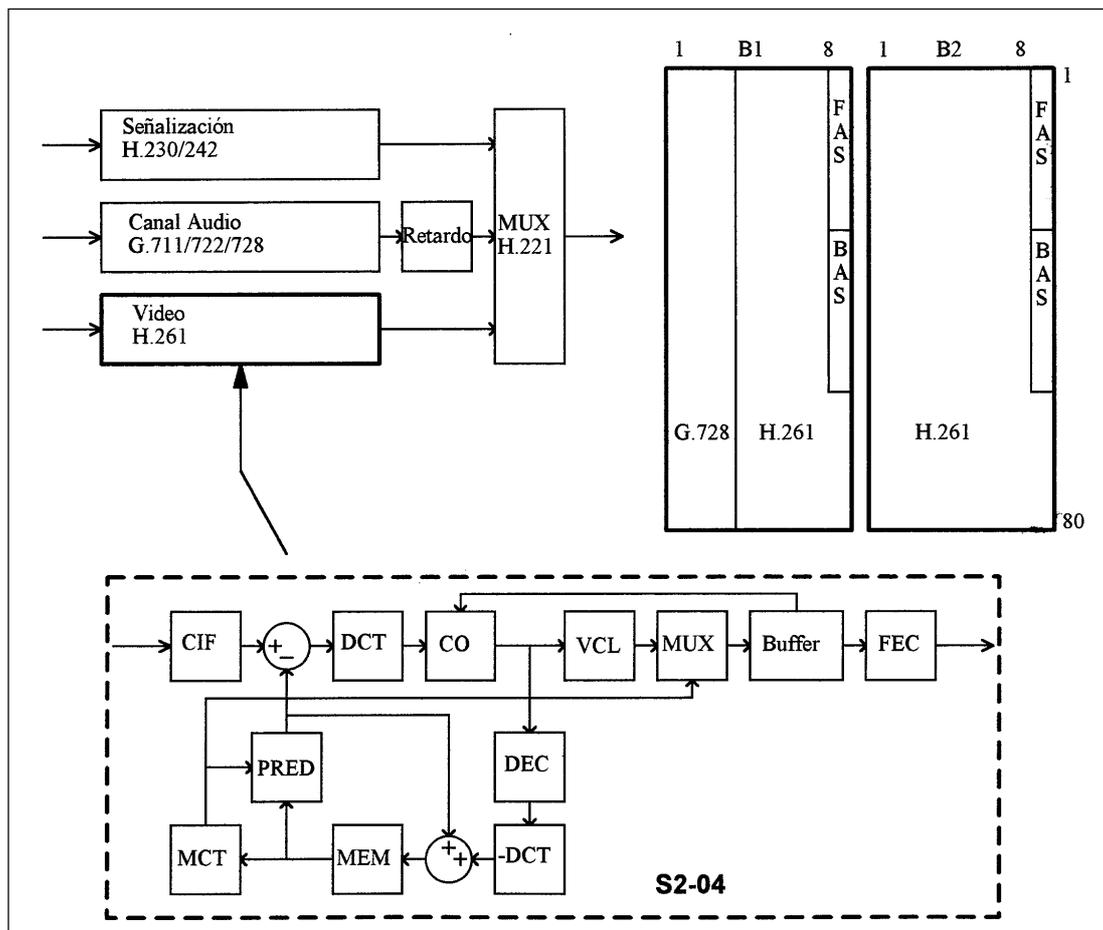


Fig 04. Procesamiento de vídeo mediante H.320.

Se determinan los componentes del sistema de videoteléfono conectado a una central privada o desde un acceso ISDN a 2x64 kb/s. El algoritmo de codificación de vídeo se indica el H.261; el algoritmo de audio en AV.250; el control de sistema en H.242 (señalización dentro de banda) y H.230 (intercambio de tramas de control); el multiplexor de las 3 señales anteriores en H.221 y el adaptador hacia la red en I.400.

COMPRESIÓN DE VIDEO MPEG-ITU

Se disponen de 3 posibilidades (el complemento a 128 kb/s se usa para señalización):

- 56 kb/s de audio según G711 68,8 kb/s de vídeo según H.261.
- 48 kb/s de audio según G722 76,8 kb/s de vídeo según H.261.
- 16 kb/s de audio según G718 108,8 kb/s de vídeo según H.261.

ITU-T H.261. Sus particularidades son:

-El proceso de codificación es del tipo MPEG. El diagrama se muestra en **Fig 04**. Se aplica el Formato Común Intermedio **CIF** de 352 pixel de luminancia por línea. Se define también el **QCIF** de 176 pixel (mitad de CIF) para sistemas de baja velocidad de acceso. Se llama formato intermedio debido a que es una mezcla de PAL, Secan y NTSC.

-Una primer etapa de codificación ADPCM usa predicción **MCT** (estimación y compensación del movimiento). La información obtenida se codifica mediante la transformada **DCT** y **VLC** (Huffman). VLC se define en H.261; en MPEG se aplica un subset de H.261 combinado con CLC. El codificador de salida posee un número de niveles cuánticos ajustable de acuerdo con el estado del buffer de VLC. El retardo producido por el buffer y VLC se introduce sobre el canal de sonido para compensar la diferencia.

-La trama de salida (H.221) involucra los datos codificados y el vector MCT de compensación de movimiento. Se utiliza un codificador FEC del tipo BCH para la corrección de errores. **ITU-T H.221** determina la trama de los canales de **videoteléfono** a **64 kb/s**. Ambos canales del acceso ISDN (B1-B2) poseen 64 kb/s. La trama dispone de 80 Bytes. Se tienen señales para alineamiento de trama **FAS** y para señalización **BAS** (**Fig 04**). El canal de señalización permite la transacción de velocidad y aplicaciones.

