

Conceptos básicos de vídeo y audio. Introducción

Alex Ribelles García

PID_00176933



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción.....	5
Objetivos.....	6
1. Características de la imagen.....	7
1.1. Fundamentando la imagen: parámetros básicos	7
1.2. Trabajando con señales de vídeo analógico	8
1.2.1. Vídeo compuesto	8
1.2.2. S-Vídeo	9
1.2.3. Vídeo analógico en componentes	9
1.3. Vídeo digital en calidad estándar, alta definición y 3D	10
1.3.1. SDI (Serial Digital Video)	11
1.3.2. HD-SDI (High Definition SDI)	12
1.3.3. 3G-SDI	12
1.3.4. DVI (Digital Visual Interface)	12
1.3.5. HDMI (High Definition Multimedia Interface)	13
1.3.6. Firewire	13
1.3.7. USB (Universal Serial Bus)	14
1.3.8. Conclusión	14
1.4. <i>Broadcast</i> frente a vídeo en Internet	15
1.5. Resoluciones, velocidad de cuadro, relaciones de aspecto y otros parámetros	19
2. Características del sonido.....	23
2.1. Audio analógico	23
2.2. Audio digital	23
3. La revolución en la gestión de las señales.....	25
4. Codificación de la imagen fija.....	27
4.1. La necesidad de comprimir	27
4.1.1. Codificación y compresión. Tipos de compresión	28
4.1.2. Redundancia y entropía	28
4.1.3. Medida de la compresión de imagen	29
4.1.4. Medida objetiva y subjetiva de la calidad de imagen	30
4.2. El JPEG	30
4.2.1. El modo secuencial	31
4.2.2. Otros modos de trabajo del JPEG	44
4.3. JPEG 2000	45
4.4. JPEG XR	47
4.5. Tendencias futuras de la codificación de imagen	48

Resumen.....	49
---------------------	-----------

Introducción

Desde el momento en el que la madurez tecnológica de los dispositivos móviles ha llegado a tal punto que los capacita para la generación, gestión y distribución de contenidos multimedia, la frontera que los separaba de los equipos dedicados se ha difuminado. Así, se han aplicado buena parte de las tecnologías ya existentes y se han generado otras nuevas que aprovechan la flexibilidad y movilidad características del nuevo medio.

Una buena captación es importante para conseguir material audiovisual que integrar en nuestros proyectos, al igual que seleccionar el formato adecuado para su almacenamiento y posterior tratamiento digital. Así pues, empezamos el módulo construyendo sobre los fundamentos básicos de la imagen y el sonido, definiendo los parámetros que los caracterizan, los estándares de interconexión y acabamos viendo el estándar de almacenamiento de la imagen estática, con el que se introducen los conceptos básicos de codificación y compresión que estarán presentes en el resto de módulos, especialmente en el de vídeo.

El módulo es eminentemente práctico, si bien es de obligada necesidad introducir conceptos teóricos que se ilustran en las actividades integradas en el texto. Como herramientas de trabajo, se van a utilizar dos programas, el **VLC de Videolan** (para todas las plataformas) y el programa **Adobe Photoshop** (Windows y Mac) que se incluye en el software de la asignatura.

Objetivos

Los objetivos que se pretende alcanzar con el estudio de este módulo son los siguientes:

1. Identificar los parámetros que caracterizan la imagen y el sonido.
2. Diferenciar los estándares de interconexión multimedia domésticos, semi-profesionales y profesionales.
3. Identificar el transporte físico más adecuado según el tipo de contenido digital, coste y prestaciones.
4. Diferenciar los conceptos de codificación y compresión.
5. Interpretar los parámetros de compresión de JPEG y JPEG2000 y seleccionar los valores adecuados en función del contenido visual de la imagen.
6. Identificar los artefactos visuales más habituales en estos estándares de imagen.

1. Características de la imagen

En la generación de contenido multimedia, es frecuente la participación de elementos audiovisuales originados por captura, es decir, digitalizados. En este primer módulo, vamos a introducir las características básicas que definen la señal de vídeo y audio y, a partir de ellas, ilustraremos y compararemos los diferentes tipos de señales y transportes de vídeo y audio existentes en el mercado industrial y *broadcast*, con los que se trabaja en el día a día. A partir de ese momento y en el resto del curso, nos vamos a centrar en las señales digitales y presentaremos cómo se codifica la imagen estática. En este recorrido, destacaremos conceptos de compresión que utilizaremos también intensivamente en otros módulos de la asignatura.

Ved también

Estos conceptos de compresión se utilizarán concretamente en los módulos "Digitalización, almacenamiento y transmisión de audio y vídeo".

1.1. Fundamentando la imagen: parámetros básicos

Cualquier texto sobre imagen o vídeo comienza con la definición de algunos términos básicos y, en nuestro caso, vamos a ser breves:

1) **Píxel**: es el elemento visual mínimo de una imagen. Es la piedra angular de todas las formas de imagen digital, incluyendo tanto la fotografía como el vídeo, y si bien el concepto es sencillo su uso no lo es, en especial en un mundo de formatos digitales de variada relación de aspecto.

2) **Luminancia o luma**: es la porción de la señal de vídeo que representa el brillo de cada píxel. La luminancia máxima de un píxel es el blanco y la mínima, el negro. El ojo humano está desarrollado para percibir una gran cantidad de niveles de luminancia.

3) **Crominancia o croma**: es la otra porción de la señal de vídeo, que representa el color de cada píxel. Los colores se destinan a cubrir todo el espectro completo del sistema visual humano, del rojo al verde y al azul, en cualquier combinación. El ojo humano, a diferencia de otros seres vivos, no posee la misma capacidad de discriminación de colores que la que posee para la luminancia. En realidad, esta porción consta de varios elementos diferentes, cuya combinación da el color.

4) **Escaneado**: es el proceso utilizado en una señal de vídeo tanto para capturar como para mostrar los valores de luminancia y crominancia de cada píxel. El escaneado sitúa la información de cada píxel en un orden específico para que un equipo pueda determinar la información que pertenece a cada píxel. Para el ojo humano, la presentación de una secuencia de imágenes escaneada en pantalla a más de 18 imágenes por segundo ya le genera un efecto de movimiento.

Escaneado

La presentación en dispositivos en matriz, como pantallas LCD, plasma o LED, ya no se realiza en escaneado, sino que se refrescan todos los píxeles de manera síncrona.

Los cuatro conceptos están íntimamente ligados. Así, como un carácter de texto o un byte de datos, cada píxel tiene un valor o conjunto de valores. Uno de ellos es el brillo (la luminancia) y, si es en color, también la crominancia. Para obtener una imagen fija, el valor de cada píxel es fijo y suficiente, aunque se necesitan los suficientes como para que la imagen sea reconocible por el destinatario. Pero para vídeo, además, la información de cada píxel se escanea (se presenta) al menos 18 veces por segundo. Más adelante, haremos algunos números para demostrar el gran problema que supone la gestión de tal cantidad de datos para lograr transmitirlo y el porqué de los estándares de compresión.

1.2. Trabajando con señales de vídeo analógico

Las señales de vídeo analógico, por su naturaleza, pueden crearse a partir de tres señales diferentes de color (R, G y B) o bien a partir de dos señales (luminancia Y y crominancias U y V).

En el quehacer diario con equipos de vídeo (como cámaras y grabadores) y programas de gestión (como tratamiento de la imagen o transmisión de vídeo), trabajamos con estas señales en sus diferentes formatos a la vez, a veces de manera inconsciente. Veámoslo con detalle.

1.2.1. Vídeo compuesto

Una señal de vídeo compuesto contiene toda la información necesaria para presentar vídeo en color en pantalla. La señal se envía normalmente en un único cable coaxial entre dispositivos como un reproductor de vídeo y un televisor. Muchos sistemas de consumo utilizan un conector amarillo, denominado RCA (aunque a veces podemos encontrar un conector BNC). Es el más extendido en el sector doméstico, pero el de menor calidad: la señal de vídeo compuesto es PAL, sistema analógico de vídeo en color de 625 líneas y 25 imágenes por segundo que también se utilizaba para las emisiones de televisión analógicas hasta el 2010. Al igual que sus sistemas hermanos NTSC y SECAM, tiene ventajas e inconvenientes, que indicamos en un cuadro resumen al final de este apartado.

Los equipos de captura de vídeo que se basaban en señales de vídeo compuesto fueron la puerta de entrada al mercado doméstico y semiprofesional en la década de 1980. Aquellos pioneros que trabajaban con un Mac, Amiga o Atari, entre otros, y un reproductor VHS o Beta necesitaban un hardware externo de



Conectores RCA

Hardware externo

Este hardware externo era, por ejemplo, Truevision TGA de Targa o Video Toaster de Newtek.

muy alto coste a veces integrado en el mismo equipo, siempre con problemas de interferencia en las señales y multitud de parámetros por ajustar en cada captura para conseguir una calidad aceptable.

1.2.2. S-Vídeo

Denominado también conector S-VHS, es analógico como el vídeo compuesto, pero la transmisión de luminancia y croma se lleva a cabo a través de diferentes cables, lo que reduce sensiblemente la degradación de la señal. Por esta razón, un cable de S-Vídeo tiene cuatro pines: un par para la señal de croma U + V y otro par para la de luminancia Y.

Su llegada a finales de la década de 1980 con los equipos SuperVHS y 8 mm facilitó mejorar tanto la calidad de imagen tanto en reproducción como en captura y el S-Vídeo aún puede encontrarse de manera testimonial en dispositivos de vídeo de consumo, incluyendo pantallas de vídeo, reproductores de DVD, videocámaras digitales, decodificadores de televisión por cable y satélite o consolas de videojuegos. En general, para aplicaciones de consumo y semiprofesionales, el S-Vídeo debe usarse siempre que sea posible en lugar del vídeo compuesto **si no hay alguna alternativa digital** como las que veremos más adelante.

1.2.3. Vídeo analógico en componentes

Superior en calidad a los dos anteriores, utiliza un cable diferente para cada una de las tres señales, por lo que la calidad de color era mejor que la anterior al sacrificar la distancia máxima entre reproductor y presentación o captura.

Hay dos versiones diferentes según el equipo:

a) **Vídeo en componentes RGB:** cada señal es de un color específico y de buena resolución, además supone un mínimo procesado para el equipo que la emite y el que la recibe, por lo que es compatible directamente con la pantalla de presentación y mejora la imagen sensiblemente.

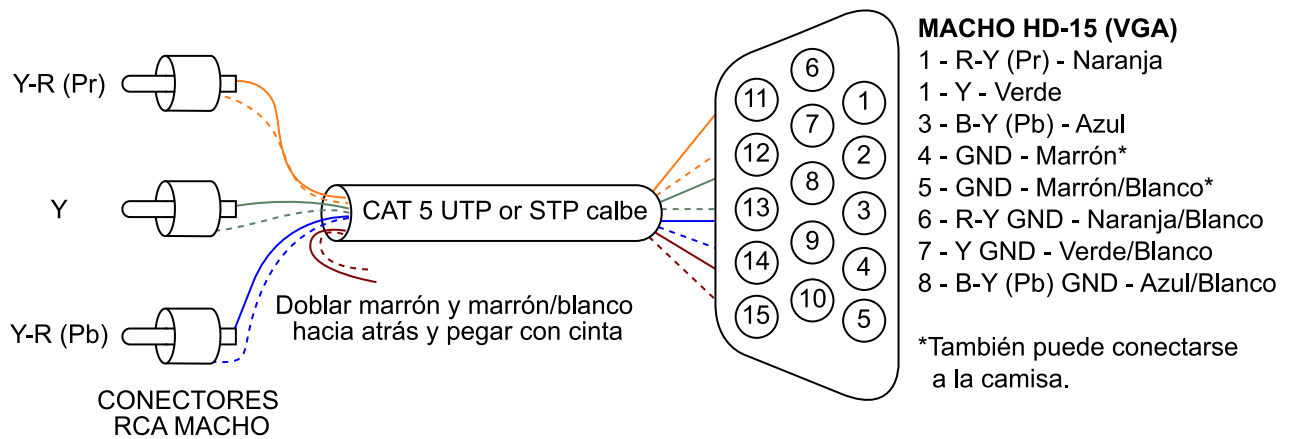
b) **Vídeo en componentes YUV:** versión que trabaja con la señal de luma y las dos de croma. Es recomendable en equipos que trabajen internamente con éstas, como el Betacam SP. Si va a ser usada para suministrar vídeo a una pantalla, deberá convertir internamente las señales a RGB.

Ved también

Al final de este subapartado, encontraréis una tabla resumen con las ventajas y desventajas de este tipo de señales.



Conectores S-Vídeo



Asignación de pines entre componentes analógicos YUV y conector VGA

Ejemplos

Un ejemplo aún vigente de este sistema de conexión es el estándar **VGA** en componentes analógicos, que añade señales de pulsos de sincronía para ajustar la señal al barrido de la pantalla.

Otro ejemplo aún vigente es el robusto **euroconector**, que de manera excepcional incluye diferentes formatos de vídeo de la misma señal: vídeo compuesto, S-Vídeo y componentes RGB analógico para dar servicio a todo tipo de pantallas desde el mismo conector, además de audio estéreo analógico. Al ser tan versátil, su vida útil se ha alargado incluso a esta década.



El robusto euroconector, un clásico desde la década de 1970

1.3. Vídeo digital en calidad estándar, alta definición y 3D

Ninguno de los sistemas analógicos de conexión anteriores sería capaz de mantener la calidad de una señal actual de vídeo digital con requerimientos de ancho de banda superiores, más si cabe si además es de alta definición o 3D, al poseer una cantidad mucho mayor de píxeles que transmitir cada segundo. Y es que el vídeo digital se compone de muestras de señal, cada una representa una porción minúscula de ésta.

La señal de vídeo digital, al igual que la analógica, se puede representar por tres señales de color independientes (R, G y B) o bien una señal de luminancia (Y) y dos de crominancia (U y V).

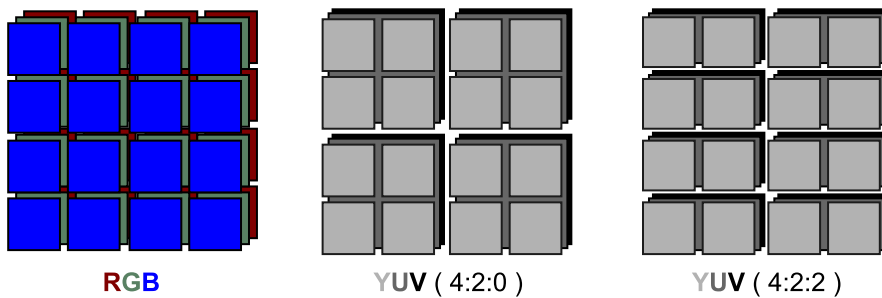
El proceso de digitalización de una señal analógica sigue siempre el **criterio de Nyquist**, es decir, se muestrea, al menos, al doble de la máxima frecuencia de la señal analógica. No entraremos en cálculos, pero un ejemplo sencillo nos puede ilustrar tan sencilla regla: en el caso del audio, si la frecuencia máxima de la señal se establece en 20.000 Hz (por limitaciones del oído humano, y esta suposición es muy optimista), se muestrea a más de 40.000 muestras/segundo. Por ejemplo, CD y DAT utilizan 44.100 muestras/segundo.

