

Conceptes bàsics de vídeo i àudio. Introducció

Alex Ribelles García

PID_00176921



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Introducció	5
Objectius	6
1. Característiques de la imatge	7
1.1. Fonamentant la imatge: paràmetres bàsics	7
1.2. Treballant amb senyals de vídeo analògic	8
1.2.1. Vídeo compost	8
1.2.2. S-vídeo	9
1.2.3. Vídeo analògic en components	9
1.3. Vídeo digital en qualitat estàndard, alta definició i 3D	10
1.3.1. SDI (<i>serial digital video</i>)	11
1.3.2. HD-SDI (<i>high definition SDI</i>)	11
1.3.3. 3G-SDI	12
1.3.4. DVI (<i>digital visual interface</i>)	12
1.3.5. HDMI (<i>high definition multimedia interface</i>)	12
1.3.6. Firewire	13
1.3.7. USB (<i>universal serial bus</i>)	13
1.3.8. Conclusió	13
1.4. Difusió enfront de vídeo a Internet	15
1.5. Resolucions, velocitat de quadre, relacions d'aspecte i altres paràmetres	18
2. Característiques del so	22
2.1. Àudio analògic	22
2.2. Àudio digital	22
3. La revolució en la gestió dels senyals	24
4. Codificació d'imatge fixa	26
4.1. La necessitat de comprimir	26
4.1.1. Codificació i compressió. Tipus de compressió	27
4.1.2. Redundància i entropia	27
4.1.3. Mesura de la compressió d'imatge	28
4.1.4. Mesura objectiva i subjectiva de la qualitat d'imatge	29
4.2. JPEG	29
4.2.1. Mode seqüencial	30
4.2.2. Altres modes de treball de JPEG	44
4.3. JPEG 2000	45
4.4. JPEG XR	47
4.5. Tendències futures de la codificació d'imatge	48

Resum..... 49

Introducció

Des del moment en què la maduresa tecnològica dels dispositius mòbils ha arribat a un punt que els capacita per a generar, gestionar i distribuir continguts multimèdia, la frontera que els separava dels equips dedicats s'ha difuminat, i s'apliquen bona part de les tecnologies ja existents i se'n generen de noves que aprofiten la flexibilitat i mobilitat característiques del nou mitjà.

Una bona captació és important per a aconseguir material audiovisual per a integrar en els nostres projectes, de la mateixa manera que seleccionar el format adequat per a emmagatzemar-lo i tractar-lo digitalment més endavant. Així doncs, comencem el mòdul construïnt sobre els fonaments bàsics de la imatge i el so, definint els paràmetres que els caracteritzen, els estàndards d'interconnexió i finalment veient l'estàndard d'emmagatzematge de la imatge estàtica, amb el qual s'introdueixen els conceptes bàsics de codificació i compressió que seran presents en la resta de mòduls, especialment en el de vídeo.

El mòdul és eminentment pràctic, si bé és obligatòriament necessari introduir conceptes teòrics que s'il·lustren en les activitats integrades en el text. Com a eines de treball s'utilitzaran dos programes, **VLC de VideoLAN** (per a totes les plataformes) i **Adobe Photoshop** (Windows i Mac), que s'inclouen al programari de l'assignatura.

Objectius

Els objectius que es pretenen aconseguir amb l'estudi d'aquest mòdul són els següents:

1. Identificar els paràmetres que caracteritzen la imatge i el so.
2. Diferenciar els estàndards d'interconnexió multimèdia domèstics, semi-professionals i professionals.
3. Identificar el transport físic més adequat segons el tipus de contingut digital, el cost i les prestacions.
4. Diferenciar els conceptes de codificació i compressió.
5. Interpretar els paràmetres de compressió de JPEG i JPEG2000 i seleccionar els valors adequats segons el contingut visual de la imatge.
6. Identificar els artefactes visuals més habituals en aquests estàndards d'imatge.

1. Característiques de la imatge

En la generació de contingut multimèdia és freqüent que hi participin elements audiovisuals originats per captura, és a dir, digitalitzats. En aquest primer mòdul introduïrem les característiques bàsiques que defineixen el senyal de vídeo i àudio i, a partir d'aquestes, il·lustrarem i compararem els diferents tipus de senyals i transports de vídeo i àudio que hi ha al mercat industrial i de difusió (*broadcast*), amb les quals es treballa en el dia a dia. A partir d'aquest moment i en la resta del curs ens centrarem en els senyals digitals, i presentarem com es codifica la imatge estàtica. En aquest recorregut destacarem conceptes de compressió que utilitzarem també intensivament en altres mòduls de l'assignatura.

Vegeu també

Aquests conceptes de compressió s'utilitzaran concretament en els mòduls "Digitalització, emmagatzematge i transmissió d'àudio i vídeo".

1.1. Fonamentant la imatge: paràmetres bàsics

Qualsevol text sobre imatge o vídeo comença amb la definició d'alguns termes bàsics, i en el nostre cas serem breus:

1) **Píxel**: és l'element visual mínim d'una imatge. És la pedra angular de totes les formes d'imatge digital, incloent tant la fotografia com el vídeo, i si bé el concepte és senzill el seu ús no ho és, especialment en un món de formats digitals d'una relació d'aspecte molt variada.

2) **Luminància**: és la porció del senyal de vídeo que representa la lluentor de cada píxel. La luminància màxima d'un píxel és el blanc, i la mínima, el negre. L'ull humà està desenvolupat per a percebre una gran quantitat de nivells de luminància.

