

CAPÍTULO 10:

VÍDEO

El vídeo empieza a ser, cada día más, un importante componente en las producciones multimedia. El método más idóneo para incluir vídeo en las producciones multimedia es capturarlo utilizando una cámara digital, no obstante, debido a la gran cantidad de información que existe en formato analógico se hace necesaria la digitalización.

En cualquier caso el principal problema al utilizar vídeo en las producciones multimedia es el gran tamaño de los ficheros generados; esto los hace poco útiles en multimedia a través de Internet. Es necesaria la compresión.

Afortunadamente existe un estándar para la compresión de vídeo el MPEG.

En este Capítulo aprenderas:

Cómo se generan las imágenes en un televisor.

Cual es la información que se transmite en una emisión televisiva.

Cómo funciona un vídeo analógico.

Cómo se digitaliza vídeo analógico y los estándares existentes.

Las técnicas de compresión de vídeo digital y en particular el estándar MPEG.

Cuales son los estándares para la difusión de vídeo digital.

ESQUEMA DEL CAPÍTULO 10: VÍDEO

10.1 LA TELEVISIÓN.

10.1.1 Funcionamiento.

10.1.2 Características de las imágenes.

10.2 EL VÍDEO.

10.2.1 Funcionamiento del vídeo.

10.2.2 Breve historia del vídeo doméstico.

10.3 DIGITALIZACIÓN DE VÍDEO.

10.4 ESTÁNDARES DE VÍDEO DIGITAL.

10.5 COMPRESIÓN DE VÍDEO DIGITAL.

10.5.1 Motion JPEG.

10.5.2 DV.

10.5.3 Vídeo MPEG.

10.5.4 Software Codecs.

10.5.5 QuickTime.

10.6 DIFUSIÓN DE VÍDEO EN TIEMPO REAL.

APÉNDICE A: ESPACIOS DE COLOR BASADOS EN DIFERENCIAS DE COLOR.

RESUMEN

10.1 LA TELEVISIÓN.

Igual que en el caso de la animación, la latencia de las imágenes en la retina es el fundamento sobre el que se basa el mecanismo de la televisión. Imágenes con pequeños cambios entre ellas y mostradas lo suficientemente aprisa una tras otras son percibidas como un movimiento continuo.

A la frecuencia mínima a la cual se han de mostrar las imágenes para que aparezcan en movimiento se le llama *frecuencia de latencia* y es de alrededor de 40 fotogramas por segundo (f.p.s) en el caso de los humanos. Si la frecuencia de presentación de las imágenes es menor que este umbral, se produce el fenómeno de parpadeo.

10.1.1 Funcionamiento.

Las imágenes en un aparato de televisión se deben a la luz producida cuando un haz de electrones golpea el material fosforescente de la pantalla. El haz de electrones recorre la pantalla de arriba hacia abajo y de izquierda a derecha, en la Figura 10.1 se muestra el recorrido del haz de electrones sobre la pantalla. El brillo se modula mediante la velocidad de los electrones que golpean la pantalla.

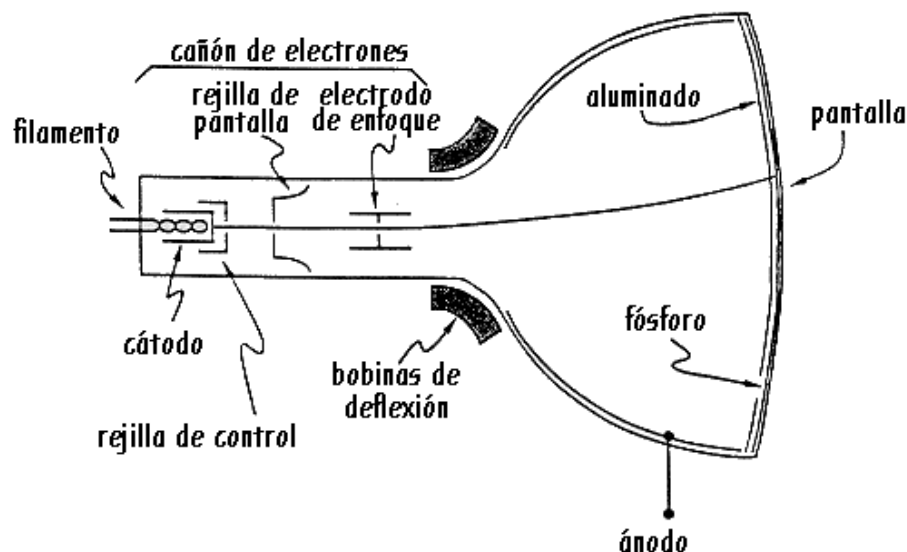


Figura 10.1: Esquema del funcionamiento de un tubo de rayos catódicos.