3.2- ITU-T H.120/324

ITU-T H.120.

Hace referencia al codificador para **videoconferencia** trabajando a **2048 kb/s**. Sus características son:

-Se trata de un sistema de 625 líneas con 50 cuadros/seg. Se dispone de un canal de audio a 64 kb/s y otro de datos con operación en dúplex.

-El muestreo de **luminancia** es de 5 MHz y la codificación a 8 bit/pixel. El predictor toma en cuenta el promedio de dos pixel. Se usa el pixel anterior en la misma línea y el posterior de la línea superior. Se codifica en 4 bits y luego en longitud variable de 2 a 9 bits. Cuando la memoria buffer se satura se introduce un submuestreo de línea y cuadro. La **chrominancia** se muestrea a 1 MHz, con codificación DPCM y VLC. El predictor equivale al punto anterior en la misma línea.

-El entramado de datos lleva el código de sincronismo de línea y cuadro. Se efectúa una corrección de errores mediante el código FEC del tipo **BCH**(4095,4035). La tasa de error se reduce desde 10^{-4} a 10^{-8} .

ITU-T H.324

Esta norma incluye la codificación H.263 para la señal de vídeo. El objetivo de **ITU-T H.263** es mejorar la calidad de H.261. Esta norma es coherente con **MPEG-4** desarrollado por ISO. Formalmente utiliza las mismas técnicas de compresión de imagen con 5 a 15 imágenes/seg. MPEG-4 utiliza estimación de movimiento para la compensación entre tramas. Por otro lado, la posición de marcas de resincronismo son periódicas en lugar de estar colocadas luego de un bloque no-periódico.

La estructura de grupos de bloques difiere en H.263 y H.261. Se define la imagen CIF con 288 líneas y 352 pixel de luminancia (la mitad para chrominancia). De esta imagen se encuentran las estructuras 4CIF y 16CIF con el doble y el cuádruple de líneas y pixel y las imágenes QCIF con la mitad de ambas.

H.324 permite la interactividad entre terminales PC-multimediales, módem de voz-datos, *Browsers* de WWW con vídeo en vivo, videoteléfonos, sistemas de seguridad, etc. Permite la conexión mediante la red telefónica convencional:

- Utiliza un módem full-duplex **V.34** a velocidad de 28.800 o 33.600 b/s, pudiendo operar a menor velocidad.
- La compresión de datos es del tipo **V.42** y el protocolo de módem es el **LAPM**.
- La codificación vocal se realiza mediante **G.723** a una velocidad de 5,3 o 6,4 kb/s (trama de 30 mseg).
- El retardo total entre extremos para el canal de audio es cercano a 97,5 mseg debido al procesamiento.
- Llega a 150 mseg con los buffer de jitter y de multiplexación, sin contar el tiempo de propagación.

COMPRESIÓN DE VIDEO MPEG-ITU

3.3- ITU-T H.323

Esta tecnología permite la transmisión en tiempo real de vídeo y audio por una red de paquetes. Es de suma importancia ya que los primeros servicios de voz sobre protocolo Internet (**VoIP**) utiliza esta norma. En la versión 1 del protocolo H.323v1 del año 1996 se disponía de un servicio con calidad de servicio (**QoS**) no garantizada sobre redes LAN. En la versión 2 del año 1998 se definió la aplicación VoIP. Una versión 3 posterior incluye el servicio de fax sobre IP (**FoIP**) y conexiones rápidas entre otros.

Los componentes del servicio H.323 son los siguientes:

- Terminal: se realiza en forma bidireccional en tiempo real; se trata de una PC o un equipo a medida.
- Gateway: provee la conectividad entre la red H.323 y otra distinta (PSTN por ejemplo).
- Gatekeeper: es el centro de control para el procesamiento de la llamada, direccionamiento, autenticación, gestión de ancho de banda, tarificación, etc). El Gatekeeper debe realizar la traslación de dirección IP a la E.164 de la red telefónica. El proceso de admisión utiliza mensajes de protocolo RAS, requerimiento de admisión, confirmación y rechazo de admisión.
- Unidad de control multipunto: para realizar conferencias entre varios usuarios.

Los protocolos especificados por H.323 para efectuar las funciones entre componentes son los siguientes:

- Codificación de audio: G.711 a velocidad de 64 kb/s; G.722 para 48, 56 y 64 kb/s; G.728 para 16 kb/s y G.729 para 8 kb/s.
- Codificación de vídeo: de acuerdo con **H.263**.
- Protocolo **RAS** (*Registration, Admission and Status*) según H.225 para la comunicación entre terminal, gateway y gatekeeper. Sirva para registración, control de admisión, control de ancho de banda, estado y desconexión.
- Señalización de llamada: para establecer la conexión y desconexión mediante protocolo H.225.
- Señalización de control: mediante protocolo H.245 para comandos, indicaciones, control de flujo, gestión de canales lógicos.
- Protocolo **RTP** (*Real-Time Transport Protocol*): usado con UDP/IP para identificación de carga útil, numeración secuencial, monitoreo, etc. Trabaja junto con **RTCP** (*RT Control Protocol*) para entregar un feedback sobre la calidad de la transmisión.

PROCEDIMIENTO DE COMUNICACION. Se describe a continuación el procedimiento de conexión, comunicación y desconexión de una llamada mediante H.323. Intervienen dos terminales T1 y T2 y el Gatekeeper en el centro.