3) **Crominància**: és l'altra porció del senyal de vídeo que representa el color de cada píxel. Els colors es destinen a cobrir tot l'espectre complet del sistema visual humà, del vermell al verd i al blau, en qualsevol combinació. L'ull humà, a diferència d'altres éssers vius, no té la mateixa capacitat de discriminació de colors que la que té per a la luminància. En realitat, aquesta porció consta de diversos elements diferents, la combinació dels quals donen el color.

4) **Escaneig**: és el procés utilitzat en un senyal de vídeo tant per a capturar com per a mostrar els valors de luminància i crominància de cada píxel. L'escaneig situa la informació de cada píxel en un ordre específic perquè un equip pugui determinar la informació que pertany a cada píxel. Per a l'ull humà, la presentació d'una seqüència d'imatges escanejada en pantalla a més de 18 imatges per segon ja li genera un efecte de moviment.

Escaneig

La presentació en dispositius en matriu, com pantalles LCD, plasma o LED, ja no es fa en escaneig, sinó que es refresquen tots els píxels de manera síncrona.

Els quatre conceptes estan íntimament lligats: així, com un caràcter de text o un byte de dades, cada píxel té un valor o conjunt de valors. Un és la lluentor (luminància) i, si és en color, també la crominància. Per a obtenir una imatge fixa, el valor de cada píxel és fix i suficient, encara que en calen prou perquè el destinatari pugui reconèixer la imatge. Però per a vídeo, a més, la informació de cada píxel s'escaneja (presenta) almenys 18 vegades per segon. Més endavant farem alguns números per a demostrar el gran problema que suposa la gestió d'aquesta quantitat de dades per a aconseguir transmetre'ls i el perquè dels estàndards de compressió.

1.2. Treballant amb senyals de vídeo analògic

Els senyals de vídeo analògic, per la seva naturalesa, poden crear-se a partir de tres senyals diferents de color (R, G i B) o bé a partir de dos senyals (luminància Y i crominàncies U i V).

En el treball diari amb equips de vídeo (càmeres, gravadors, etc.) i programes de gestió (tractament d'imatge, transmissió de vídeo, etc.) treballem amb aquests senyals en formats diferents alhora, de vegades de manera inconscient. Vegem-ho amb detall.

1.2.1. Vídeo compost

Un senyal de vídeo compost conté tota la informació necessària per a presentar vídeo en color en pantalla. El senyal s'envia normalment en un únic cable coaxial entre dispositius com un reproductor de vídeo i un televisor. Molts sistemes de consum utilitzen un connector groc, denominat RCA (encara que de vegades podem trobar un connector BNC). És el més estès en el sector domèstic, però el de menys qualitat: el senyal de vídeo compost és PAL, sistema analògic de vídeo en color de 625 línies i 25 imatges per segon que també s'utilitzava per a les emissions de televisió analògiques fins a 2010. Com els seus sistemes germans NTSC i SECAM, té avantatges i inconvenients, que indiquem en un quadre resum al final d'aquest apartat.

Els equips de captura de vídeo que es basaven en senyals de vídeo compost van ser la porta d'entrada al mercat domèstic i semiprofessional en els anys vuitanta. Aquells pioners que treballaven amb un Mac, Amiga, Atari, etc. i un reproductor VHS o Beta necessitaven un maquinari extern d'un cost molt alt de vegades integrat en el mateix equip, sempre amb problemes d'interferència en els senyals i amb molts paràmetres per a ajustar en cada captura per a aconseguir una qualitat acceptable.



Connectors RCA

Maquinari extern

Aquest maquinari extern era, per exemple, Truevision TGA de Targa, Video Toaster de Newtek, etc.

Vegeu també

Al final d'aquest subapartat trobareu una taula resum amb els avantatges i desavantatges d'aquest tipus de senyals.

1.2.2. S-vídeo

Denominat també *connector S-VHS*, és analògic com el vídeo compost, no obstant això la transmissió de luminància i crominància es duu a terme per diferents cables, cosa que redueix sensiblement la degradació del senyal. Per aquesta raó un cable de S-vídeo té quatre pins: un parell per al senyal de crominància U + V i un altre parell per al de luminància Y.



Connectors S-vídeo

La seva arribada a la fi dels anys vuitanta amb els equips Super-VHS i 8 mm va facilitar millorar tant la qualitat d'imatge tant en reproducció com en captura, que encara es pot trobar S-vídeo de manera testimonial en dispositius de vídeo de consum, incloent pantalles de vídeo, reproductors de DVD, càmeres de vídeo digitals, descodificadors de televisió per cable i satèl·lit, consoles de videojocs, etc. En general, per a aplicacions de consum i semiprofessionals, S-vídeo s'ha d'usar sempre que sigui possible en lloc de vídeo compost **si no hi ha cap alternativa digital** com les que veurem més endavant.

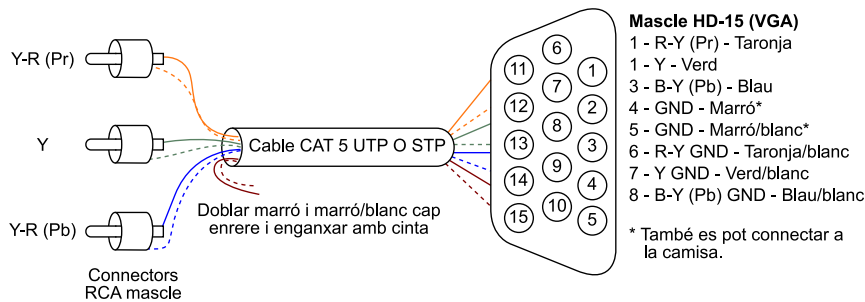
1.2.3. Vídeo analògic en components

Superior en qualitat als dos anteriors, utilitza un cable diferent per a cadascun dels tres senyals, amb la qual cosa la qualitat de color era millor que l'anterior a canvi de sacrificar la distància màxima entre reproductor i presentació o captura.