En el caso del vídeo, si llevamos a cabo la misma operación con la señal de vídeo en RGB, el volumen de datos que resulta es inmenso, pues para cada muestra de rojo R necesitamos otra muestra de verde G y una tercera de azul B, pero si utilizamos la señal YUV se reduce sensiblemente. Esto se debe a una pequeña argucia: el ojo humano tiene buena resolución en blanco y negro, pero pobre en color.

Mala resolución para el color

Un ejemplo esclarecedor es intentar adivinar el color de un único hilo de tejido que hayamos situado a varios metros: es posible que detectemos si hay un hilo, pero no su color.

Así, podemos **submuestrear** las señales de color U y V, es decir, generar menos muestras de las que realizaremos para la luminancia Y, sin que el usuario final se percate:



Representación visual de la asignación de muestras sobre píxeles según diferentes formatos

La proporción denominada 4:2:2 supone que por cada dos muestras de señal Y realizaremos una de U y otra de V, mientras que la 4:2:0 supone una muestra de U y otra de V por cada cuatro de Y. Todo depende de la calidad de color que se desee.

En todo caso, los flujos de datos generados son de tal volumen (centenares de Mbits) que necesitamos una nueva generación de sistemas de conexión, así destacan los SDI y HD-SDI para el entorno industrial y *broadcast* y los DVI y HDMI para el entorno doméstico.

1.3.1. SDI (Serial Digital Video)

El estándar de facto en el entorno profesional sigue la norma CCIR 601 que opera a 270 Mbps y contiene la señal de vídeo en calidad estándar 720x576 a 25 imágenes/segundo y además hasta 16 señales de audio digital **embebidas**, todo a través de un único cable coaxial con conector BNC y hasta una distancia de cientos de metros, ideal tanto para el estudio como para el despliegue de equipos en un gran acontecimiento.

La señal digital que transmite es por componentes (Y, U, V) digitalizados en proporción 4:2:2 que se alternan en el tiempo, de manera que el cable sólo transporta una de estas tres señales en cada instante de tiempo, así pues no existe interferencia alguna. Los equipos Betacam SP, Betacam Digital, Betacam SX, DVCAM y DVCPRO, entre otros, y todos los sistemas de enrutamiento y mezclado de imagen que existen se interconectan con este sistema si desean ser comerciales.

1.3.2. HD-SDI (High Definition SDI)

La evolución natural del estándar SDI, compatible con señales de alta definición, sigue la norma SMPTE 292M y opera a 1,5 Gbps pero es mecánicamente similar al anterior: un cable coaxial (evidentemente más caro) con conectores BNC. Tiene cinco veces la velocidad del anterior para acoger señales de 1920x1080 o 1440x1080 a 25 imágenes/segundo (modo denominado 1080i o 1080 entrelazado) o 1280x720 a 50 imágenes/segundo (modo 720p o 720 progresivo), siempre sin comprimir.

Formatos de vídeo como HDCAM o DVCPRO HD y todas las pantallas de alta definición industrial y *broadcast* se interconectan así.

1.3.3. 3G-SDI

Sin embargo, para señales HD-SDI en 3D se necesita el doble de velocidad de datos, ya que en *broadcast* el 3D se genera con dos flujos 1080i HD-SDI independientes que han de circular por un mismo cable (modo 1080p o 1080 progresivo), de manera que el nuevo estándar 3G (denominado así porque llega casi a 3 Gbps) ya aparece como sustituto natural del HD-SDI.

1.3.4. DVI (Digital Visual Interface)

En un entorno doméstico, las señales SDI o HD-SDI no son utilizadas al no necesitarse tal calidad, ya que para cuando la señal sale de su origen y llega a su destino a través de la red de transporte utilizada (por TDT, satélite, *streaming* por Internet, en soporte DVD o Blu-ray), está **comprimida**, es decir, se ha reducido su calidad visual para reducir su peso en bits y hacer comercial su difusión. En realidad, buena parte de esta asignatura versa sobre cómo se debe comprimir en todas esas redes de distribución.

Así pues, como soluciones de conexionado más sencillas contamos con el DVI, que puede tratar con señales de vídeo tanto digitales (estándar o alta definición) como analógicas, pero habitualmente sin el sonido asociado, que requiere un cableado aparte. La novedad es que es bidireccional, de manera que el equipo emisor y el receptor pueden acordar la resolución más adecuada para trabajar entre sí y posibilita también encriptar la señal, aunque raramente se usa en ese aspecto. Sin duda, es la evolución natural del VGA y se encuentra



El conector BNC llega al siglo XXI llevando HD y sonido multicanal en un solo cable.



Conector DVI

en la mayoría de monitores de vídeo del hogar, tarjetas gráficas e incluso en televisores planos, aunque en este último sector está perdiendo terreno frente a su evolución, el HDMI.

Equipos emisor y receptor

Un equipo Blu-ray que desee reproducir a 1080p conectado a una pantalla HD Ready reproducirá en calidad 1080i o 720p, ya que la pantalla comunicará vía DVI su incapacidad para presentar señales en 1080p, sin necesidad de ninguna acción por parte del usuario. Si, en cambio, la pantalla es Full HD 1080, la reproducción tendrá lugar en máxima resolución.

1.3.5. HDMI (High Definition Multimedia Interface)

El HDMI lleva tanto vídeo digital como señales de audio. Se basa en DVI mediante la adición de audio digital en el mismo cable, pero conserva la compatibilidad con la especificación eléctrica DVI-Digital, es decir, un mero adaptador convierte un conector HDMI en DVI, por lo que pierde el sonido. La protección anticopia se basa en una tecnología llamada HDCP (protección de contenido de gran banda digital), que obliga a que el equipo reproductor y el equipo monitor acuerden que sólo se visualizará la señal y no se ejecutarán copias. Si el acuerdo fracasa, se reproduce en baja calidad o sencillamente no hay reproducción posible.

De las diferentes versiones que han ido apareciendo en el mercado desde su nacimiento en el 2002, actualmente se trabaja con la 1.4, la única capaz de trabajar con señales HDTV-3D y que abre la posibilidad de trabajar con resoluciones superiores a 2.000x2.000 píxeles aún en estudio (Quad HD, Super Hi-Vision, Ultra HDTV).

1.3.6. Firewire

Si bien el origen de Firewire se asocia con la transmisión en serie de datos digitales en entornos informáticos, al trabajar con vídeo en estas plataformas también son utilizados, usualmente para transferencia de vídeo digital entre equipos.

Firewire

Originalmente fue denominado IEEE 1394 o, por parte de Sony, i.Link.

Firewire fue desarrollado por Apple en 1995 y actualmente existe en diferentes versiones según la velocidad de transferencia (400, 800, 1.600, 3.200 Mbps). Este soporte físico se utiliza con el protocolo DV para conectar grabadores de vídeo y cámaras domésticas e industriales a ordenadores e incluye comandos de control (como rebobinado, *play* o *stop*) y datos como el código de tiempo.



Conector HDMI



Arriba, conectores 400 (estándar y mini); abajo, conectores 800

1.3.7. USB (Universal Serial Bus)

Si por uso extensivo y versatilidad se tuviese que seleccionar un sistema de transmisión de señales, sin duda el USB sería el que centraría nuestra atención. Nacido a mediados de los noventa principalmente para el mercado informático, en la actualidad se extiende su uso en entornos móviles e incluso en sistemas industriales. Su resistencia, facilidad de mantenimiento, interoperabilidad entre equipos, altas prestaciones de velocidad y fiabilidad, así como una constante mejora en velocidad resultan idóneos para las crecientes necesidades de transferencias de datos: desde 1,5 Mbps (USB 1.0), pasando por los 12 Mbps (USB 1.1) y los 480 Mbps (USB 2.0), hasta llegar a los 4,8 Gbps en la revisión 3.0.



El omnipresente USB

1.3.8. Conclusión

A modo de conclusión, presentamos la siguiente tabla con un resumen de los estándares de transporte de vídeo más habituales.

Tabla resumen de los estándares de transporte de vídeo más habituales

Transporte de vídeo analógico	Ventajas	Inconvenientes
Vídeo compuesto	<ul style="list-style-type: none"> Gran número de dispositivos lo poseen en todo el mundo. Un único cable para transmitir la señal de vídeo. 	<ul style="list-style-type: none"> La señal de luminancia Y y las de crominancia U y V se interfieren entre sí al compartir el cable, lo que degrada la imagen resultante. Es analógico, por lo que queda fácilmente afectado por interferencias externas y ruido.
S-Vídeo	<ul style="list-style-type: none"> Gran número de dispositivos lo poseen en todo el mundo. La luma y la croma van por caminos separados, no se interfieren y da mejor señal de imagen. 	<ul style="list-style-type: none"> Conector mecánicamente más delicado. Sigue siendo analógico, por lo que se utiliza en exclusiva para distancias inferiores a los 2-3 metros.
Vídeo en componentes	<ul style="list-style-type: none"> Máxima calidad en analógico. Aceptable para edición y posproducción profesional. 	<ul style="list-style-type: none"> Añadiendo las señales de sincronía H y V, son necesarios hasta cinco cables. Todos ellos han de ser exactamente iguales o los retardos no coinciden y aparecen artefactos en la imagen. Exclusivamente para distancias cortas.
Transporte de vídeo digital	Ventajas	Inconvenientes
SDI / HDSDI / 3G	<ul style="list-style-type: none"> Señal digital, por lo que es poco sensible al ruido y las interferencias. Acepta múltiples generaciones de copiado. El audio, el código de tiempo y otras señales van incrustadas en la misma señal de vídeo. De fácil conexión con un solo cable. Es un estándar generalizado en entornos industriales y <i>broadcast</i>. 	<ul style="list-style-type: none"> Requiere equipamiento industrial o <i>broadcast</i> de alto coste. Los 270 Mbps que genera no son de fácil gestión ni almacenamiento.
DVI	<ul style="list-style-type: none"> Muy compatible, pues puede trabajar con señales analógicas o digitales. En el caso digital, puede trabajar con cableado de hasta 50 metros. 	<ul style="list-style-type: none"> Dedicado exclusivamente para visualización de vídeo, no a la interconexión de equipos. Es un conector multipin, por lo que es complejo y puede sufrir torsión. El audio necesita de cableado aparte.

Transporte de vídeo analógico	Ventajas	Inconvenientes
HDMI	<ul style="list-style-type: none"> • Conector compacto. • Incluye la señal de audio. • Mediante extensores, puede llegar a los 250 metros. • Un sencillo adaptador lo convierte en DVI. • Incluye protección anticopia HDCP. 	<ul style="list-style-type: none"> • Existen varias versiones del HDMI con capacidades de transmisión diferentes. • El conector no posee anclaje, de modo que es fácilmente desconectable. • Si bien podría utilizarse para transferir vídeo entre equipos, el sistema anticopia es un <i>handicap</i> que lo relega a conexiones a monitores digitales.
Firewire	<ul style="list-style-type: none"> • Conector sencillo y compacto. • De uso universal, aunque es el más adoptado para la transferencia de audio y vídeo a alta velocidad en entornos domésticos y semiprofesionales. • Mayor transferencia sostenida de datos que el USB, necesaria para vídeo. 	<ul style="list-style-type: none"> • No posee anclajes. • Es propietario (Apple). • No está tan extendido como el USB.
USB	<ul style="list-style-type: none"> • Conector sencillo y compacto. • De uso universal. • Se utiliza para fuentes de vídeo de baja velocidad, como webcams y escáneres. 	<ul style="list-style-type: none"> • Orientado a la interconexión de periféricos. • Rendimiento lento para vídeo de alta definición. • No posee anclajes.

Actividades

Actividad 1

Revisad el equipo informático que tenéis actualmente. ¿Qué tipo de entradas y salidas pueden ser interesantes para capturar vídeo?

Actividad 2

Revisad las entradas de señal del monitor. ¿Qué señales de vídeo acepta? Pensad que, si tenéis un portátil multimedia, puede adelantarse que la entrada es de componentes digitales RGB, aunque conectada directamente a la placa gráfica del portátil.

Actividad 3

Si tenéis una cámara de vídeo o webcam, ¿con qué estándar de conexión se conecta?

Actividad 4

Si tenéis software de captura (como Adobe Premiere, Camtasia o software gratuito), ¿es compatible con las entradas disponibles?

Actividad 5

Ahora revisad el televisor del comedor. Si es de reciente adquisición, el número de entradas que posee suele ser apabullante. ¿Cuáles son compatibles con el equipo informático? Por lo general, se acaba creyendo realmente en la tantas veces pronosticada conversión de la televisión en un centro multimedia digital.

1.4. Broadcast frente a vídeo en Internet

Hasta ahora, en este módulo nos hemos dedicado a las señales de vídeo tradicionales, las que pueden ser entregadas directamente a una pantalla de vídeo o a un ordenador a través de un simple cable, a distancias no superiores a los 100 metros. Por el contrario, los contenidos de vídeo transmitidos o compartidos a través de Internet han de ser obligatoriamente digitales (si no se capturan

previamente) y requieren un dispositivo (como un ordenador o un descodificador dedicado) o un software específico para reconstruirlas correctamente antes de ser visibles.

Todos los ficheros o *streams* de vídeo de la Red destinados a los consumidores finales **se comprimen** mediante alguno de los muchos tipos de compresión disponibles en la actualidad, por lo que ambas señales de audio y vídeo deben ser **descomprimidas** mediante hardware o software antes de poder ser entregadas al destinatario. Si se tiene TDT, televisión por cable o por satélite, se descomprime por hardware, mientras que si se tiene un equipo Windows Media Center o sencillamente se es usuario de YouTube, por software.

Ved también

Los diferentes tipos de compresión se tratan en el módulo "Digitalización, almacenamiento y transmisión de audio y vídeo".

En el caso del software, el más interesante y con mayor futuro, concretemos que hay pequeñas aplicaciones denominadas extensiones o **plugins** que, integradas en el programa navegador de uso habitual lo capacitan para ejecutar esa descodificación al vuelo: Adobe Flash Player, QuickTime de Apple y Windows Media Player, por citar algunas de las más populares. Las señales de vídeo por Internet pueden ser esencialmente de cualquier número de píxeles horizontales y verticales y hay una amplia selección de sistemas de protección de contenidos y técnicas de entrega de los datos a través de la red.

Con todas las distintas combinaciones disponibles, los proveedores de vídeo por Internet son capaces de desarrollar sus propias combinaciones de formatos, por lo que el vídeo producido en un sitio puede no ser directamente compatible con otros sitios de alojamiento. Por ello, la flexibilidad es un elemento clave de los reproductores de vídeo de Internet: los televidentes deberían tener libertad para detener, iniciar, rebobinar y pausar la reproducción en cualquier momento, controlar el tamaño del vídeo en sus pantallas y elegir entre una variedad de formatos de pantalla del reproductor. Cuando queráis distribuir vuestro producto audiovisual por la Red, no os aferréis a un estándar en particular porque reduciréis el impacto.