Las primeras emisiones de televisión no disponían del suficiente ancho de banda como para transmitir las imágenes con una frecuencia superior a la frecuencia de latencia, se ideó entonces el sistema del entrelazado, para generar una imagen primero se barren las líneas impares y luego las líneas pares, con ellos la información necesaria para generar una imagen es la mitad y la frecuencia de refresco se duplica.

10.1.2 Características de las imágenes.

No existe un único estándar para la difusión de imágenes de televisión. En Europa, Australia y Nueva Zelanda se utiliza el estándar PAL (Phase Alternating Line) con 625 líneas para cada imagen de las cuales únicamente 576 corresponden a datos de la

imagen. el resto son señales de control. La frecuencia de refresco se hace coincidir con la frecuencia de la red eléctrica. En Europa es de 50 Hz.¹

En Estados Unidos, Japón y Taiwan, se utiliza el estándar NTSC abreviatura de National Television System Committee. Bajo este estándar una imagen de televisión contiene 525 líneas, de las cuales únicamente 480 contienen datos de la imagen. La frecuencia de refresco en estos países es de 60 Hz.

Finalmente, en Francia utilizan su propio estándar llamado SECAM (Séquential Couleur avec Mémoire) también con 625 líneas 576 correspondientes a información de la imagen, con una frecuencia de refresco de 50 Hz.

En los tres estándares la relación ancho : alto de una imagen es 4:3.

Como ya hemos comentado, la señal de televisión no porta únicamente información de la imagen sino también otro tipo de información, como señales de sincronismo, en la Figura 10.2 se muestra la forma de onda de una señal típica de televisión para un estándar cualquiera.

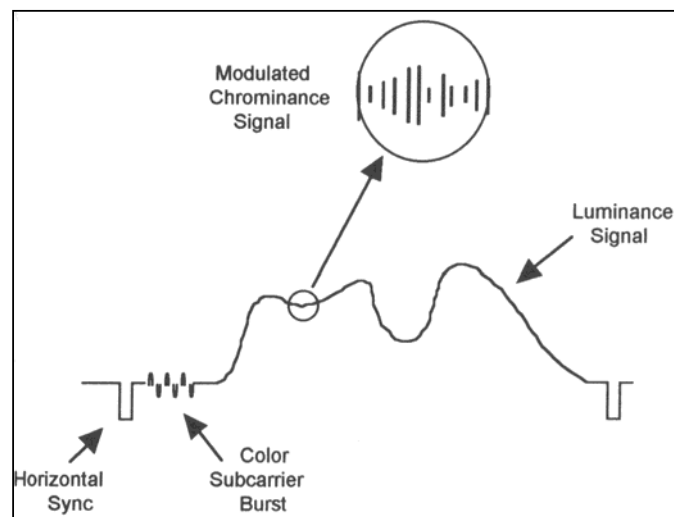


Figura 10.2: Señal analógica de una emisión de televisión.

Ya que el ojo humano es más sensible a los cambios de intensidad luminosa que a los cambios de color, el ancho de banda utilizado para transportar los valores de luminancia es el doble que el utilizado para los de crominancia. En la Figura 10.2 se puede observar que la señal de crominancia es una señal de mayor frecuencia que la señal de luminancia que además la modula.

10.2 EL VÍDEO.

La principal fuente de vídeo digital es la digitalización de vídeo analógico. Esto es debido a la gran cantidad de material documental que existe en este formato.

¹ La frecuencia de refresco no es exactamente la utilizada en la red eléctrica, ya que esto causa problemas de interferencias. La frecuencia de refresco de las emisiones televisivas se hace coincidir con el 1000/1001 de la frecuencia de la red eléctrica.