Hi ha dues versions diferents segons l'equip:

a) Vídeo en components RGB: cada senyal és d'un color específic i de bona resolució, a més suposa un processament mínim per a l'equip que l'emet i el que la rep, i és compatible directament amb la pantalla de presentació, cosa que millora la imatge sensiblement.

b) Vídeo en components YUV: una versió que treballa amb el senyal de luminància i les dues de crominància. Més recomanable en equips que treballin internament amb aquestes, com Betacam SP. Si ha de servir per a subministrar vídeo a una pantalla, aquesta haurà de convertir internament els senyals a RGB.



Assignació de pins entre components analògics YUV i connector VGA

Exemples

Un exemple encara vigent d'aquest sistema de connexió és l'estàndard **VGA** en components analògics, que afegeix senyals de polsos de sincronia per a ajustar el senyal a l'escombratge de la pantalla.

Un altre exemple encara vigent és el robust **Euroconnector**, que de manera excepcional inclou diferents formats de vídeo del mateix senyal: vídeo compost, S-vídeo i components RGB analògic per a donar servei a tot tipus de pantalles pel mateix connector, a més d'àudio estèreo analògic. Com que és tan versàtil, la seva vida útil s'ha allargat fins a aquesta dècada.



El robust Euroconnector, un clàssic des dels setanta

1.3. Vídeo digital en qualitat estàndard, alta definició i 3D

Cap dels sistemes analògics de connexió anteriors seria capaç de mantenir la qualitat d'un senyal actual de vídeo digital amb requeriments d'amplada de banda superiors, més i tot si a sobre és d'alta definició o 3D, ja que té una quantitat molt més gran de píxels per a transmetre cada segon. I és que el vídeo digital es compon de mostres de senyal, que cada una representa una porció minúscula d'aquest.

El senyal de vídeo digital, com l'analògic, es pot representar per 3 senyals de color independents (R, G i B) o bé un senyal de luminància (Y) i dos de crominància (U i V).

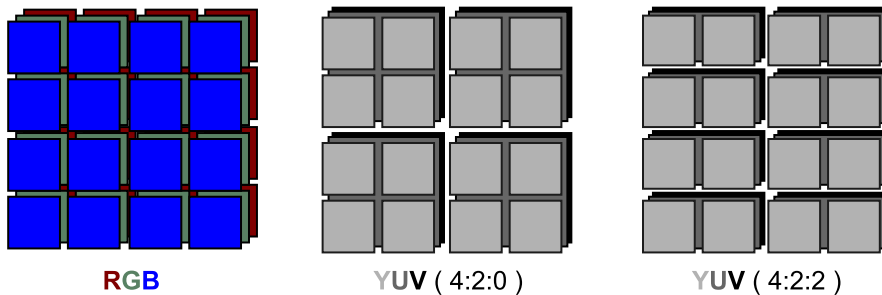
El procés de digitalització d'un senyal analògic segueix sempre el **criteri de Nyquist**, és a dir, es mostreja, almenys, al doble de la màxima freqüència del senyal analògic. No farem càlculs, però un exemple senzill ens pot il·lustrar aquesta regla tan senzilla: en el cas de l'àudio, si la freqüència màxima del senyal s'estableix en 20.000 Hz (per limitacions de l'oïda humana, i aquesta suposició és molt optimista), es mostreja a més de 40.000 mostres/segon. Per exemple, CD i DAT utilitzen 44.100 mostres/segon.

En el cas de vídeo, si fem la mateixa operació amb el senyal de vídeo en RGB el volum de dades que resulta és immens, ja que per a cada mostra de vermell R necessitem una altra mostra de verd G i una tercera de blau B, però si utilitzem el senyal YUV es redueix sensiblement. Això és a causa d'una petita argúcia: l'ull humà té bona resolució en blanc i negre, però és pobre en color.

Mala resolució per al color

Un exemple aclaridor és intentar endevinar el color d'un únic fil de teixit que hàgim situat a diversos metres: detectarem, o no, l'existència del fil, però no pas el color.

Així, podem **submostrejar** els senyals de color U i V, és a dir, generar menys mostres que les que farem per a la luminància Y, sense que l'usuari final se n'adoni:



Representació visual de l'assignació de mostres sobre píxels segons formats diferents

La relació denominada 4:2:2 suposa que, per cada 2 mostres de senyal Y, en farem una de U i una altra de V, mentre que la 4:2:0 suposa 1 mostra de U i una altra de V per cada 4 de Y. Tot depèn de la qualitat de color que es desitgi.

En tot cas, els fluxos de dades generades són de tal volum (centenars de Mbits) que necessitem una generació nova de sistemes de connexió, en la qual destaquen SDI i HD-SDI per a l'entorn industrial i de difusió, i DVI i HDMI per a l'entorn domèstic.

1.3.1. SDI (*serial digital video*)

L'estàndard *de facto* en l'entorn professional segueix la norma CCIR 601, que opera a 270 Mbps i conté el senyal de vídeo en qualitat estàndard 720x576 a 25 imatges/segon i a més fins a 16 senyals d'àudio digital **embeguts**, tot per mitjà d'un únic cable coaxial amb connector BNC i fins a una distància de centenars de metres, ideal tant per a l'estudi com per al desplegament d'equips en un gran esdeveniment.