En una red como Internet, pensada para la transmisión de datos, las señales de vídeo no son fáciles de transportar, básicamente por dos razones:

a) La primera es que no es una red pensada para vídeo, el tránsito a través de la red del vídeo y audio comprimidos se realiza mediante paquetes de datos genéricos (paquetes IP), como todos los demás tipos de datos (textos, imágenes) que fluyen por la Red. Estos paquetes de datos circulan por las mismas redes y son encaminados por los mismos centros de distribución (*routers*) como cualquier otro dato, a veces desordenándose o perdiéndose.

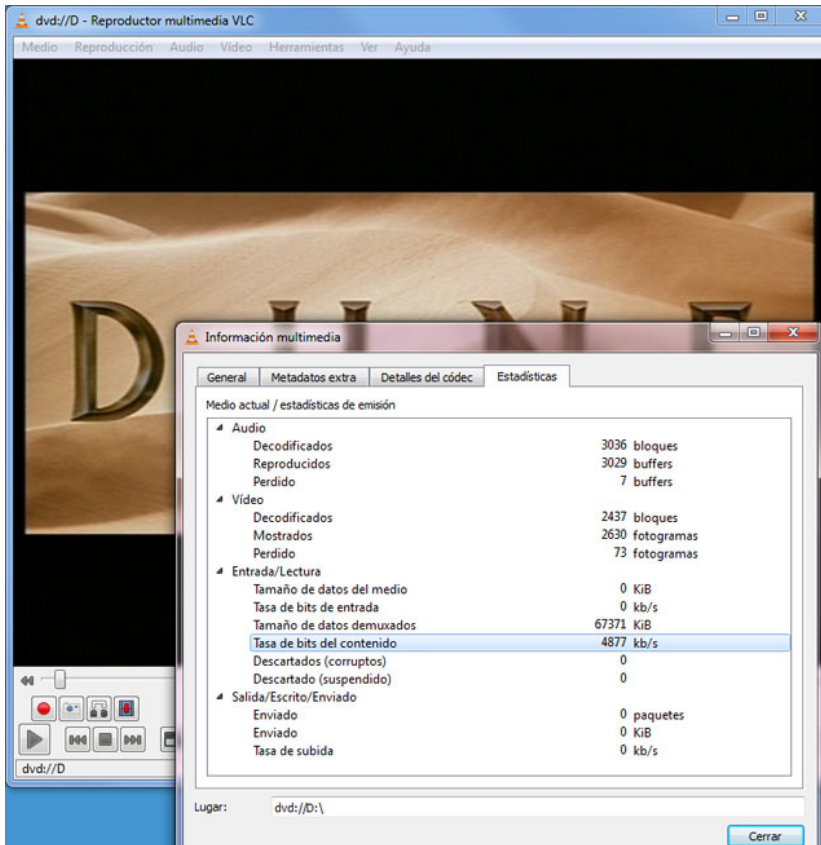
b) La segunda razón es un flujo de datos muy alto, casi constante, por lo que necesita gran cantidad de paquetes necesarios para asegurar la reproducción con la cadencia necesaria: por ejemplo, un vídeo de YouTube de un minuto puede exigir la entrega de unos mil paquetes sin perder el orden y con una cadencia fija. Todo esto lo veremos con detalle en otros módulos.

Sin duda, entre el vídeo por Internet y el vídeo *broadcast* hay diferencias críticas en su distribución que obligan a retos técnicos en parte superados. Sin embargo, más allá de la tecnología, hay una revolución más profunda y de consecuencias más intensas: el usuario elige lo que desea ver desde el punto en el que desea verlo, en contraposición a ver lo que ofrecen tal como indica la programación estática del canal. Esta nueva capacidad de decisión por parte del usuario final obliga a replantear la atracción del producto y a crear nuevos mercados en los que aprovechar la interacción. Esta revolución ya está pasando, por mucho que se niegue en algunos sectores, aferrados a argumentos tales como la falta de calidad de imagen o de tratamiento profesional asegurado de los contenidos de la Red, pero estos argumentos van perdiendo fuerza año tras año al incorporarse profesionales preparados para trabajar en este medio como su entorno natural. Esperemos aportar nuestro grano de arena.

Actividades

Actividad 6

Comparemos los niveles de compresión que se utilizan habitualmente en el mercado. Para ello, vamos a instalar el programa de reproducción de vídeos VLC (<http://www.videolan.org>).



Tasas de bits informadas por VLC

Primero veamos la compresión de vídeo utilizada para poder almacenar una película en un DVD doméstico; para ello, mediante el menú Medio-Abrir disco, seleccionad la unidad de disco y reproducid una película. Accediendo en cualquier momento al menú Herramientas-Información multimedia-Detalles del códec donde en el apartado Vídeo indicará el formato de codificación, usualmente MPEG-2, el ancho y el alto de la imagen en píxeles y los fotogramas por segundo. Es fácil calcular, pues, que el flujo de datos de lectura en bits/s debería ser, sin compresión, de:

$$\text{Flujo de datos} = \text{ancho (píxeles)} \times \text{alto (píxeles)} \times 8 \frac{\text{bits}}{\text{color}} \times 3 \text{colores} \times \frac{\text{fotogr.}}{\text{seg.}}$$

Ahora, en la solapa Estadísticas, podéis ver la Tasa de bits del contenido, los bits por segundo necesarios para reproducir correctamente el fichero comprimido:

4 Entrada/Lectura	
Tamaño de datos del medio	9079 KiB
Tasa de bits de entrada	973 kb/s
Tamaño de datos demuxados	6801 KiB
Tasa de bits del contenido	1030 kb/s
Descartados (corruptos)	0

Tasas de bits en la solapa Estadísticas de VLC

La compresión que ha conseguido, usualmente muy alta, se calcula como:

$$\text{Factor de compresión} = \frac{\text{flujo de datos sin comprimir (bps)}}{\text{flujo de datos comprimido (bps)}}$$

Actividad 7

Realizad la misma operación, pero en este caso con un fichero de vídeo que se haya descargado de Internet (WMV, MP4, AVI). Para visualizarlo con VLC, seleccionad el menú Medio-Abrir archivo y recoged los datos que presenta. Calculad el factor de compresión y comparadlo con el del caso DVD. ¿Por qué razón hay tal diferencia?

Si en vez de DVD se utiliza Blu-ray, comparadlo con un fichero de vídeo que sea en alta definición o la comparativa no tendría sentido.

1.5. Resoluciones, velocidad de cuadro, relaciones de aspecto y otros parámetros

El **cuadro** es la unidad de imagen en vídeo, una imagen con todo el contenido visual captado.

Sin embargo, a la hora de presentarlo en pantalla, puede hacerse de dos maneras:

- **En modo progresivo:** por completo, es decir, presentando todas y cada una de sus filas de píxeles una debajo de la otra hasta finalizar la presentación del cuadro.
- **En modo entrelazado:** el cuadro se descompone en dos subimágenes, los campos, uno con las filas impares y el otro con las pares. A continuación, se presenta el primer campo y luego el segundo.

Modo progresivo

En Europa, los modos 720p, 1080p y 1080p50, a 25, 24 y 50 cuadros/segundo o Hz respectivamente.

Modo entrelazado

En Europa, los modos 576i, PAL y 1080i, a 50 campos/segundo o Hz (25 cuadros/segundo).

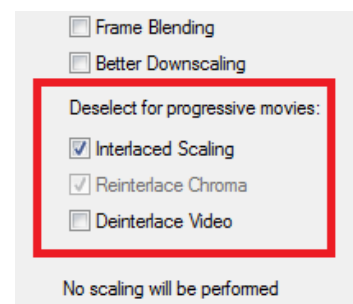
Por historia, el modo entrelazado posibilitó emitir televisión utilizando un ancho de banda pequeño (era más sencillo emitir dos subimágenes alternativamente que una gran imagen), pero a costa de generar un efecto de vibración visual en las zonas de imagen con detalle brillante, denominado *flicker*. Algunas personas son más receptivas a este problema que otras.

La técnica más habitual para reducir el *flicker* de una señal entrelazada ha sido simular con mayor o menor gracia una frecuencia de cuadro mayor en el receptor: un monitor que prometa 100 Hz, 200 Hz e incluso 600 Hz intenta mejorar la experiencia visual, a la vez que reduce el *flicker*, pero introduce artificios varios que, en general, generan otros problemas visuales colaterales. En vídeo digital, que es nuestro campo de trabajo, todo software de tratamiento de vídeo provee de alguna herramienta para reducir los efectos visuales del entrelazado.

Es curiosa la existencia de la nueva frecuencia de cuadro de 24 imágenes/segundo en televisión digital. Hasta ahora, toda transferencia de material cinematográfico a vídeo había sido a costa de perder calidad visual. Si hablamos de película en 16 mm o 35 mm, rodada a 24 imágenes por segundo, al transferirla a vídeo con una cámara de 25 imágenes/segundo aparecían transiciones visuales oscuras constantemente, por lo que se optó por reproducirla a 25 imágenes/segundo y acelerar así la acción en unos 4 minutos de cada 100 (una película de 100 minutos en el cine dura 96 en televisión). Gracias a la intro-



El rayado en los perfiles de los objetos es indicativo de entrelazado.



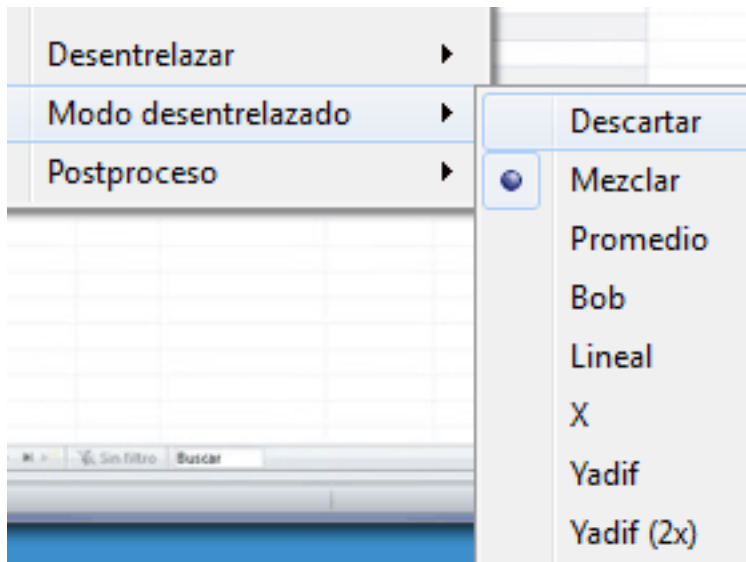
Típicas opciones contra el entrelazado

ducción de la frecuencia de vídeo de 24 imágenes/segundo en los formatos 720 y 1080, estos fenómenos desaparecen y se puede capturar, almacenar y reproducir digitalmente el material de celuloide tal como se hizo.

El modo entrelazado y el progresivo conviven en la actualidad. En realidad, cualquier sistema que supere los 18 cuadros/segundo es factible, pero **cuanto mayor es la velocidad de cuadro más realista es la experiencia**, aunque requiera más velocidad de datos. En casos específicos como la videoconferencia, sistema puramente progresivo, este límite es aconsejable aunque, si no se cumple, no pierde el valor añadido de una comunicación a gran distancia por muy bajo coste.

Actividad 8

Abrid el VLC y, mediante el menú Medio-Abrir archivo, visualizad los vídeos indicados en el aula. Este programa posee capacidad de desentrelazar utilizando diferentes técnicas enumeradas en el menú Vídeo-Modo desentrelazado.



Opciones de desentrelazado de VLC

Una vez seleccionada una de éstas, se puede aplicar y ver el resultado mediante Vídeo-Desentrelazar-Activar. Seleccionad el que se crea más adecuado para cada vídeo.

Puede desactivarse mediante Vídeo-Desentrelazar-Desactivar.

Respecto a la **resolución de imagen**, solemos hablar de la resolución vertical, que se define como el número de líneas horizontales de imagen que pueden distinguirse al recorrer en vertical la imagen.

Así, en PAL o DVD son 576 líneas de un total de 625 (el resto son líneas de sincronía sin imagen) y en Blu-ray son 720 (modo 720p) y 1.080 (modos 1080i y 1080p).

La videoconferencia

En la videoconferencia, por ejemplo, se utiliza a menudo el formato denominado CIF (Common Intermediate Format), de 352 píxeles de ancho y 288 líneas de resolución

vertical, el 4CIF (que cuadriplica este tamaño hasta llegar a 704x576) o el QCIF (un cuarto del CIF, 176x144).

Sin embargo, la resolución de la imagen vendrá limitada a continuación por la resolución de la pantalla sobre la que se muestre y viceversa. Así, tamaños de 640x480 (el ya casi inexistente VGA) hasta 1600x1200 (UXGA) son los más habituales en el entorno informático, mientras que resoluciones de 1920x1080 son las propias de monitores Full HD domésticos, aun cuando estos dos segmentos de mercado parecen tender a un destino común (la unificación augurada del televisor doméstico como centro digital).

Finalmente, el tercer parámetro que debemos destacar es la **relación de aspecto**, que indica la proporción entre el ancho y el alto de la imagen.

En la actualidad, el estándar en vídeo es 16:9 (más cómodo que decir 1,78). Estudios han indicado que este rectángulo podía abarcar los formatos de aspecto más frecuentes que han existido hasta ahora (4:3 convencional, 1.85:1 y 2.39:1 cinematográfico, 3:2 fotográfico), ya sea utilizando el zoom o mediante barras negras a los lados (*pillarbox*) o arriba y abajo (*letterbox*).

David Sarnoff Center

Estos estudios fueron dirigidos en 1980 por el doctor Kerns Powers del David Sarnoff Center y miembro del SMPTE.



Comparativa de los diferentes formatos de imagen

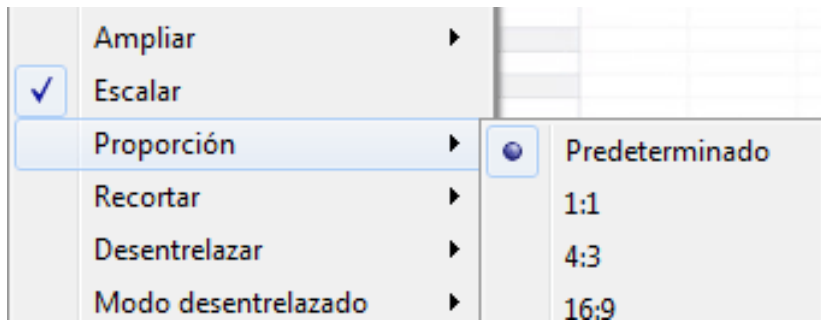
La relación de aspecto está relacionada con la proporción del píxel: en el mundo de la informática, los píxeles son cuadrados. Así, una imagen 4:3 de 1024x768 (VGA) tiene una proporción exacta de 4:3 y una imagen 16:9 de 1600x900 tiene una proporción de 16:9. Sin embargo, en el mundo de la imagen no es así: el vídeo convencional de 720x576 es 4:3 pero la relación de sus píxeles no:

$$\frac{720}{576} = \frac{5}{4} = 1,25$$

Con frecuencia, trabajar con señales de resolución estándar en entornos informáticos supone tener que encoger horizontalmente la imagen para verla en la proporción correcta, cosa que los propios programas de captura ya realizan internamente. Este problema no existe con las señales de alta definición, ya que la mayoría de sus formatos son de píxel cuadrado: 1280x720 (720p) y 1920x1080 (1080i y 1080p) (16:9). El formato 1440x1080 (1080i) es la excepción, con un píxel representado en pantalla ligeramente rectangular.