10.2.1 Funcionamiento del vídeo.

La grabación de las señales de vídeo primero se intentó de modo longitudinal en la cinta (como en las cintas de audio), lo que implicaba a una gran velocidad de arrastre y una enorme longitud de cinta.

Para evitar este problema se idearon otros sistemas (helicoidal, transversal, etc.), lo que se conseguía moviendo la cabeza lectora además de la cinta. Estos métodos permiten almacenar más fotogramas por metro de cinta.

10.2.2 Breve historia del vídeo doméstico.

La historia del vídeo doméstico se remonta al siglo pasado. Estos son los principales hitos dentro de esta historia:

- En 1927, John Logie Baird inventó la "fonovisión", para grabar imágenes de televisión en discos de fonógrafo. La calidad era muy mala pero fue un primer intento.
- En 1951, Bing Crosby Enterprises construyó el primer magnetoscopio en blanco y negro. Al año siguiente la RCA desarrolló un prototipo en color.
- En los sesenta se idearon cientos de formatos y tipos distintos (Sony, Panasonic, Philips, General Electric, etc.), fundamentalmente para grabación y redifusión de programas de TV. La cinta era bastante ancha, de dos pulgadas.
- En los setenta empezaron a aparecer los primeros aparatos domésticos (CBS, Philips, etc) y se redujo el ancho de la cinta hasta las 3/4 de pulgada del sistema U-matic.
- En 1975 apareció el sistema Betamax con cintas de media pulgada, diseñado por Sony.
- En 1977 JVC comercializó el VHS.
- En 1979 Philips y Grundig lanzaron al mercado el Vídeo 2000.
- La política de VHS, mucho más abierta que la de Sony, junto a la mayor duración de las cintas, provocaron que el mercado se inclinara significativamente por el VHS que, como sabemos, es el sistema que finalmente se ha impuesto en el mercado doméstico.
- En los noventa está surgiendo la siguiente generación de vídeo: el digital, DV. Y con él el DVD. El vídeo digital de consumo es de calidad similar al vídeo de estudio, con 500 líneas de resolución.

10.3 DIGITALIZACIÓN DE VÍDEO.

El problema principal en la digitalización de vídeo es el tamaño de los ficheros generados. Por ejemplo, un segundo de vídeo digital obtenido a partir de una señal analógica con formato PAL entrelazado da un tamaño de fichero de $768 \times 576 \times 26 = 1'85$ Gb.

Existen dos posibilidades para obtener vídeo digital bien digitalizar imagen analógica, opción que introduce una gran cantidad de ruido en la imagen final o bien utilizar un sistema de captura de la imagen digital, una cámara digital.

Aún teniendo en cuenta la ventaja que supone utilizar una cámara de vídeo digital para la adquisición de imágenes, la gran cantidad de material documental existente en

el formato de vídeo analógico hace que esta sea un fuente importante para el vídeo digital.

Al digitalizar vídeo nos podemos encontrar con una serie de problemas, uno de ellos deriva de que, en general, los monitores de ordenador no generan la imagen de modo entrelazado, como es el caso de las señales de televisión y vídeo, si no que la imagen es generada sobre un buffer y una vez generada se muestra, de una sola vez, por pantalla. Esto evita, entre otros, el problema del parpadeo.

El problema del entrelazado se puede resolver con varias estrategias. Una de ellas es promediar las líneas pares con las impares, pero con este método no se obtiene una buena calidad en la imagen final. Otra posibilidad es tomar tan solo las pares e interpolar la información en las impares.

Pero la estrategia que mejores resultado suele dar es obtener a partir de una imagen de vídeo analógico dos de vídeo digital, la correspondiente a las líneas pares y la correspondiente a las líneas impares. El resultado es que hemos generado el doble de imágenes con lo que hemos dividido por dos la velocidad del movimiento que se observa en la filmación. Este es un problema menor fácil de solventar.

10.4 ESTÁNDARES DE VÍDEO DIGITAL.

Existe un estándar para digitalización de vídeo, la norma Rec. ITU-R BT.601 o abreviadamente CCIR 601. Según este estándar, cada línea horizontal se muestrea utilizando 720 muestras para la luminancia y 320 muestras para las crominancias U y V. En el apéndice A se detalla el espacio de color basado en diferencias de color.