El senyal digital que transmet és per components (Y, U, V) digitalitzats en relació 4:2:2, i que s'alternen en el temps, de manera que el cable només transporta un d'aquests tres senyals en cada instant de temps, sense que hi hagi, doncs, cap interferència. Equips Betacam SP, Betacam Digital, Betacam SX, DVCAM, DVCPRO, etc. i tots els sistemes d'encaminament i mesclat d'imatge que hi ha s'interconnecten amb aquest sistema si volen ser comercials.

1.3.2. HD-SDI (*high definition SDI*)

L'evolució natural de l'estàndard SDI, compatible amb senyals d'alta definició, segueix la norma SMPTE 292M i opera a 1,5 Gbps, però és mecànicament similar a l'anterior: un cable coaxial (evidentment més car) amb connectors BNC. Cinc vegades la velocitat de l'anterior per a acollir senyals de 1.920×1.080 o



El connector BNC arriba al segle XXI amb HD i so multicanal en un sol cable

1.440 × 1.080 a 25 imatges/segon (mode denominat *1080i* o *1080 entrelaçat*) o 1.280 × 720 a 50 imatges/segon (mode denominat *720p* o *720 progressiu*), sempre sense comprimir.

Formats de vídeo com HDCAM i DVCPRO HD, i totes les pantalles d'alta definició industrial i de difusió s'interconnecten així.

1.3.3. 3G-SDI

No obstant això, per a senyals HD-SDI en 3D cal el doble de velocitat de dades, ja que per a difusió el 3D es genera amb dos fluxos 1080i HD-SDI independents que han de circular per un mateix cable (mode anomenat *1080p* o *1080 progressiu*), de manera que el nou estàndard 3G (denominat així perquè arriba gairebé a 3 Gbps) ja apareix com a substitut natural de l'HD-SDI.

1.3.4. DVI (*digital visual interface*)

En un entorn domèstic els senyals SDI o HD-SDI no són utilitzats perquè no cal tanta qualitat, ja que quan el senyal surt del seu origen i arriba a la destinació per la xarxa de transport utilitzada (per TDT, satèl·lit, reproducció en temps real per Internet, en suport DVD, Blu-ray, etc.), està **comprimit**, és a dir, s'ha reduït la seva qualitat visual per a reduir el seu pes en bits i fer comercial la seva difusió. En realitat, bona part d'aquesta assignatura versa sobre la manera com s'ha de comprimir en totes aquestes xarxes de distribució.

Així doncs, com a solucions de connexió més senzilles hi ha el DVI, que pot tractar senyals de vídeo tant digitals (estàndard o alta definició) com analògiques, però habitualment sense el so associat, que requereix un cablatge a banda. La novetat és que és bidireccional, de manera que l'equip emissor i el receptor poden acordar la resolució més adequada per a treballar entre si, i permet també encriptar el senyal, encara que rarament s'usa per a això. Sens dubte és l'evolució natural del VGA i és en la majoria de monitors de vídeo de la llar, targetes gràfiques i fins i tot en televisors plans, encara que en aquest últim sector està perdent terreny enfront de la seva evolució, l'HDMI.

1.3.5. HDMI (*high definition multimedia interface*)

L'HDMI porta tant vídeo digital com senyals d'àudio. Es basa en DVI mitjançant l'addició d'àudio digital en el mateix cable, però conserva la compatibilitat amb l'especificació elèctrica DVI-Digital, és a dir, un mer adaptador converteix un connector HDMI en DVI, i es perd el so. La protecció anticòpia es basa en una tecnologia anomenada *HDCP* (protecció de contingut de gran banda digital), que obliga que l'equip reproductor i l'equip monitor acordin que només es visualitzarà el senyal i que no se'n faran còpies. Si l'acord fracassa, es reproduceix en qualitat baixa o senzillament no hi ha reproducció possible.



Connector DVI

Equips emissor i receptor

Un equip Blu-ray que vulgui reproduir a 1080p connectat a una pantalla HD Ready reproduirà en qualitat 1080i o 720p, ja que la pantalla comunicarà via DVI la seva incapacitat per a presentar senyals en 1080p, sense necessitat que l'usuari faci res. Si, en canvi, la pantalla és Full HD 1080 la reproducció serà en màxima resolució.



Connector HDMI

De les versions diferents que han anat apareixent al mercat des que va néixer el 2002, actualment es treballa amb la 1.4, l'única capaç de treballar amb senyals HDTV-3D i que obre la possibilitat de treballar amb resolucions superiors a 2000×2000 píxels encara en estudi (Quad HD, Super Hi-Vision, Ultra HDTV, etc.).

1.3.6. Firewire

Si bé l'origen de Firewire s'associa amb la transmissió en sèrie de dades digitals en entorns informàtics, en treballar amb vídeo en aquestes plataformes també s'utilitzen, usualment per a transferència de vídeo digital entre equips.

Firewire

Originalment va ser denominat *IEEE 1394* o *i.Link*, per Sony.

Firewire va ser desenvolupat per Apple el 1995 i actualment hi ha diferents versions segons la velocitat de transferència (400, 800, 1600, 3200 Mbps, etc.). Aquest suport físic s'utilitza amb el protocol DV per a connectar gravadors de vídeo i càmeres domèstiques i industrials a ordinadors, i inclou comandaments de control (rebobinatge, *play*, *stop*, etc.) i dades com el codi de temps.