Actividad 9

Mediante VLC, se puede experimentar con los dos formatos habituales de trabajo en vídeo digital. El menú es Vídeo-Proporción, cuyos valores van desde Predeterminado (el programa elige la proporción según los datos del propio fichero de vídeo) hasta proporciones más rectangulares. Comprobad que los vídeos en alta definición siempre son en proporción 16:9.



Selección de proporciones en VLC

2. Características del sonido

La distribución de imagen en casi cualquier aspecto va asociada a la distribución de audio, por ello requiere tratarlo en este apartado. Como el audio tiene características diferentes a la imagen, su tratamiento es también diferente, en general más sencillo. Comencemos con el audio analógico y luego el digital para ver los estándares de compresión en el siguiente apartado.

2.1. Audio analógico

Convertido el audio en una señal eléctrica, ésta puede alimentar directamente el equipo, el altavoz o los cascos, por ejemplo, donde se vaya a reproducir. Salvando la potencia necesaria, este sistema de audio abarca desde micrófonos hasta sistemas de bafles de conciertos con éxito desde hace décadas, sobreviviendo en la era digital a través de las salidas RCA de los equipos CD, DVD, decodificadores de cable y satélite, o de las salidas de auriculares de 3,5 mm de todo reproductor MP3. Sin embargo, el ruido, las interferencias, la calidad de los conectores y del cable hacen mella en su resultado final; además, no incluye señalización que facilite su sincronía con el vídeo que pueda acompañar, por lo que se circunscribe actualmente al entorno de consumo.



Conectores de audio estéreo analógico: minijack (arriba) y pareja RCA (abajo)

2.2. Audio digital

El sonido siempre será analógico, pero la transmisión y almacenamiento se han digitalizado, por lo que así lo han protegido del ruido y lo han hecho inmune a las interferencias, a la vez que lo compatibiliza con los sistemas digitales. Además, del CD al MP3, el audio digital ha pasado de ser no comprimido a utilizar técnicas de compresión que veremos más adelante. Necesita conectores Canon XLS, aunque el formato AES también es el utilizado para empaquetar el audio digital en la señal de vídeo SDI, HD-SDI o 3G, transmitiendo todo el conjunto por un único cable coaxial.

La señal de audio digital sigue el formato AES/EBU, con cada muestra de 16 bits y a velocidades de 32.000, 44.100 y 48.000 muestras según el ancho de banda analógico que se desee representar (hasta 16 kHz, calidad FM; hasta 22 kHz, calidad CD, y hasta 24 kHz, calidad de estudio profesional, respectivamente). Sin duda, sigue el mencionado criterio de Nyquist. Así, una señal estéreo digital AES supone un flujo de datos de:

Flujo de audio = 44.100 muestras x 16 bits/muestra x 2 canales = 1,41 Mbps

Ved también

Las técnicas de compresión se tratan en el módulo "Digitalización, almacenamiento y transmisión de audio y vídeo" de esta asignatura.



Conectores Canon de audio digital (AES)

Si además contamos con que el estándar AES reserva 32 bits para transportar cada muestra y otros datos asociados (como el código de tiempo), resulta lo siguiente:

Flujo de transporte de audio = flujo de audio x 2 = 2,82 Mbps

Eso supondría casi una conexión de 3 Mbps sólo para reproducir el audio estéreo a través de una red, el caso de audio multicanal o envolvente multiplica por 3 o 6 esta cifra. Evidentemente, se ha de poder comprimir sin perder gran calidad para poder distribuir audio por red, cosa que veremos más adelante.

Actividad 10

Mediante VLC, abrid un fichero de audio y, mientras está en reproducción, acceded a Herramientas-Información multimedia. En la pestaña Detalles del códec indicará el formato en el que está comprimido el audio y en la pestaña Estadísticas se presentará una serie de datos de los que en especial destaca la Tasa de bits de entrada, es decir, los kilobits por segundo que el programa va leyendo del fichero de audio, la Tasa de bits de contenido, que indica cuántos de los kilobits por segundo leídos realmente son para audio (los demás son de control, calidad, sincronía y demás).

Entrada/Lectura	
Tamaño de datos del medio	1504 KiB
Tasa de bits de entrada	128 kb/s
Tamaño de datos demuxados	1458 KiB
Tasa de bits del contenido	125 kb/s
Descartados (corruptos)	0

Datos proporcionados por VLC de un fichero de audio

Comparad la tasa de bits de contenido con el flujo de audio (no el flujo de transporte de audio) que debería tener en formato AES para dar una idea de la compresión conseguida con el formato de audio del fichero perdiendo la calidad de sonido que se pueda detectar.

Probadlo con diferentes formatos (por ejemplo MP3, WMA o WAV).

3. La revolución en la gestión de las señales

Una vez hemos caracterizado las señales principales, nos quedan otras informaciones adicionales que tienen definido su sitio en el entorno doméstico y profesional, como la audiodescripción para personas con discapacidad visual, el teletexto de información general o los subtítulos para el colectivo con diferentes grados de discapacidad auditiva. Estas señales deberían poder subsistir o tener equivalentes funcionales en cualquier plataforma de distribución digital, siempre a costa de un consumo mayor de ancho de banda, pero los beneficios para la audiencia son evidentes. Incluso en ciertos países son de obligada existencia.

La gestión de este cúmulo de información ligada (vídeo, audio, servicios adicionales) en una productora audiovisual se realiza mediante equipos denominados **matrices o enrutadores** de vídeo y audio, que seleccionan las señales adecuadas del conjunto de señales existentes para conseguir un producto resultante coherente. Sin entrar en detalles, han existido matrices analógicas y actualmente todas son digitales.

Al tener que trabajar con señales totalmente digitales, las condiciones son las mismas que las de trabajo de un *router* IP, aparato concentrador de datos que encamina los paquetes digitales de datos desde un origen hasta un destino. La fusión entre las matrices digitales de audio/vídeo y los *routers* IP no ha tardado en llegar, habida cuenta de las funciones avanzadas de los *routers* IP de última generación en entornos de muy alta velocidad. Así, las instalaciones tradicionales de audio y vídeo con cables coaxiales dedicados desde cada equipo hasta la matriz van desapareciendo y se sustituyen por una red de alta velocidad en la que se conectan todos los equipos al *router* IP central. Es una red escalable, es decir, de tamaño fácilmente ampliable si se requieren nuevos equipos para la producción.

Este cambio sustancial en el núcleo de producción unifica además la gestión del resto de señales, incluyendo la telefonía interna, la gestión remota de los equipos, la integración de la producción en la red de distribución al cliente, a otros centros de producción o directamente al mercado Internet. Sin duda, hay otro tipo nuevo de problemas, entre ellos la lucha contra el retardo de los datos y la mejora de la robustez del sistema ante la congestión de la red y su consecuente pérdida de paquetes de datos.



Una matriz de vídeo y audio que presenta las conexiones de entradas y salidas.

Si en el campo de la distribución audiovisual la Red ya tenía ganada la batalla, en el campo de la producción audiovisual la victoria es clara a medio plazo. Muchos equipos *broadcast* ya incluyen conexiones de red (fibra o cobre) de alta velocidad (4 Gbps en adelante) preparados para este cambio de proceso. El transporte de vídeo y audio por IP irá en aumento vertiginoso.

4. Codificación de la imagen fija

4.1. La necesidad de comprimir

Las imágenes requieren considerablemente más espacio que el texto y el audio y el vídeo aún demandan más en términos de almacenamiento y velocidad de transmisión. Unos pocos números nos ilustran en este aspecto e introducen así algunos conceptos básicos que vamos a usar a lo largo de todo el curso a partir de este punto:

- Una página de este libro contiene unas 35 líneas de 80 caracteres; si cada carácter puede representarse, almacenarse y transmitirse digitalmente con 1 byte (8 bits), cada página ocupa 22.400 bits.
- Una imagen de vectores (es decir, creada con líneas, como los diseños de Autocad o una fuente escalable) contiene unos 500 trazos. Si cada trazo se representa por sus coordenadas de inicio y final (cuatro números en total) y cada número por 2 bytes, la imagen ocupa 32.000 bits.
- La música en calidad CD supone 44.100 muestras/segundo, si las muestras son de 16 bits y es estéreo, se generan 1.411.200 bits cada segundo.
- Una imagen fotográfica de 720 por 576 píxeles contiene 414.720 píxeles en total, si la imagen es en blanco y negro cada píxel se representa por un byte, por lo que ocupa 3.317.760 bits.
- Si la imagen anterior fuese en color, cada píxel se representa por 3 bytes (uno para el rojo, otro para el verde y otro para el azul), por lo que ocupa el triple, 9.953.280 bits.
- El sonido digital envolvente puede llegar a suponer casi 9 Mbps sin tratamiento alguno.
- En el caso del vídeo, una señal PAL de 625 líneas a 25 cuadros/segundo se descompone en luminancia y dos señales de crominancia que necesitan unos 216 Mbps.
- Si es HD a 1080i, aumenta a 1.15 Gbps.

Los requerimientos de memoria y velocidad de transmisión para estas informaciones deben reducirse de manera notable para que puedan ser gestionados por los equipos actuales, con capacidades de almacenamiento del orden de

Gbytes y memorias buffer del orden de Mbytes, y de sistemas de transmisión en el entorno de algunos Mbps. Necesitamos comprimir y, en algunos casos, de manera drástica.

4.1.1. Codificación y compresión. Tipos de compresión

Es usual que se utilicen los términos codificación y compresión de manera equivalente, pero **codificación** es sencillamente expresar una información de una manera diferente utilizando un nuevo código, mientras que **compresión** es reducir su tamaño total en bits. La compresión es un caso particular de codificación, donde el nuevo código utiliza menos elementos para representar la misma información.

Así, la tabla ASCII codifica los caracteres y los expresa en formato binario, sin existir ninguna compresión en el proceso. Por el contrario, una codificación que utiliza una herramienta informática que represente un fichero de datos con otro de menor tamaño en bits (como .zip o .rar en Windows, .sit en OSX o .tg en Linux) es un caso de compresión.

No obstante, aún podemos perfilar más la definición de compresión: cuando indicamos que ha de representar la misma información, no significa en todos los casos que deba ser exactamente la misma información final que la original, es decir, la compresión puede ser **sin pérdidas** o **con pérdidas**. En el primer caso, podemos realizar el proceso inverso para volver a tener la misma información exacta original: de todo fichero .zip o .rar puede recuperarse el fichero original. En el segundo caso, aceptamos una degradación del mensaje en aras de conseguir una compresión efectiva, siempre y cuando no haga inservible la información que lleva.

4.1.2. Redundancia y entropía

¿A qué nos referimos con información inservible? En muchos casos, la información por transmitir tiene **datos redundantes**, información predecible o repetitiva que se puede comprimir de forma intensa, como por ejemplo varios espacios en blanco consecutivos en un texto o varios píxeles consecutivos de un mismo color en un fax. Su reconstrucción al descomprimir es exacta. En la compresión sin pérdidas se elimina la redundancia de la información, sin tocar un ápice la información fundamental. Es la ideal para ficheros de datos (como zip, rar, sit y tg), algunos formatos de imagen como GIF, PIC, un modo de trabajo específico del formato de compresión de imagen JPEG 2000, entre otros.

Sin embargo, estas compresiones sin pérdidas no rinden mucho. Conseguir un 50% de reducción es generalmente un éxito, por lo que la compresión con pérdidas es la que domina en el mundo de la imagen y el sonido, donde se necesitan tasas de reducción mucho más altas.

La base de la compresión con pérdidas es la eliminación de los **datos irrelevantes**, como por ejemplo las pequeñas variaciones de color entre píxeles cercanos (ya que el ojo humano tiene poca resolución para el color) o los sonidos de frecuencias superiores a los 16.000 Hz (pues pocos seres humanos adultos son capaces de oírlos). La información irrelevante se suprime completamente, al modificar de manera permanente la información por transmitir, aunque de una manera tolerable para el destinatario al que vaya dirigida. Ejemplos de formatos estándar de compresión con pérdidas son los siguientes:

- formatos de **compresión de imagen** como JPEG, EZW, SPIHT y casi todos los modos de trabajo de JPEG 2000,
- formatos de **compresión de audio** como AAC o MP3,
- formatos de **compresión de vídeo** como la familia MPEG.

Es evidente que el éxito de cualquier compresión dependerá de cuánta información redundante o irrelevante haya en el mensaje, por lo que es habitual que el rendimiento de la compresión varíe casi constantemente. La entropía es una medida de la información esencial de un mensaje (la diferencia entre la cantidad total de datos que contiene el mensaje y su redundancia). A mayor entropía, más datos no pueden eliminarse, por lo que la codificación con pérdidas rendirá poco. A menor entropía, la codificación con pérdidas será muy efectiva.

4.1.3. Medida de la compresión de imagen

Para medir la compresión conseguida en una imagen y así poder comparar diferentes sistemas de compresión como haremos en el curso, utilizamos el factor de compresión que ya vimos antes, un valor numérico que aumenta con la compresión lograda. Así, dada una imagen original y su equivalente comprimido, el factor de compresión es el siguiente:

$$\text{Factor de compresión} = \frac{\text{Peso de la imagen "original"}}{\text{Peso de la imagen "comprimida"}}$$

Otra manera de representar la compresión es indicando el número de bits que necesita cada píxel de la imagen, el bit por píxel (bpp):

$$\text{bpp} = \frac{\text{Peso de la imagen (expresado en bits)}}{\text{Número total de píxeles de la imagen}}$$

Nota

Recordad indicar ambos pesos siempre en la misma unidad (bits, bytes, Kbytes).

De manera sencilla podemos deducir que el bpp de una imagen original y el bpp de su imagen comprimida se relacionan con el factor de compresión entre ellas:

$$\text{Factor de compresión} = \frac{\text{bpp de la imagen "original"}}{\text{bpp de la imagen "comprimida"}}$$

Actividad 11

Si una imagen es de 1600x1200 píxeles con 24 bits de color/píxel y se comprime en un formato tal que el fichero resultante es de 230 Kbytes, ¿cuál es el bpp de la imagen original? ¿Y el de la comprimida? ¿Cuál será el factor de compresión conseguido? Atención con las unidades, no mezcléis bits con bytes.

4.1.4. Medida objetiva y subjetiva de la calidad de imagen

En la actualidad, se siguen investigando técnicas cuantitativas que puedan asignar un valor numérico a la calidad visual de cualquier imagen, así nos da una idea de lo degradada que ha resultado tras aplicarle cierta compresión con pérdidas y nos ayuda a elegir el estándar de compresión más adecuado.