El estándar de televisión y vídeo analógico PAL contiene 576 líneas con información de la imagen y teniendo en cuenta la relación ancho : alto de 4 : 3, el número de muestras en cada línea es $576 \cdot 4/3 = 768$. Este número de muestras asume que un píxel tiene la misma longitud que altura. Pero, si se va a utilizar el estándar CCIR 601, los píxeles son un poco mas anchos que altos.

En el caso del estándar NTSC los píxeles, al utilizar el estándar de digitalización CCIR 601 son, al contrario que en PAL, un poco más altos que anchos.

La forma de muestrear la señal en el estándar CCIR 601 para la luminancia y las crominancias se muestra en la Figura 10.3.

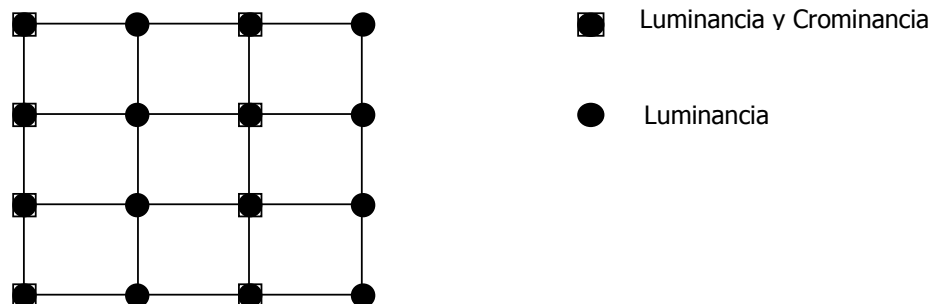


Figura 10.3: Posiciones de muestreo de luminancia y crominancia siguiendo el estándar de digitalización CCIR 601.

10.5 COMPRESIÓN DE VÍDEO DIGITAL.

Como hemos visto en §10.3, el principal inconveniente en la manipulación de vídeo digital es el enorme volumen de los ficheros generados. Es de capital importancia en este tipo de media, más que en ningún otro, el uso de la compresión de los ficheros.

Existen dos técnicas en la compresión de vídeo, técnicas de compresión espacial en la que cada fotograma se comprime de modo independiente, y técnicas de compresión temporales, técnica en la que se toman de toda la secuencia algunos fotogramas clave y el resto de fotogramas es almacenado como diferencia con respecto del clave.

Además las técnicas de compresión temporal se pueden subdividir a su vez en técnicas de compresión con pérdidas y sin pérdidas.

10.5.1 Motion JPEG.

Lo primero que debemos decir de la compresión Motion JPEG, abreviadamente MJPEG, es que no tiene nada que ver con la compresión MPEG.

La compresión MJPEG es una técnica de compresión espacial de video. Cada uno de los fotogramas es comprimido utilizando la compresión JPEG de modo independiente. Los ficheros con este tipo de compresión pueden ser reproducidos por QuickTime.

Si la frecuencia de transferencia de información está limitada a 3Mb/s. es necesario que la tasa de compresión sea de alrededor de 7:1

10.5.2 Digital Video (DV).

En esta técnica de compresión se parte de una imagen digitalizada por submuestreo de crominancia 4:1:1. Sobre esta imagen digitalizada se efectúa una compresión espacial por transformada del coseno y, únicamente en el caso es que la imagen original esté entrelazada, se efectúa una compresión temporal por compensación de la imagen, técnica que se describirá en la sección dedicada a la compresión MPEG.

Para alcanzar una tasa de transferencia de información de 3'25 Mb. el ratio de compresión debe ser de 5:1.

10.5.3 Vídeo MPEG.

MPEG es un estandar de compresión de vídeo internacional propuesto por el *Motion Picture Expert Group*.

Esta técnica de compresión, y su especificación, es tremendamente compleja, aquí describiremos a grandes rasgos cual es el procedimiento.

Inicialmente, cada imagen digital se divide en bloques de tamaño 16x16 píxeles a los que se llaman técnicamente macrobloques.

Para la siguiente imagen en la secuencia se intenta predecir como contribuye cada uno de los macrobloques de la imagen anterior, el fotograma clave. Esto se hace por tanteo, se prueban todos los posibles desplazamientos de los macrobloques de la imagen anterior y se toma el que mejor se ajusta.