1.3.7. USB (*universal serial bus*)

Si per ús extensiu i versatilitat s'hagués de seleccionar un sistema de transmissió de senyals, sens dubte l'USB seria el que centraria la nostra atenció. Nascut a mitjan anys noranta principalment per al mercat informàtic, actualment s'estén en entorns mòbils i fins i tot en sistemes industrials. La seva resistència, facilitat de manteniment, interoperabilitat entre equips, altes prestacions de velocitat i fiabilitat i la seva millora constant en velocitat resulten idònies per a les necessitats creixents de transferències de dades: des d'1,5 Mbps (USB 1.0), passant pels 12 Mbps (USB 1.1), els 480 Mbps (USB 2.0) i arribant a 4,8 Gbps en la revisió 3.0.

1.3.8. Conclusió

A manera de conclusió, presentem la taula següent amb un resum dels estàndards de transport de vídeo més habituals.

Taula resum dels estàndards de transport de vídeo més habituals

Transport de vídeo analògic	Avantatges	Inconvenients
Vídeo compost	<ul style="list-style-type: none"> Gran nombre de dispositius en tenen a tot el món. Un únic cable per a transmetre el senyal de vídeo. 	<ul style="list-style-type: none"> El senyal de luminància Y i els de crominància U i V s'interfereixen entre si en compartir el cable, i això degrada la imatge resultant. És analògic, per la qual cosa les interferències externes i el soroll l'afecten fàcilment.



A dalt, connectors 400 (estàndard i mini). A baix, connectors 800



L'omnipresent USB

S-vídeo	<ul style="list-style-type: none"> • Gran nombre de dispositius en tenen a tot el món. • La luminància i la crominància van per camins separats, no s'interfereixen i dóna més bon senyal d'imatge. 	<ul style="list-style-type: none"> • Connector mecànicament més delicat. • Segueix essent analògic, per la qual cosa s'utilitza exclusivament per a distàncies inferiors als 2-3 metres.
Vídeo en components	<ul style="list-style-type: none"> • Màxima qualitat en analògic. • Acceptable per a edició i postproducció professional. 	<ul style="list-style-type: none"> • Afegint els senyals de sincronia H i V, calen fins a cinc cables. • Tots han de ser exactament iguals, o els retards no coincideixen i apareixen artefactes en la imatge. • Exclusivament per a distàncies curtes.
Transport de vídeo digital	Avantatges	Inconvenients
SDI/HDSDI/3G	<ul style="list-style-type: none"> • Senyal digital, per la qual cosa és poc sensible al soroll i interferències. • Accepta múltiples generacions de copiat. • L'àudio, el codi de temps i altres senyals van incrustades en el mateix senyal de vídeo. • Fàcil de connectar amb un sol cable. • És un estàndard generalitzat en entorns industrials i de difusió. 	<ul style="list-style-type: none"> • Requereix equipament industrial o de difusió d'alt cost. • Els 270 Mbps que genera no són fàcils de gestionar ni d'emmagatzemar.
DVI	<ul style="list-style-type: none"> • Molt compatible, ja que pot treballar amb senyals analògics o digitals. • En el cas digital pot treballar amb cablatge de fins a 50 metres. 	<ul style="list-style-type: none"> • Dedicat exclusivament per a visualitzar vídeo, no per a interconnectar equips. • És un connector multipin, per la qual cosa és complex i pot sofrir torsió. • L'àudio necessita cablatge a part.
HDMI	<ul style="list-style-type: none"> • Connector compacte. • Inclou el senyal d'àudio. • Mitjançant extensors pot arribar a 250 metres. • Un adaptador senzill el converteix en DVI. • Inclou protecció anticòpia HDCP. 	<ul style="list-style-type: none"> • Hi ha diverses versions d'HDMI amb capacitats de transmissió diferents. • El connector no té ancoratge, de manera que es pot desconnectar fàcilment. • Si bé podria utilitzar-se per a transferir vídeo entre equips, el sistema anticòpia és un desavantatge que el relega a connectar monitors digitals.
Firewire	<ul style="list-style-type: none"> • Connector senzill i compacte. • D'ús universal, encara que és el més adoptat per a transferir àudio i vídeo a alta velocitat en entorns domèstics i semiprofessionals. • Més transferència sostinguda de dades que l'USB, necessària per a vídeo. 	<ul style="list-style-type: none"> • No té ancoratges. • És propietari (Apple). • No està tan estès com l'USB.
USB	<ul style="list-style-type: none"> • Connector senzill i compacte. • D'ús universal. • És utilitzat per a fonts de vídeo de velocitat baixa, com càmeres web, escàners, etc. 	<ul style="list-style-type: none"> • Orientat a interconnexió de perifèrics. • Rendiment lent per a vídeo d'alta definició. • No posseeix ancoratges.

Activitats

Activitat 1

Revisu l'equip informàtic que teniu actualment: quin tipus d'entrades i sortides poden ser interessants per a capturar vídeo?

Activitat 2

Revisu les entrades de senyal del monitor: quins senyals de vídeo accepta? Penseu que si teniu un portàtil multimèdia, pot avançar-se que l'entrada és de components digitals RGB, encara que connectada directament a la placa gràfica del portàtil.

Activitat 3

Si teniu una càmera de vídeo, etc. amb quin estàndard de connexió es connecta?

Activitat 4

Si teniu programari de captura (com Adobe Premiere, Camtasia, programari gratuït, etc.), és compatible amb les entrades disponibles?

Activitat 5

Ara reviseu el televisor del menjador. Si fa poc que l'heu comprat tindrà un bon nombre d'entrades. Quines són compatibles amb l'equip informàtic? Usualment s'acaba creient realment en la conversió, que tan sovint s'ha pronosticat, del televisor en un centre multimèdia digital.