En todo caso, presentamos una medida, el **error cuadrático medio**, una manera de calcular las diferencias que existen entre una imagen y su comprimida. Básicamente es una aproximación para diferenciar en cuánto son diferentes, algo que el ojo humano hace de manera natural en instantes pero que matemáticamente es muy difícil de concretar, pues no se conocen con detalle los mecanismos del cerebro humano que posibilitan tal hazaña. La idea es restar píxel a píxel una de otra y sumar las diferencias, el resultado es un valor que, cuanto más se acerca a cero, más parecidas son las imágenes.

Sin embargo, es una medida no del todo fiable, pues los errores visuales son percibidos por el ojo humano en mayor medida en función del entorno en el que aparecen, cosa que no se valora con esta herramienta. Tras ver el formato JPEG, veremos una manera mucho más visual y entretenida de detectar los errores mediante la aplicación Photoshop, capaz de calcular la imagen diferencia, es decir, una resta visual píxel a píxel entre, por ejemplo, una imagen sin comprimir y su equivalente comprimido, lo que pone en evidencia la información perdida en la compresión.

4.2. EL JPEG

Comencemos viendo el primer **sistema de codificación de imagen fija** e introduciendo una serie de conceptos que utilizaremos a lo largo de todo el curso. El JPEG es un comité (Joint Photographic Experts Group), no un estándar de codificación, que a finales de la década de 1980 ideó un proceso por el que una imagen digital podía comprimirse y mantener un compromiso con la calidad visual. En aquellos años ya comenzaba a ser necesario algún sistema para que las imágenes no ocupasen los reducidos espacios de almacenamiento digital existentes en la época y para poder ser transmitidas por líneas de telefonía

digital de baja velocidad. En general, el proceso es suficientemente robusto y de resultados aceptables, en especial con fotografías naturales e imágenes realistas, con lo que se llega a un promedio de compresión de 10:1 con pérdidas imperceptibles visualmente, por eso sigue siendo un formato muy utilizado.

Por primera vez en este curso, hemos de distinguir entre el sistema de codificación (código) de la imagen, que es JPEG, y el **contenedor** o formato de fichero donde se guardan los datos resultantes de la codificación, del que hay dos tipos, JFIF y EXIF. Así, JPEG/JFIF es el formato clásico, mientras que JPEG/EXIF posibilita añadir **metadatos**. En la actualidad, algunos móviles con cámara y GPS integrados añaden la longitud y latitud en la que una foto fue tomada. Nuestro interés se centra en el código.

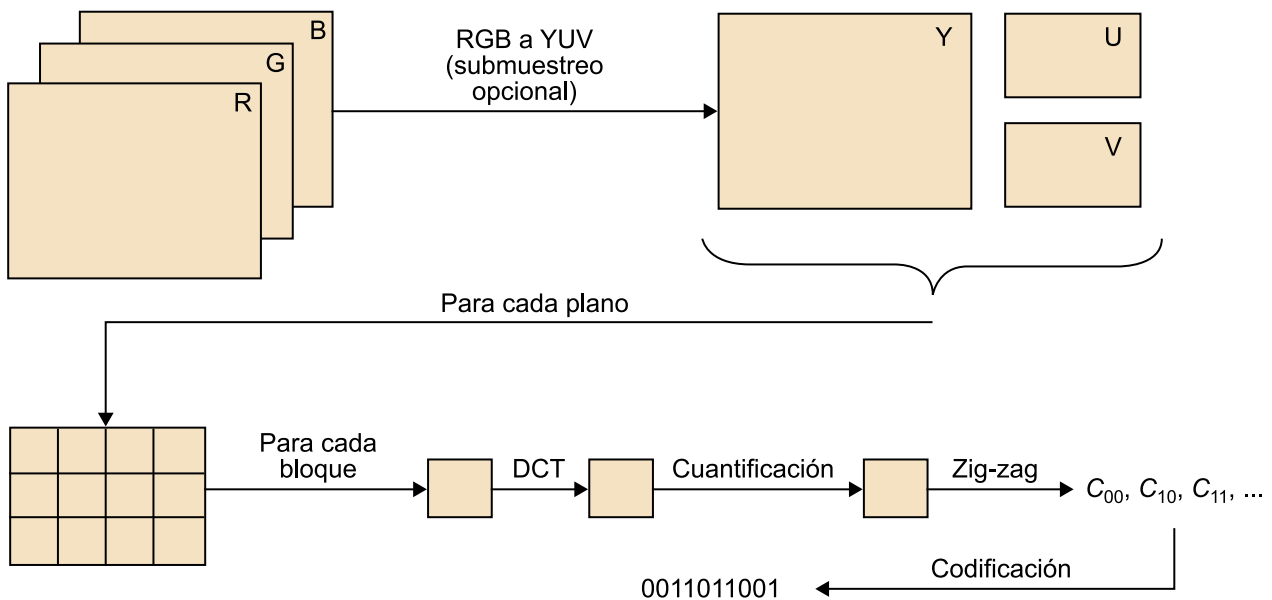
Los metadatos

Son los datos adicionales asociados a la imagen como, por ejemplo, la fecha y hora de la captura, la velocidad del obturador, el nombre de la cámara o los datos de configuración.

El JPEG puede trabajar en varios modos diferentes, aunque sólo vamos a detallar dos modos **con pérdidas**: el más utilizado, el **modo secuencial**, con el que introduciremos varios conceptos de codificación relevantes en cada paso del proceso de codificación, y el **modo progresivo**.

4.2.1. El modo secuencial

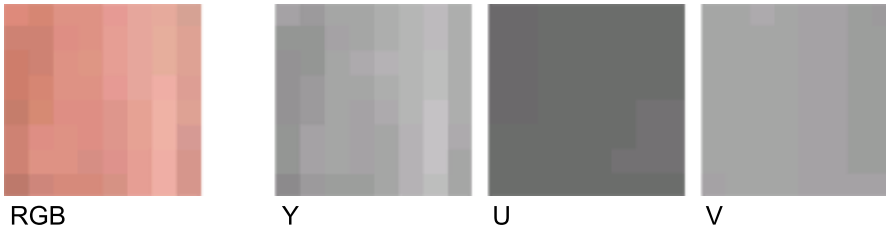
Un esquema general de los diferentes pasos que se realizan en este modo secuencial es el siguiente:



Esquema general de los pasos en el modo secuencial

1) Paso 1: cambio de espacio de color y submuestreo del color

Si la imagen es RGB, en primer lugar se transforma a YUV. Como ya vimos al principio de este módulo, dado que el ojo humano tiene menor resolución de color, las imágenes U y V se submuestran y así reducen su número de píxeles. La imagen Y (blanco y negro) se mantiene inalterada.



Reducir el número de píxeles

La reducción se realiza en proporción 4:4:4, 4:2:2 o 4:2:0 en función de la calidad elegida.

2) Paso 2: división en bloques y secuenciado

Cada uno de los tres componentes se comprime por separado y se divide en bloques de 8x8 píxeles. Así, cada componente se convierte en una secuencia de bloques, comenzando por el de la esquina superior izquierda y acabando con el de la esquina inferior derecha. A partir de este punto, el proceso se ejecuta de forma individual por cada bloque.

3) Paso 3: codificación de cada bloque

Cada bloque de imagen de 8x8 píxeles se transforma matemáticamente en otro bloque de 8x8 valores (también enteros) que ya no representan una pequeña imagen sino las **frecuencias espaciales** de ésta. El bloque, pues, contiene la misma información que tenía pero con otros valores: un puro ejemplo de codificación. Y sin pérdidas hasta este punto, pues podríamos realizar el proceso inverso y recuperar exactamente la pequeña imagen original.

Bloques de píxeles

Los píxeles son de valor entero de 0 a 255, pero en este punto se le resta a cada uno 128 para que estén en el margen de -128 a 127. Esta curiosa operación previa a la codificación se explica en el paso siguiente.

La transformación utilizada se denomina **DCT** (*discrete cosinus transform*) y es la base de la compresión de imagen y vídeo hoy en día.



Valors de píxel del bloc visual original								→	Valors de coeficient del bloc transforma							
40	31	35	33	46	68	71	47		240	35	23	-13	4	-6	3	2
22	24	40	37	42	68	71	47		36	-4	-21	4	2	0	-1	1
23	23	35	43	55	70	76	47		-68	4	9	21	9	5	-1	-1
23	30	36	34	43	68	76	47		4	0	-7	12	-2	0	-1	1
22	39	35	31	41	67	80	47		-18	-3	6	6	3	-2	0	0
25	33	37	31	36	68	80	50		1	3	2	2	1	-1	0	0
32	33	35	30	35	68	80	35		-2	0	0	-1	0	-1	0	0
10	31	28	28	35	69	73	35		1	-1	3	1	0	0	0	0

4) Paso 4: cuantificación de cada bloque

Si ambos bloques contienen la misma información, ¿por qué buena parte de los valores o coeficientes del segundo bloque son de valor muy reducido, incluso nulo, respecto a los restantes? Éste es el quid del proceso: la transformación DCT compacta la información asociada a la imagen en los primeros coeficientes (situados cerca de la esquina superior izquierda), con lo que nos posibilitaría eliminar buena parte del resto con valores nulos o de bajo valor (indicados en rojo en la imagen adjunta) y reducir el número total de coeficientes que representan el bloque.

Pero ¿qué criterio elegir para eliminar los coeficientes pequeños? ¿Y qué quiere decir pequeño? Para automatizar la selección de los coeficientes por eliminar en cada bloque de una manera razonable, se utiliza una tabla de referencia denominada **de cuantificación**, que contiene 8x8 valores de peso o ponderación, definida tras muchas pruebas de calidad subjetiva realizadas con grupos de observadores. Esta tabla contiene pesos pequeños en las posiciones cercanas a la esquina superior izquierda y pesos crecientes cuanto más lejanos están de ésta. Cada coeficiente de la imagen transformada se divide por el peso que le corresponde por posición, de manera que los alejados de la esquina superior izquierda quedan divididos por pesos elevados y dan resultado cero (se desprecian los decimales).

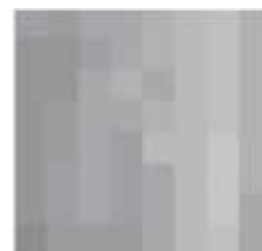
Ejemplo

Veamos un ejemplo de tabla de cuantificación utilizada en JPEG (Losheller):

16	11	10	10	24	40	51	61
----	----	----	----	----	----	----	----

12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Si hiciésemos la transformación inversa, denominada IDCT, para recuperar el bloque de imagen original a partir del bloque transformado y cuantificado, **el efecto visual de la pérdida de coeficientes es mínimo**. En el ejemplo siguiente, hay que descartar que los indicados en rojo no afectan sensiblemente al resultado destransformado.



Valores de coeficiente del bloque transformado y debidamente cuantificado								→	Valor de píxel del bloque visual destransformado resultante							
15	3	2	-13	0	0	0	0	40	30	33	32	42	59	65	44	
3	0	-1	0	0	0	0	0	22	23	41	33	45	61	69	44	
-4	0	0	1	0	0	0	0	23	22	36	46	51	73	73	43	
0	0	0	1	0	0	0	0	22	31	36	33	41	70	75	44	
-1	0	0	0	0	0	0	0	23	38	33	31	40	63	81	43	
0	0	0	0	0	0	0	0	24	32	34	29	32	63	75	53	
0	0	0	0	0	0	0	0	31	29	33	32	32	60	76	35	
0	0	0	0	0	0	0	0	8	29	25	25	34	67	75	35	

Sin duda, almacenar en un fichero los coeficientes transformados que nos interesan en vez de todos los píxeles originales daría una buena tasa de compresión. Si además descartamos aquellos que tienen valor despreciable, la tasa mejora sensiblemente al minimizar el efecto visual de las pérdidas introducidas cuando se invierte la transformación y se recupera la imagen. Además, la DCT no necesita de un procesador potente de cálculo y se puede ejecutar en hardware sencillo. Sin embargo, el proceso de compresión JPEG aún no ha acabado.

Hagamos una pausa en el proceso de explicación y razonemos las siguientes cuestiones:

a) **¿Cómo es posible que la DCT haga insensible visualmente la eliminación de buena parte de los 64 valores?** Tras años de pruebas y comparativas con todas y cada una de las transformaciones conocidas, la DCT es la que mejor condensa lo importante en menos datos sin necesidad de mucha potencia de cálculo (factor que en los años ochenta era crítico). Hay transformaciones mejores, pero todas necesitan cálculo intenso superior, como por ejemplo JPEG2000, que veremos más adelante en este mismo módulo.

b) **¿Qué significa visualmente cada uno de los 64 coeficientes?** Tal como se ha indicado, estos datos generados por la DCT representan frecuencias espaciales, es decir, cada coeficiente, por sí solo, generaría un bloque de imagen simple correspondiente a una oscilación visual.

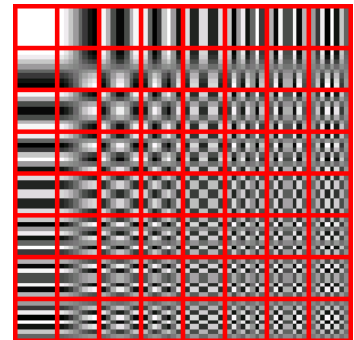
La imagen que tenéis en el margen es un resumen del tipo de bloque de imagen generado por cada coeficiente en función de su posición en el bloque transformado. Por ejemplo, el coeficiente que se ubica en la esquina superior izquierda genera por sí solo un bloque de imagen de 8x8 píxeles como el indicado en esa misma esquina, un bloque visualmente homogéneo. El segundo coeficiente genera un bloque de imagen de 8x8 píxeles similar a una onda horizontal de medio periodo, el tercero una onda horizontal de un periodo y así sucesivamente.

Pues bien, cualquier fragmento de fotografía en blanco y negro de 8x8 píxeles que se elija puede representarse como una suma de estos 64 bloques debidamente ponderados (multiplicados) por un valor. Si se tuviese tiempo (años) y paciencia, se podrían ir probando sumas de estos bloques debidamente amplificadas cada uno por un valor a nuestra elección. El día que consiguiésemos la suma perfecta que reprodujese exactamente el fragmento de fotografía, descubriríamos que los 64 valores seleccionados coinciden con los generados por la sencilla DCT.

El primer coeficiente del bloque es el peso o valor de la frecuencia cero que denominados **componente de continua o baja frecuencia (visual)** y que, como podéis ver, recrea el valor medio de brillo de todo el bloque y genera por sí solo un bloque visual de 8x8 homogéneo. Así, si borrásemos todos los componentes del bloque salvo éste y aplicáramos la transformada inversa para volver a tener un bloque de imagen, tendríamos lo que denominamos un efecto mosaico.

Potencia de cálculo en la década de 1980

El 80386 de Intel era el procesador del momento hasta 1989.