Al macrobloque real se le substraee el predicho y se almacena la diferencia entre ellos y el desplazamiento utilizado.

Esta es la técnica que se conoce como *compensación de movimiento*. Finalmente los fotogramas son comprimidos utilizando la compresión JPEG

Existen cuatro tipos de fotogramas en el formato MPEG:

Fotogramas I (*Intra-coded*): Estos son los fotograma clave o de referencia. Aparecen cada 10 ó 15 fotogramas y están comprimidos utilizando el algoritmo JPEG sobre imágenes estáticas.

Fotogramas P (*Predictive-coded*): Se codifican por compensación de movimiento a partir de un fotograma I o P anterior a ellos.

Fotogramas B (*Bidirectional predictive-coded*): Se codifican por compensación de movimiento a partir de un fotograma I o P tanto anterior como posterior, de ahí su nombre. No pueden ser tomados como fotograma clave.

Fotogramas D (*DC-coded*): Son utilizados en el caso particular en que se desee compresión sin pérdidas, son como fotogramas I y no hacen referencia a ningún otro fotograma de la secuencia.

Una posibilidad para la cadena de fotogramas comprimidos es la siguiente:

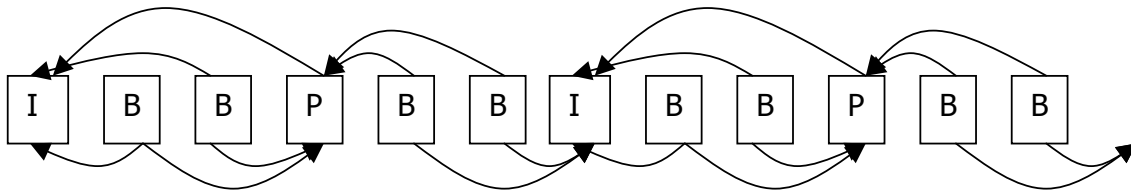


Figura 10.4: Una posible ordenación en los fotogramas de una secuencia MPEG.

Un problema con esta ordenación es que si se desea transmitir para su reproducción la secuencia, para mostrar el segundo fotograma de la secuencia, que es de tipo B, necesitamos esperar a que llegue el cuarto, de tipo P, al cual hace referencia.

Este problema se resuelve reordenando los fotogramas antes de su transmisión del siguiente modo:

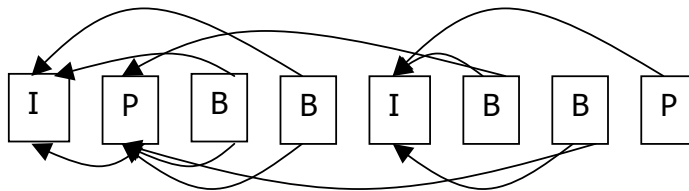


Figura 10.5: Ordenación en la que cada fotograma sólo hace referencia a los anteriores a él.

Los fotogramas P se comprimen con un factor tres veces mayor que los I y los B con un factor 1'5 veces mayor que los P.

Para un vídeo de tamaño de fotograma 352x240 píxeles con submuestreo cromático 4:2:0 y 30 f.p.s. la tasa de transferencia necesaria es de tan solo 1'86 Mb.

Es evidente que tanto la compresión como la descompresión de los vídeos en formato MPEG necesitan gran cantidad de cálculo, con lo que se hace necesario el uso de hardware para su compresión de reproducción.

10.5.4 Software Codecs.

Los software codecs son programas que efectúan todos los cálculos necesarios para descomprimir vídeo comprimido en alguno de los siguiente formatos: MPEG, Cinepak, Intel Indeo Y Sorensen.

El formato MPEG ya lo hemos comentado, los tres restantes usan para la compresión una técnica llamada *compresión por vector de cuantización*.

En la compresión por vector de cuantización se divide cada imagen en bloques. Cada uno de estos bloques se compara con una lista estándar y es sustituido por aquel con el que guarda un mayor parecido, los bloques originales son sustituidos por índices a una tabla. Ya que la decodificación únicamente consiste desreferenciar un índice, son muy rápidos.

Además también utilizan fotogramas clave y fotogramas diferencia.