1.4. Difusió enfront de vídeo a Internet

Fins ara, en aquest mòdul ens hem dedicat a senyals de vídeo tradicionals, que poden ser enviats directament a una pantalla de vídeo o a un ordinador amb un simple cable, a distàncies no superiors als 100 metres. En canvi, els continguts de vídeo transmesos o compartits per Internet han de ser obligatòriament digitals (si no es capturen prèviament), i requereixen un dispositiu (com un ordinador o un descodificador dedicat) o un programari específic per a reconstruir-les correctament abans de ser visibles.

Tots els fitxers o *streams* de vídeo de la Xarxa destinats als consumidors finals es **comprimeixen** mitjançant algun dels molts tipus de compressió que hi ha disponibles, per la qual cosa els dos senyals d'àudio i vídeo han de ser **descomprimides** mitjançant maquinari o programari abans de ser lliurats al destinatari. Si tenim TDT, televisió per cable o per satèl·lit, descomprimim per maquinari; si tenim un equip Windows Media Center o senzillament som usuaris de YouTube, per programari.

En el cas de programari, el més interessant i amb més futur, concretem que hi ha petites aplicacions denominades *extensions* o **connectors** que, integrades al programa navegador d'ús habitual, el capaciten per a fer aquesta descodificació al vol: Adobe Flash Player, QuickTime d'Apple i Windows Media Player, per dir algunes de les més populars. Els senyals de vídeo per Internet poden ser essencialment de qualsevol nombre de píxels horitzontals i verticals, i hi ha una selecció molt àmplia de sistemes de protecció de continguts i tècniques de lliurament de les dades mitjançant la Xarxa.

Amb totes les combinacions diferents que hi ha disponibles, els proveïdors de vídeo per Internet són capaços de desenvolupar les seves pròpies combinacions de formats, per la qual cosa el vídeo produït en un lloc pot no ser directament compatible amb altres llocs d'allotjament. Per això, la flexibilitat és un element clau dels reproductors de vídeo d'Internet: els televidents haurien de tenir llibertat per a aturar, iniciar, rebobinar i posar la pausa a la reproducció en qualsevol moment, controlar la mida del vídeo a la pantalla i triar entre

Vegeu també

Els diferents tipus de compressió es tracten en el mòdul "Digitalització, emmagatzematge i transmissió d'àudio i vídeo".

una varietat de formats de pantalla del reproductor. Quan vulguem distribuir un producte audiovisual per la Xarxa, no ens hem d'aferrar a un estàndard en particular si no volem reduir el nostre impacte.

En una xarxa com Internet, pensada per a transmetre dades, els senyals de vídeo no són fàcils de transportar, bàsicament per dues raons:

a) La primera és que no és una xarxa pensada per a vídeo. El trànsit per mitjà de la xarxa del vídeo i àudio comprimits es fa mitjançant paquets de dades genèriques (paquets IP), com tots els altres tipus de dades (textos, imatges, etc.) que hi flueixen. Aquests paquets de dades circulen per les mateixes xarxes i són encaminats pels mateixos centres de distribució (encaminadors) com qualsevol altra dada, i de vegades es desordenen o es perden.

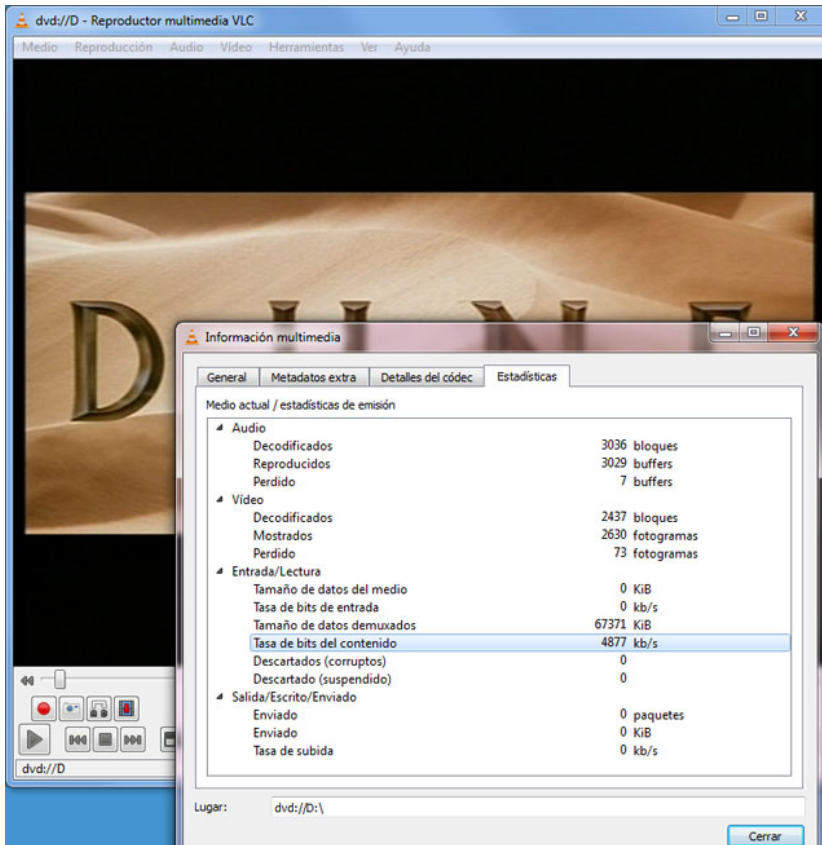
b) La segona raó és un flux de dades molt alt, gairebé constant, per la qual cosa necessita gran quantitat de paquets per a assegurar la reproducció amb la cadència necessària: per exemple, un vídeo de YouTube d'un minut pot exigir lliurar uns mil paquets sense perdre l'ordre i amb una cadència fixa. Tot això ho veurem amb detall en altres mòduls.