Ejemplo completo de la imagen generada por cada coeficiente por separado

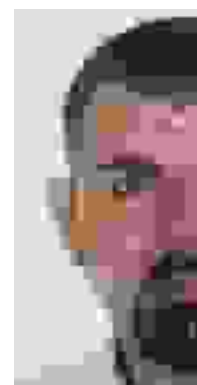


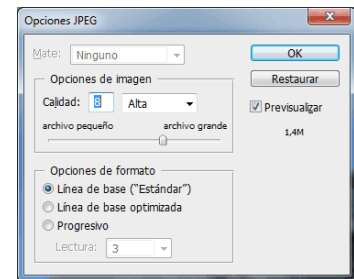
Imagen codificada sólo con los tres primeros componentes de cada bloque

El ojo humano es muy sensible a este valor medio y se percata de cualquier error en este valor, de tal manera que cualquier variación que hubiese entre dos cuadros consecutivos genera un efecto mosaico. Por ello, el JPEG mima el valor de este coeficiente como ningún otro, tal como vamos a ver en el paso siguiente.

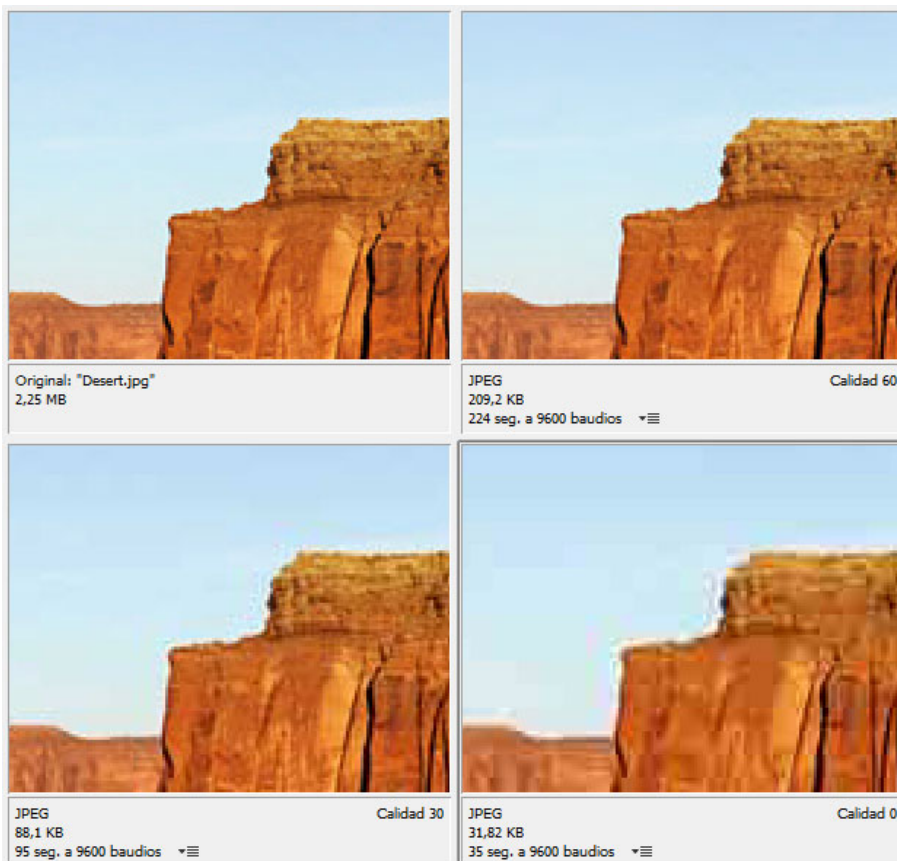
Los 63 coeficientes restantes se denominan **de alterna o alta frecuencia** y son responsables de las variaciones de brillo a partir de este nivel de base. Los bloques de imagen con mucho detalle visual, al aplicárseles la DCT, tienen muchos coeficientes de este tipo con gran valor, mientras que los bloques de poco detalle al contrario.

c) Calidad de imagen y tablas de cuantificación

Cuando se desea codificar una imagen en formato JPEG, el software (por ejemplo, Photoshop) pregunta el nivel de calidad que deseamos mantener de la imagen original en la nueva imagen JPEG. Usualmente es un parámetro cuantitativo (como un valor de 0 a 10 o de 0 a 12, incluso un porcentaje de 0 a 100%) y el primero es la peor calidad y máxima compresión, mientras que el último es la máxima calidad y mínima compresión. En otras ocasiones, posibilita seleccionar la calidad o compresión mediante valores cualitativos (de menor a mayor detalle), como por ejemplo en las configuraciones de las cámaras digitales.



Opciones de compresión JPEG de Photoshop



Cuadro comparativo de las compresiones JPEG sobre una misma imagen

Este parámetro de calidad seleccionado por el usuario está íntimamente ligado a la tabla de cuantificación que debemos aplicar para eliminar los coeficientes de alta frecuencia de todos los bloques transformados de la imagen por codificar. Así, a mayor valor de calidad, los valores de la tabla de cuantificación se reducen, por lo que la división da un resultado mayor y sobreviven más coeficientes de alta frecuencia. Cuando la calidad es máxima, los valores de la tabla de cuantificación son todos 1 y no teniendo ningún efecto.

También el valor de calidad puede hacer seleccionar al codificador utilizar técnicas de submuestreo de color 4:2:2 o 4:2:0 para reducir sensiblemente el número de bloques de color, lo que mejora la compresión sin necesidad de ninguna acción por parte del usuario.

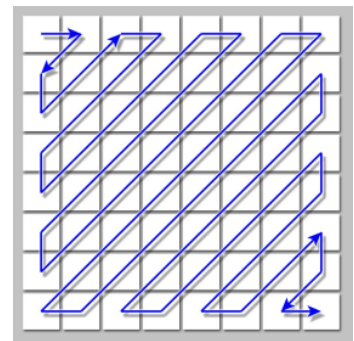
5) Paso 5: codificación de los coeficientes

Si bien los 64 coeficientes son enteros, en este paso se convierten a binario y se almacenan en fichero mediante una codificación que reduce la necesidad de bits para su almacenamiento. Como el coeficiente de continua es tan importante, se trata con una codificación específica.

En primer lugar, se ordenan siguiendo un orden en zigzag establecido. A continuación, se codifican de la siguiente manera:

- **Coefficiente de baja frecuencia:** guardar la diferencia entre dos valores ocupa casi siempre menos espacio que guardar los dos valores, siempre y cuando estos valores sean parecidos. Es lo que se denomina **codificación diferencial (DPCM)** y se aplica al coeficiente de continua de cada bloque: como en la mayoría de imágenes el valor medio de un bloque es usualmente similar al del bloque inmediatamente anterior, se almacena el valor del primero y posteriormente las diferencias de cada uno con el siguiente coeficiente de continua. Si además se almacena en el fichero JPEG antes que el resto de coeficientes, tendremos la imagen mosaico de manera inmediata como vista preliminar de ésta.
- **Coefficientes de alta frecuencia:** en la ristra secuencial de valores generada por la ordenación en zigzag aparecen usualmente muchos valores repetidos, la mayoría ceros, por lo que se aplican técnicas de compresión que aprovechen esta característica (**Huffman**).

Según como se almacenen en el fichero los componentes de color, se habla de **componentes entrelazados** (donde se van almacenando los bloques de luminancia y de color alternativamente) o **no entrelazados** (primero se almacenan los de luminancia y luego los de color).



El camino en zigzag de ordenación de los coeficientes

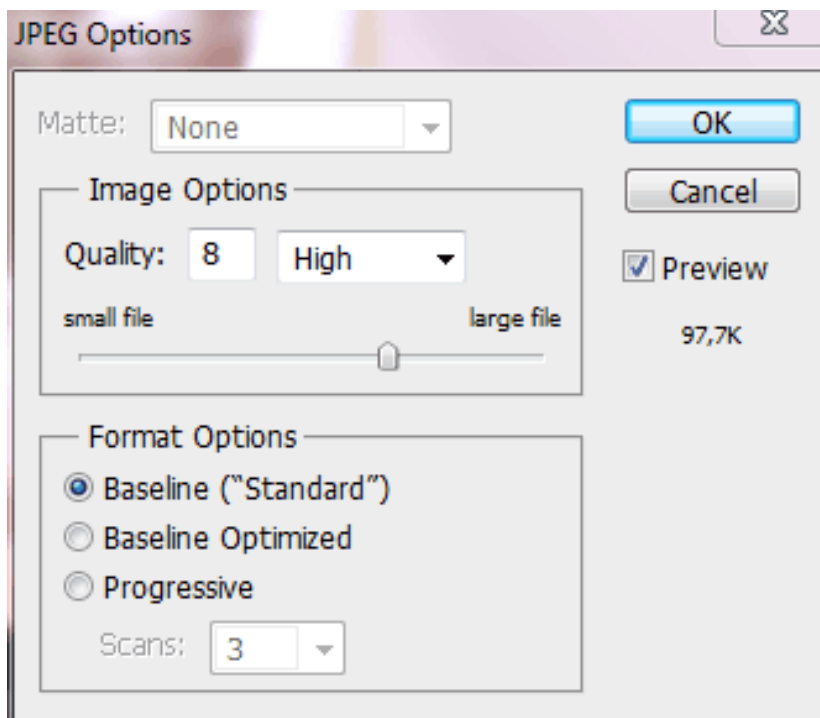
Descodificación de una imagen JPEG

El fichero JPEG posee toda la información necesaria para recuperar la imagen comprimida según el valor de calidad con el que se generó, ya que no sólo incluye los coeficientes transformados de todos los bloques, sino la información necesaria para recuperarlos a partir de sus versiones codificadas, el valor de calidad seleccionado y la tabla de cuantificación seleccionada para eliminar elementos.

Actividades

Actividad 12

Vamos a comprimir una imagen en formato JPEG mediante Photoshop para su efecto en la calidad de la imagen y en el tamaño del fichero resultante. Para ello, abrid la aplicación y leed una imagen, a continuación mediante Archivo-Guardar como seleccionad un nuevo nombre y como formato de grabación elegid JPEG (*.JPG, *.JPEG, *.JPE, las tres extensiones habituales que se añaden al nombre del fichero). Tras aceptar, aparece la siguiente ventana, donde se pregunta la calidad visual en la que se desea comprimirla (un valor de 0 a 12). Por ahora, mantened como opción el modo *baseline* (estándar) del que se hablará en el siguiente apartado.

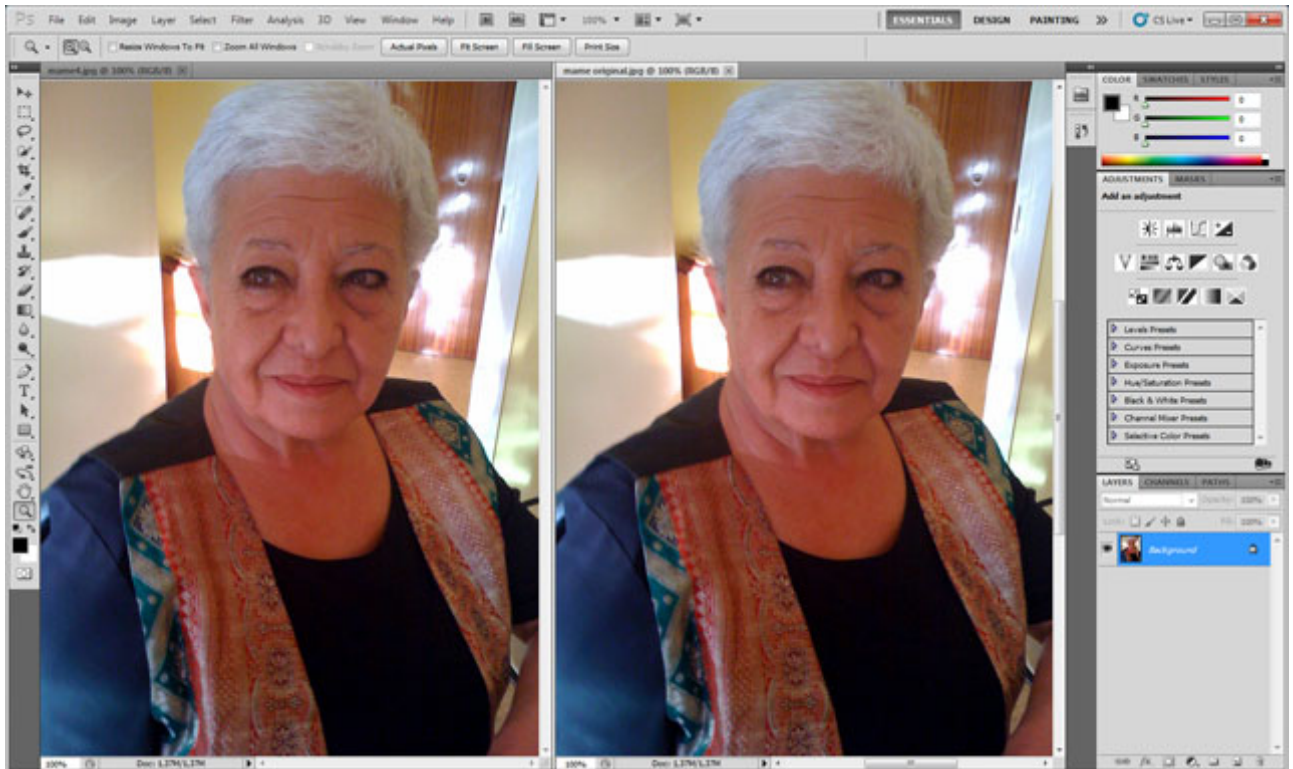


Tras aceptar, el Photoshop almacenará el fichero comprimido. Visualizad el resultado y mediante el sistema operativo que se use (Windows, OSX) tomad nota del tamaño en Kbytes del fichero resultante. Podéis calcular sencillamente el factor de compresión.

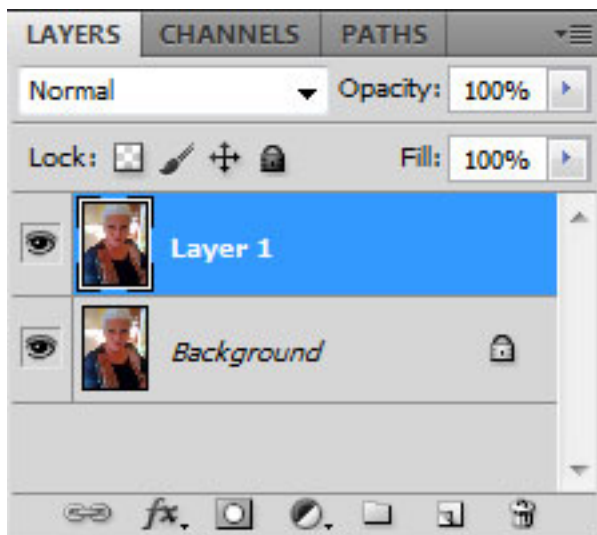
Actividad 13

Como os percatáis, el factor de compresión puede dar una idea aproximada del resultado, pero no es útil si se desea descubrir realmente dónde se está generando la pérdida de información visual. Para ello, se puede crear una imagen diferencia que sea la resta píxel a píxel entre la imagen sin comprimir y la comprimida, en este caso, en JPEG.