Cinepak es el que mayores tasas de compresión proporciona. Sorensen es el que ofrece mayor calidad.

10.5.5 QuickTime.

Existe una gran cantidad de formatos de vídeo y no cabe esperar que se opte por estandarizar uno de ellos, lo que constituye un problema desde el punto de vista de la producción multimedia.

QuickTime, de Apple, es un intento de utilizar una arquitectura estándar para todos los formatos de vídeo existentes, y que se está convirtiendo en un estándar de hecho.

Que QuickTime sea una arquitectura estándar y no otro formato significa que una película en QuickTime no sólo contiene los datos de la imagen, sino que además contiene la información para organizar, acceder y manipular estos datos. El modo de reproducir una determinada película es parte de esta información.

Cada vez que un productor de vídeo lanza al mercado un nuevo formato y quiere que este forme parte de QuickTime, debe escribir el plugin para QuickTime y hacer que este haga el resto del trabajo.

10.6 DIFUSIÓN DE VÍDEO EN TIEMPO REAL. STREAMING.

El *streaming* se puede definir como *difusión en tiempo real*. En el caso del vídeo, por ejemplo, si un cliente solicita la visualización de una determinada película, a través de una línea de comunicación de datos, esta le es servida para ser visualizada de modo inmediato, sin necesidad de ser almacenada.

Otro ejemplo de difusión de vídeo en tiempo real es la video conferencia donde las imágenes mostradas no se almacenan.

Existen dos formatos para difusión de vídeo en tiempo real: QuickTime y Real Networks' RealVideo.

APÉNDICE A: ESPACIOS DE COLOR BASADOS EN DIFERENCIAS DE COLOR.

Cuando se inventaron los televisores en color los televisores en blanco y negro ya existían. Las nuevas emisiones en color se debían poder ver también en un televisor en blanco y negro. En un televisor en blanco y negro sólo percibimos valores de intensidad luminosa o luminancia. Los modelos de color basados en diferencia se idearon para solucionar, entre otros, el problema presentado con los televisores en color.

En un espacio en color basado en diferencia se tiene el valor de la luminancia y dos diferencias de color con respecto de la luminancia. Para calcular la luminancia a partir de las tres componentes de color R, G y B se utiliza la siguiente fórmula en los actuales televisores en color:

$$Y = 0'2125R + 0'7154G + 0'0721B \text{ donde } Y \text{ es la luminancia.}$$

Que los factores que multiplican a las componentes de color no sean los mismos se debe a que el ojo tiene muchos mas fotosensores sensibles al verde que a cualquier otra componente de color.

Las diferencias de color que se utilizan son B-Y y R-Y a las que se llama crominancias. Dando el valor de la luminancia y las crominancias se pueden obtener los valores de las componentes de color R, G y B.

RESUMEN

La percepción de las imágenes en movimiento en televisión es debida al fenómeno de la latencia del ojo humano.

Existen diferentes tipos de estándares para las emisiones de televisión. Algo común a todas ellas es que la señal transmitida no contiene únicamente información de la imagen sino además otro tipo de información como señales de sincronismo.

Las señales de televisión están entrelazadas, primero se difunden la información correspondiente a las líneas pares y luego a las impares.

En los últimos años ha surgido una gran cantidad de estándares de vídeo, de entre ellos el más popular en uso doméstico es el VHS.

La principal fuente de video digital es el video analógico debido a la gran cantidad de información documental que existe en este formato. Para poder digitalizar vídeo analógico es necesario desentrelazar la señal.

Existe un estándar de digitalización de vídeo, la norma CCIR 601. El problema con el video digital es gran tamaño del fichero de datos que se genera. Se hace necesaria la compresión.

Las técnicas de compresión de vídeo pueden ser espaciales, cada fotograma se comprime de modo independiente (Motion JPEG), y temporales en las que cada fotograma se comprime con referencia a algunos otros fotogramas de referencia (DV, MPEG).

La compresión MPEG es el estándar de compresión de vídeo. Es una técnica que combina la compresión espacial y la temporal.

QuickTime es un estándar de facto para formato de vídeo comprimido. Una película QuickTime no contiene únicamente los datos de las imágenes, sino además el modo en que estos datos se manipulan para, por ejemplo, ser visualizados.