Sens dubte entre el vídeo per Internet i el vídeo de difusió hi ha diferències crítiques en la seva distribució que obliguen a reptes tècnics en part superats. Però més enllà de la tecnologia hi ha una revolució més profunda i de conseqüències més intenses: l'usuari tria el que vol veure des del punt en què vol veure-ho, en contraposició a veure el que t'ofereixen tal com indica la programació estàtica del canal. Aquesta nova capacitat de decisió per part de l'usuari final obliga a replantejar l'atracció del producte i a crear mercats nous per a aprofitar la interacció. Aquesta revolució ja està passant, encara que alguns sectors ho neguin amb arguments com la falta de qualitat d'imatge o de tractament professional assegurat dels continguts de la Xarxa; però aquests arguments perden força any rere any en incorporar-se professionals preparats per a treballar en aquest mitjà com el seu entorn natural. Esperem aportar-hi el nostre gra de sorra.

Activitats

Activitat 6

Comparem els nivells de compressió que s'utilitzen habitualment al mercat. Per a això, instal·lem el programa de reproducció de vídeos VLC (<http://www.videolan.org>).



Taxes de bits informades per VLC

Primer vegem la compressió de vídeo utilitzada per a poder emmagatzemar una pel·lícula en un DVD domèstic, per a això amb el menú "Medio / Abrir disco" seleccionem la unitat de disc i reproduïm una pel·lícula. Accedint en qualsevol moment al menú "Herramientas / Información multimedia / Detalles del codec", en el qual a l'apartat VÍDEO s'indicarà el format de codificació, usualment MPEG-2, l'ample i l'alt de la imatge en píxels i els fotogrames per segon. És fàcil calcular, doncs, que el flux de dades de lectura en bits hauria de ser, sense compressió, de:

$$\text{Flux de dades} = \text{amplada (píxels)} \times \text{alçària (píxels)} \times 8 \frac{\text{bits}}{\text{color}} \times 3 \text{ colors} \times \frac{\text{fotogr.}}{\text{segon}}$$

Ara, a la solapa Estadístiques, es pot veure la Taxa de bits del contingut, els bits per segon necessaris per a reproduir correctament el fitxer comprimit:



Taxes de bits a la solapa Estadístiques de VLC

La compressió que ha aconseguit, usualment molt alta, es pot calcular de la manera següent:

$$\text{Factor de compressió} = \frac{\text{flux de dades sense comprimir (bps)}}{\text{flux de dades comprimit (bps)}}$$

Activitat 7

Feu la mateixa operació, però en aquest cas amb un fitxer de vídeo que s'hagi descarregat d'Internet (WMV, MP4, AVI, etc.). Per visualitzar-lo amb VLC, seleccioneu el menú Suport/Obre el fitxer i recolliu les dades que presenta. Calculeu el factor de compressió i compareu-lo amb el del cas DVD. Per què hi ha aquesta diferència?

Si en comptes de DVD feu servir Blu-ray, compareu-lo amb un fitxer de vídeo que sigui en alta definició, o la comparativa no tindrà sentit.

1.5. Resolucions, velocitat de quadre, relacions d'aspecte i altres paràmetres

El **quadre** és la unitat d'imatge en vídeo, una imatge amb tot el contingut visual captat.

No obstant això, a l'hora de presentar-lo en pantalla es pot fer de dues maneres:

- **En mode progressiu:** completament, és a dir, presentant totes i cadascuna de les files de píxels l'una sota l'altra, fins a finalitzar la presentació del quadre.
- **En mode entrelaçat:** el quadre es descompon en dues subimatges, els camps, l'un amb les files imparelles, i l'altre amb les parelles. A continuació es presenta el primer camp i després el segon.

Històricament, el mode entrelaçat va possibilitar emetre televisió amb una amplada de banda petita (era més senzill emetre dues subimatges alternativament que una imatge gran) però a costa de generar un efecte de vibració visual a les zones d'imatge amb detall brillant, denominat *parpelleig*. Algunes persones són més receptives a aquest problema que d'altres.

La tècnica més habitual per a reduir el pampallugueig d'un senyal entrelaçat ha estat simular amb més o menys gràcia una freqüència de quadre major en el receptor: un monitor que prometi 100 Hz, 200Hz, i fins i tot 600Hz, intenta millorar l'experiència visual, reduint el pampallugueig, però fent artificis que, en general, generen altres problemes visuals col·laterals. En vídeo digital, que és el nostre camp de treball, tot programari de tractament de vídeo té alguna eina per a reduir els efectes visuals de l'entrelaçat.

És curiosa l'existència de la freqüència de quadre nova de 24 imatges/segon en televisió digital. Fins ara, tota transferència de material cinematogràfic a vídeo havia estat a costa de perdre qualitat visual. Si parlem de pel·lícula en 16 mm o 35 mm, rodades a 24 imatges per segon, en transferir-la a vídeo amb una càmera de 25 imatges/segon apareixien transicions visuals fosques constantment, per la qual cosa es va optar per a reproduir-la a 25 imatges/segon, i accelerar així l'acció en uns 4 minuts de cada 100 (una pel·lícula de 100 minuts al cinema dura 96 en televisió). Gràcies a la introducció de la freqüència de vídeo de 24 imatges/segon en els formats 720 i 1080 aquests fenòmens desapareixen i es pot capturar, emmagatzemar i reproduir digitalment el material de cel·luloide tal com es va fer.