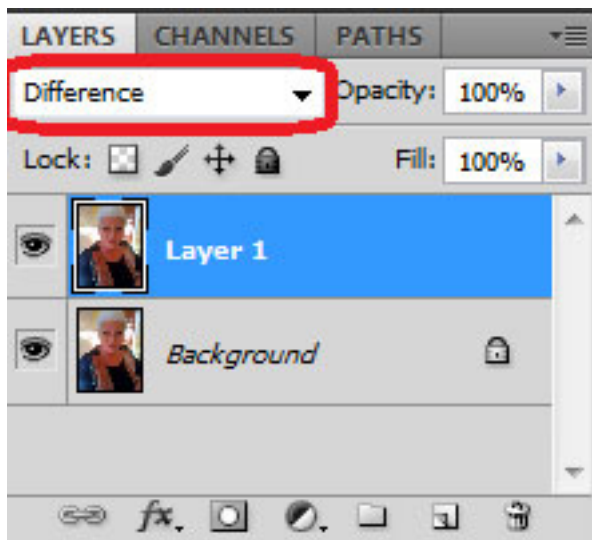
Abrid el Photoshop, leed la imagen original y leed también la imagen comprimida (aconsejamos elegir una imagen comprimida en un factor de 4 a 6, así los errores serán más evidentes).



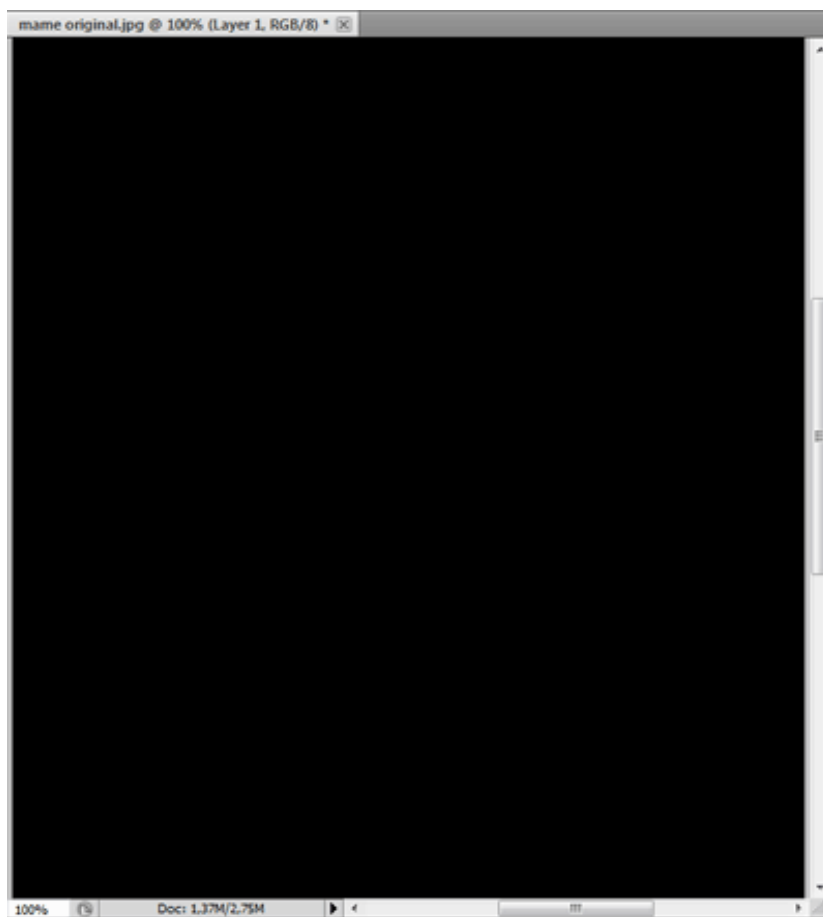
A continuación, seleccionad con un clic la imagen original sin comprimir y mediante CTRL-J (Windows) o COMMAND-J (Mac) duplicad su capa Background en la paleta de capas. El resultado son dos capas: la capa Background (o Fondo) original abajo y una copia que el Photoshop denomina Layer 1 (o Capa 1 en función del idioma del programa).



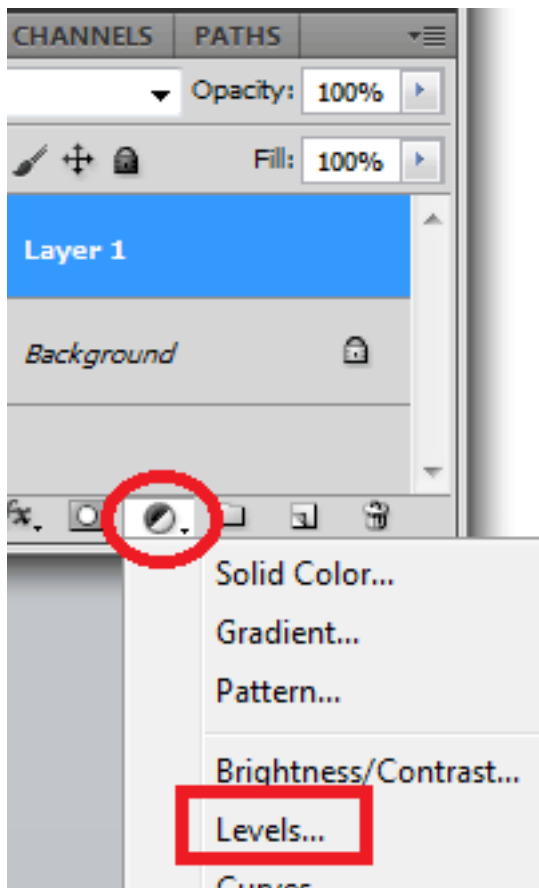
Ambas capas son idénticas, como demostración presentaremos la imagen diferencia: se puede avanzar que, si ambas son iguales, la diferencia entre píxeles correspondientes será 0, con lo que todos los píxeles de la imagen diferencia serán de valor 0, es decir, totalmente negros. En el desplegable de la ventana de capas, elegid Diferencia:



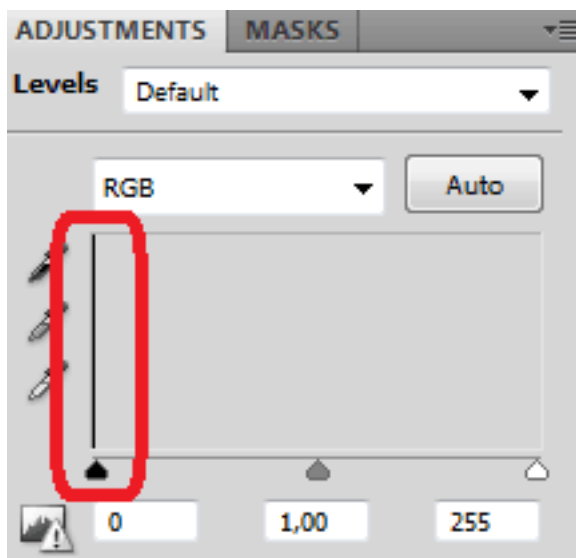
Tal como era previsible, la **imagen diferencia** aparece totalmente negra.



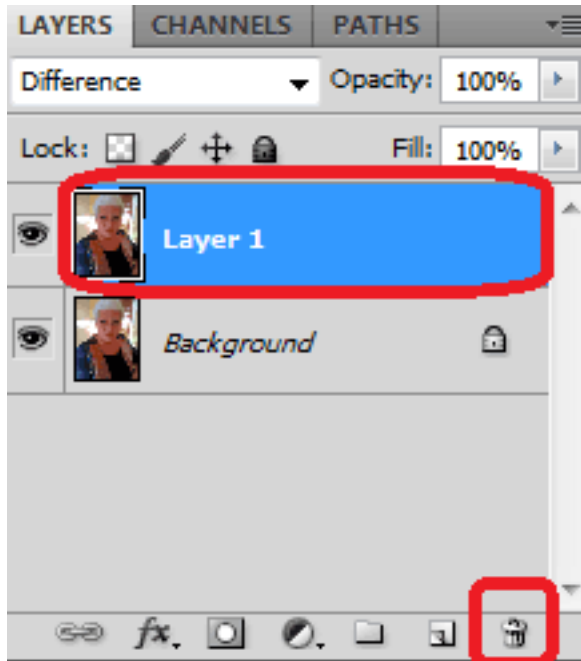
Seamos escépticos y confirmad que es así de una manera más explícita; para ello, seleccionad la herramienta de ajuste y, de las opciones existentes, elegid Levels (Niveles).



Activará una ventana de diálogo donde se refleja el histograma de la imagen diferencia. Es tan estrecho porque todos los píxeles son del mismo valor, no se distribuyen en otros valores más que el cero.



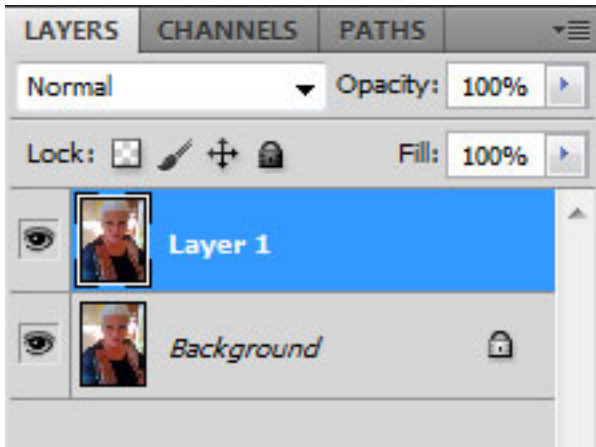
Una vez demostrada que la imagen diferencia entre dos imágenes iguales es negra, borrad Layer 1 (Capa 1) seleccionándola con un clic en la ventana de capas y pulsando la papelera situada en esa misma ventana.



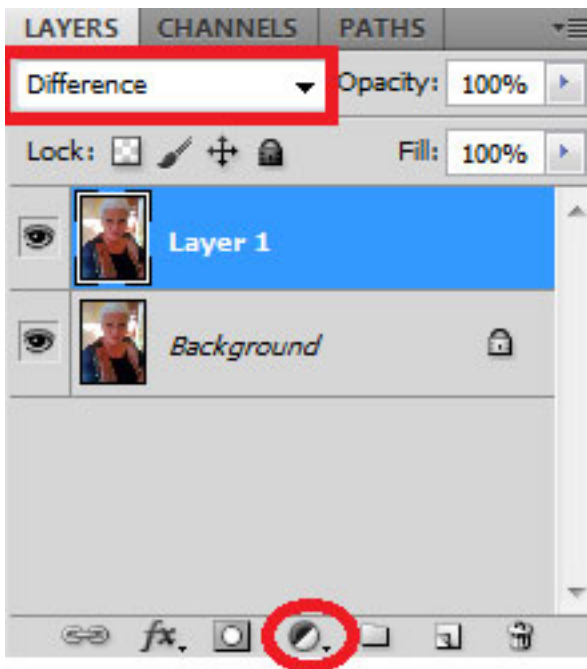
Ahora que se sabe cómo realizar una imagen diferencia y que hemos vuelto al punto de partida, vuelve a ejecutar la operación pero con las dos imágenes que aún están abiertas en el Photoshop. Seleccionad la herramienta Mover en la barra de herramientas, pulsad el ratón sobre la **imagen comprimida** sin soltarlo, arrastrad el ratón hasta la imagen sin comprimir, pulsad y mantened la tecla MAYUS y finalmente soltad el ratón.



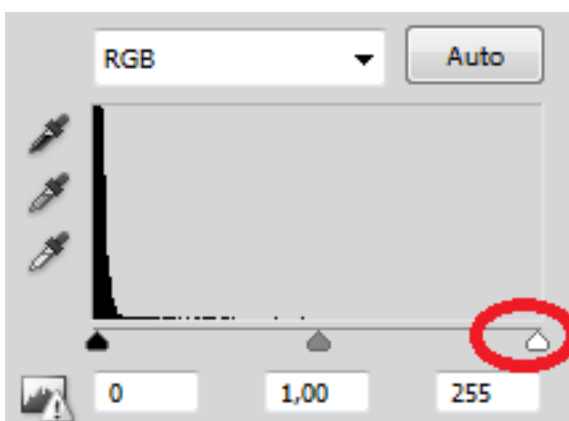
Si bien visualmente no parece haber habido ningún efecto, en la ventana de capas se verá que se ha creado una segunda capa, Layer 1 o Capa 1, que es la imagen comprimida, mientras que Background o Fondo es la imagen sin comprimir.



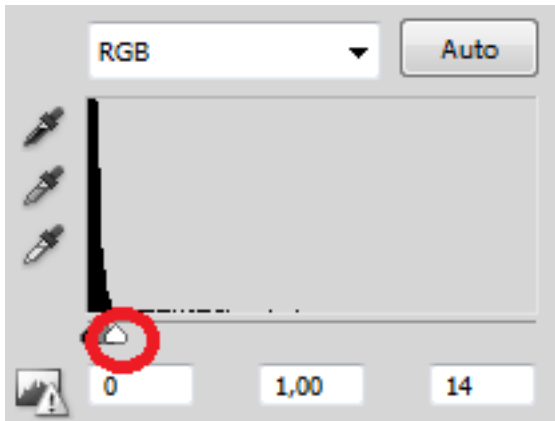
Ya se tienen ambas imágenes en capas, por lo que se puede volver a ejecutar la operación Diferencia (aparecerá ésta, aproximadamente negra) y a continuación la herramienta de capas, eligiendo Niveles:



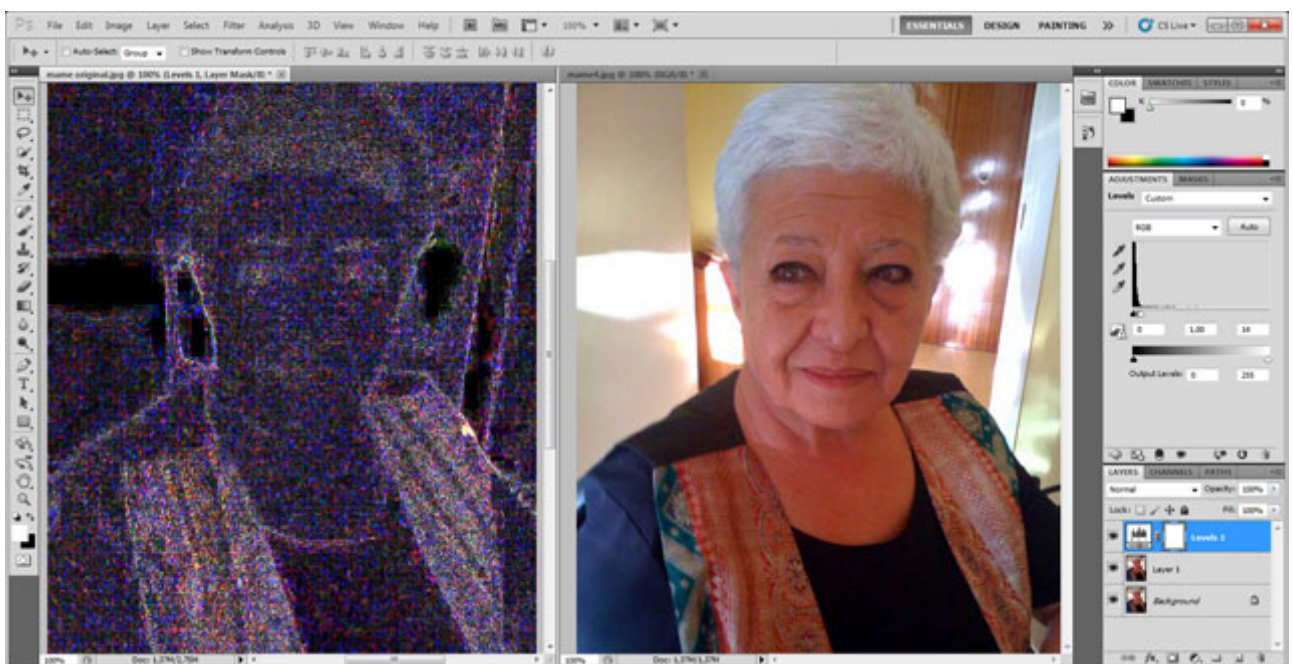
Casi con seguridad, el histograma ya no será tan uniforme:



Para poder discernir con claridad los píxeles diferencia, desplazad el *gadget* blanco hacia la izquierda y así realzaréis visualmente sus valores.



De este modo, todos los errores generados por la codificación con pérdidas de JPEG se reflejan claramente situados en zonas específicas de la imagen, como los bordes de los objetos y en general las zonas con gran detalle visual.



Realizando el proceso con niveles de compresión más elevados aparecen igualmente errores en zonas de color homogéneo y un claro efecto de bloques en general, intrínsecos a la manera como el JPEG trabaja la imagen.

4.2.2. Otros modos de trabajo del JPEG

El modo secuencial base explicado es válido para la mayoría de aplicaciones de tratamiento de la imagen, donde éstas poseen 8 bits por cada uno de los componentes (RGB) e incluso una cuarta capa de transparencia.

Sin embargo, existen otros modos (hasta 44) para aplicaciones específicas. Por ejemplo, para imágenes de muy alta calidad con 12 bits por componente o para aplicar otros sistemas de codificación de coeficientes más efectivos (codificación aritmética en vez de Huffman).

Una versión mejorada del modo secuencial base, el **modo secuencial base optimizado**, permite una mejora del 2% al 8% en el tamaño del fichero final. Sin embargo, algunas aplicaciones pueden no ser compatibles con este modo.

Actividad 14

Visualizad la imagen diferencia en el caso de compresión JPEG en modo Línea de base optimizado/Baseline optimized y valorad la diferencia con JPEG Standard/Baseline utilizando la misma imagen original del ejercicio anterior y comprimiéndola a la misma calidad. Valorad la calidad visual y el factor de compresión de la resultante.

El tercer modo interesante, el **modo progresivo**, muy utilizado para imágenes de sitios web cuyos servidores tienen conexiones de baja velocidad o tienden a estar muy saturados, posibilita ver una imagen mosaico (es decir, la imagen con sólo el coeficiente de continua de cada bloque) y sus sucesivas mejoras de calidad en función del orden de recepción de los datos, por lo que el espectador puede tener vistas parciales cada vez más detalladas de la imagen que va recibiendo. Hay varias maneras de implementar esta solución, aunque la más generalizada es almacenar en el fichero JPEG en primer lugar los coeficientes de continua y los primeros coeficientes de alterna de todos los bloques, luego coeficientes de alterna de menor relevancia visual (mayor detalle) y después de relevancia aún menor (y más detalle). El proceso suele realizar tres, cuatro o cinco fases de separación de los coeficientes. Un fichero JPEG en modo progresivo tiene normalmente un tamaño en bytes superior al de su equivalente en modo secuencial o secuencial optimizado.

Actividad 15

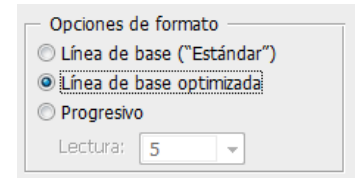
Realizad el ejercicio anterior eligiendo en este caso el modo Progresivo/Progressive con cuatro lecturas. Mediante un navegador o programa de visualización, abrid el fichero de imagen generado y comprobad que se visualiza con una mejora gradual.

Para finalizar, el **modo JPEG-LS** (lossless JPEG, JPEG sin pérdidas) es un modo creado unos años más tarde (1993) que codifica la imagen **sin pérdidas**. Utiliza una DCT para codificar, pero no utiliza tablas de cuantificación, sino que calcula previsiones de los píxeles a partir de sus píxeles vecinos. Destinado a aplicaciones médicas de imagen, consigue unos factores de compresión de 2:1, sin tener relevancia fuera de este campo.

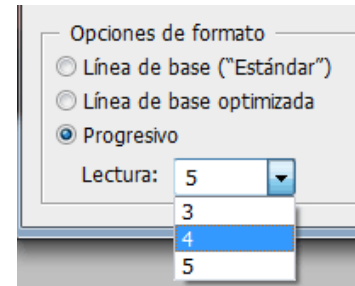
4.3. JPEG 2000

Una década más tarde de la aparición del JPEG, el escenario había cambiado sensiblemente: se habían investigado con éxito nuevas transformaciones que superaban la clásica DCT, aunque obligaban a un cálculo intensivo, pero la capacidad de cálculo de los equipos domésticos y profesionales quintuplicaba el de diez años antes, lo que hacía factible un nuevo formato de codificación que diese respuesta a las nuevas exigencias del mercado, entre otras:

- la codificación con pérdidas o sin pérdidas,
- el alto factor de compresión con menores artefactos visibles,



Modo secuencial base optimizado



Modo progresivo

- la robustez contra errores para poder transmitir a baja velocidad, por ejemplo,
- la capacidad de añadir datos adicionales o metadata,
- la mejor codificación de imágenes sintéticas o bitono.

Lamentablemente, la situación legal del formato JPEG2000 fue siempre controvertida, ya que existen patentes sobre partes del proceso, lo que ha frenado desde el principio su popularización.

El JPEG da calidades subjetivas de imagen muy bajas en compresiones menores de 0,25 bpp, donde el JPEG2000 aún puede dar calidad incluso hasta en 0,1 bpp. Y en los rangos de compresión habituales, este nuevo sistema de codificación supera al JPEG.

El secreto es la transformación Wavelet, que **trata la imagen como un todo** (no la descompone en bloques), le aplica varios filtros espaciales y aprovecha de manera muy eficiente la codificación de los resultados. Según los filtros que se utilicen, esta transformación puede ser **sin pérdidas** o **con pérdidas**, en este último caso mejora la compresión a costa de algún efecto visual en los contornos (la alta frecuencia espacial).

Así, si el JPEG2000 es capaz de comprimir con calidad a mayor tasa, es mejor para transmitir imagen a baja velocidad. Como además el **tipo de descodificación es progresivo**, la imagen recibida se descodifica a partir de una primera versión básica (baja frecuencia espacial) que va mejorando en detalle, perfecto para transmisión: si se interrumpe en cierto momento, al menos tendremos una imagen completa parcialmente detallada.

El estándar posibilita incluir datos adicionales (metadata) en el fichero de imagen. El JPEG ya lo posibilitaba en su formato EXIF, pero el JPEG2000 lo hace de manera más estandarizada siguiendo el formato XML.

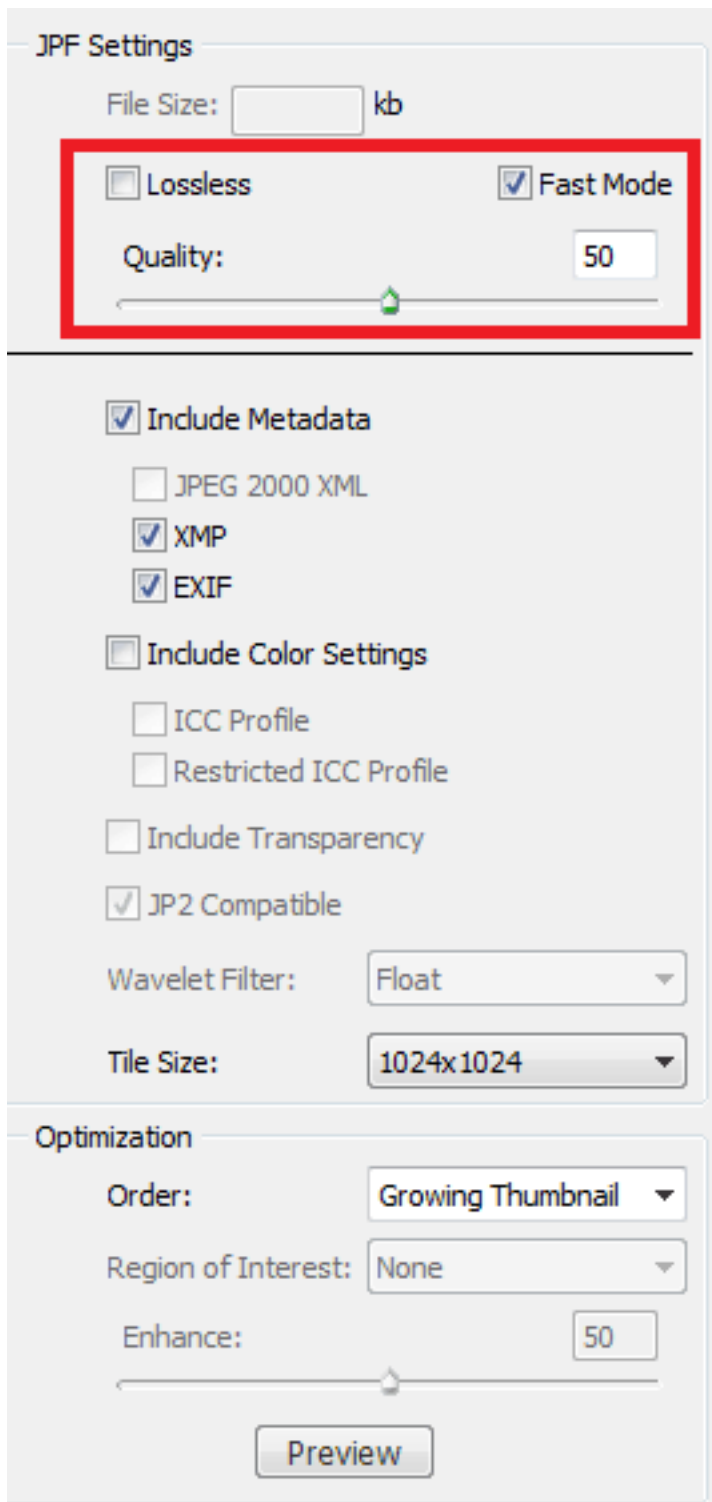
Actividad 16

El Photoshop lee y graba en formato JPEG2000 (extensiones *.JPF, *.JPX, *.JP2, *.J2C y *.JPC) aceptando una configuración personalizada en cada grabación, que destaca si se desea una compresión “sin pérdidas/lossless” en una calidad indicada por un valor de 0 a 100.

Realizad algunas compresiones con este formato en modo con pérdidas y probad el formato sin pérdidas, así valoraréis el nivel de compresión logrado en el ámbito del tamaño de fichero.

Problemas legales con JPEG 2000

Buena parte del software de tratamiento de imágenes no lo incluye o lo hace opcionalmente, como el Adobe Photoshop. En el caso de formatos de compresión sin pérdida, actualmente PNG es el más extendido.



The image shows a 'JPF Settings' dialog box with the following sections and options:

- JPF Settings**
 - File Size: kb
 - Lossless Fast Mode
 - Quality: (with a slider below it)
- Include Metadata
 - JPEG 2000 XML
 - XMP
 - EXIF
- Include Color Settings
 - ICC Profile
 - Restricted ICC Profile
- Include Transparency
- JP2 Compatible
- Wavelet Filter: (dropdown)
- Tile Size: (dropdown)

Optimization

- Order: (dropdown)
- Region of Interest: (dropdown)
- Enhance: (with a slider below it)

Opciones de codificación con o sin pérdidas de JPEG2000

4.4. JPEG XR

En el 2006, Microsoft introdujo un códec propietario de imagen inicialmente denominado HD Photo y dirigido exclusivamente al sector de la fotografía digital que desde el 2009, año en el que se convirtió en estándar ISO, es conocido como JPEG XR (Extended Range). La extensión de los ficheros de imagen es .wpd.

Sus características principales son las siguientes:

- codificación con pérdidas o sin pérdidas,
- decodificación progresiva,
- mejor compresión que el JPEG,
- un abanico de formatos amplio: 8, 16 o 32 bits por canal de color, los píxeles pueden tener valor entero o incluso flotante para adaptarse a cámaras fotográficas de rango dinámico muy alto y un formato en blanco y negro de 16 bits,
- el formato contenedor puede incluir metadata y un canal alfa,
- no está destinado a ser utilizado en imágenes sintéticas o con bitonos (texto o gráficos).

Si bien este formato sólo está soportado por Windows desde su versión Vista y existen algunos plugins comerciales (por ejemplo para el Adobe Photoshop ofrecidos por Microsoft), su avance es lento frente al estándar abierto de imagen DNG (Adobe, 2004). La liberación del JPEG XR en el 2010 por parte de Microsoft podría darle una posibilidad de éxito al facilitar su uso en aplicaciones de código abierto.

4.5. Tendencias futuras de la codificación de imagen

A mediados de la década de 1980, el matemático Michael Barnsley presentó la base de un sistema de compresión de imágenes con pérdidas basado en fractales que generó gran expectación.

Un **fractal** es un elemento sencillo que, iterado sobre sí mismo repetidas veces, genera un nuevo elemento complejo. La idea de Barnsley es generar una imagen original a partir de un elemento visual básico mediante iteraciones. Así, sólo se debería almacenar o transmitir el fractal o grupo de fractales, no la imagen. Incluso es posible reconstruir la imagen original a resoluciones superiores a las de partida con muy poca capacidad de cálculo.

Sin embargo, el problema principal reside en cómo encontrar de manera automática el elemento visual básico (o el conjunto de elementos básicos) que pudiese lograrlo. La necesidad de una gran capacidad de cálculo para conseguir deducirlos y la falta de un algoritmo eficiente para todo tipo de imágenes ha frenado hasta la fecha este sistema de compresión, aunque la investigación sigue abierta incluso para su aplicación en vídeo.

Resumen

Se han descrito los diferentes mecanismos de transporte de imagen y sonido para su captura, se han diferenciado los analógicos de los digitales y, de entre ellos, los de entorno puramente informático y se ha valorado la calidad conseguible en cada uno de ellos. En esta visión, se han introducido los conceptos de luminancia y crominancia, base de los sistemas de compresión de imagen y vídeo actuales.

Hemos tenido un primer contacto con la problemática de la transmisión de vídeo y audio digital a través de Internet, hemos valorado los problemas de retraso y pérdidas que puedan ocasionarse y hemos demostrado la existencia de la compresión en todo sistema de almacenamiento en archivos multimedia.

Describimos a continuación el sistema de codificación de imagen estática JPEG, sus fases de trabajo, los parámetros que la configuran y los diferentes modos. Este caso nos posibilita diferenciar entre los conceptos de codificación y compresión y nos ilustra algunas maneras de valorar tanto la compresión conseguida como los errores introducidos en el caso de compresión con pérdidas. Finalmente, damos una visión superficial de los estándares JPEG2000 y JPEG XR, al compararlos con el estándar original.