Mode progressiu

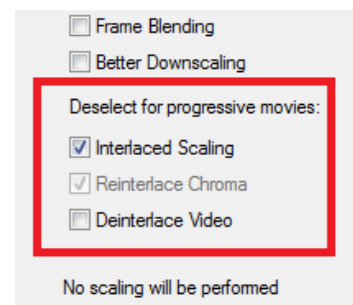
A Europa, els modes 720p, 1080p i 1080p50, a 25, 24 i 50 quadres/segon o Hz respectivament.

Mode entrelaçat

A Europa, els modes 576i, PAL i 1080i, a 50 camps/segon o Hz (25 quadres/segon).



El ratllat en els perfils dels objectes és indicatiu d'entrelaçat

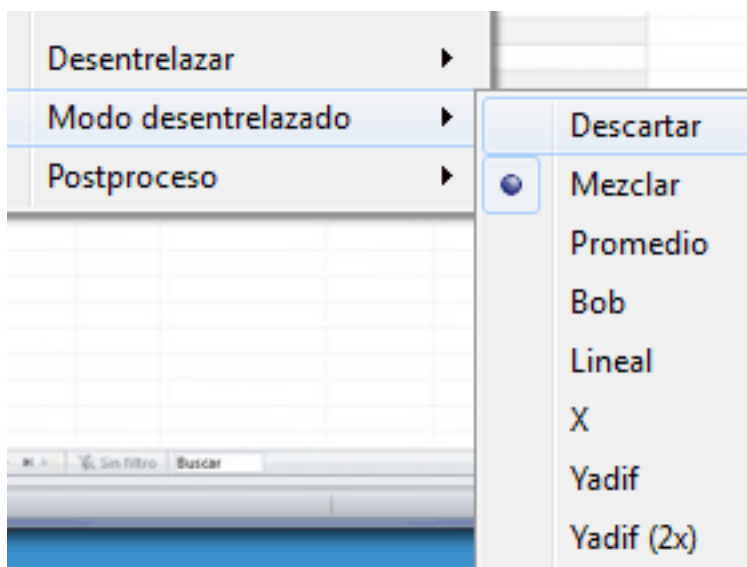


Opcions típiques contra l'entrelaçat

Avui, el mode entrellaçat i el progressiu conviuen. En realitat, qualsevol sistema que superi els 18 quadres/segon és factible, però **com més gran és la velocitat de quadre més realista és l'experiència**, encara que requereixi més velocitat de dades. En casos específics com la videoconferència, un sistema purament progressiu, aquest límit és aconsellable encara que si no es compleix no perd el valor afegit d'una comunicació a gran distància per un cost molt baix.

Activitat 8

Obriu el VLC i amb el menú Suport/Obre el fitxer visualitzeu els vídeos indicats a l'aula. Aquest programa té capacitat de desentrellaçar amb diferents tècniques enumerades en el menú Vídeo/Mode desentrellaçat.



Opcions de desentrellaçat de VLC

Una vegada seleccionada una d'aquestes tècniques, es pot aplicar i veure el resultat amb Vídeo/Desentrellaça/Activat. Seleccioneu el que trobeu més adequat per a cada vídeo.

Pot desactivar-se amb Vídeo/Desentrellaça/Inactiu.

Respecte a la **resolució d'imatge**, usualment parlem de la resolució vertical, que es defineix com el nombre de línies horitzontals d'imatge que es poden distingir en recórrer la imatge en vertical.

Així, en PAL o DVD són 576 línies d'un total de 625 (la resta són línies de sincronia sense imatge), i en Blu-ray són 720 (mode 720p) i 1080 (modes 1080i i 1080p).

La resolució de la imatge, no obstant això, està limitada per la resolució de la pantalla sobre la qual es mostra, i viceversa. Així, mides de 640 × 480 (el ja gairebé inexistent VGA) fins a 1.600 × 1.200 (UXGA) són les més habituals en l'entorn informàtic, mentre que resolucions de 1.920 × 1.080 són les pròpies de

Videoconferència

En videoconferència, per exemple, s'utilitza sovint el format denominat CIF (*common intermediate format*) –de 352 píxels d'ample i 288 línies de resolució vertical–, o el 4CIF –que quadruplica aquesta mida fins a 704 × 576) o el QCIF –un quart del CIF, 176 × 144).

monitors Full HD domèstics, tot i que aquests dos segments de mercat sembla que tendeixen a una destinació comuna (la unificació augurada del televisor domèstic com a centre digital).

Finalment, el tercer paràmetre que s'ha de destacar és la **relació d'aspecte**, que indica la relació entre l'ample i l'alt de la imatge.

Actualment l'estàndard en vídeo és 16:9 (més còmode que dir *1,78*). Estudis van indicar que aquest rectangle podia abastar els formats d'aspecte més freqüents que hi ha hagut fins ara (4:3 convencional, 1.85:1 i 2.39:1 cinematogràfic, 3:2 fotogràfic), ja sigui utilitzant el zoom o amb barres negres als costats (*pillarbox*) o a dalt i a baix (*letterbox*).

David Sarnoff Center

Aquests estudis els va dirigir el 1980 el doctor Kerns Powers del David Sarnoff Center i membre de l'SMPTE.



Comparativa dels diferents formats d'imatge

La relació d'aspecte depèn de la proporció del píxel: al món de la informàtica, els píxels són quadrats. Així, una imatge 4:3 de 1.024×768 (VGA) té una relació exacta de 4:3, i una imatge 16:9 de 1.600×900 té una relació de 16:9. No obstant això, al món de la imatge no és així: el vídeo convencional de 720×576 és 4:3 però la relació dels seus píxels no:

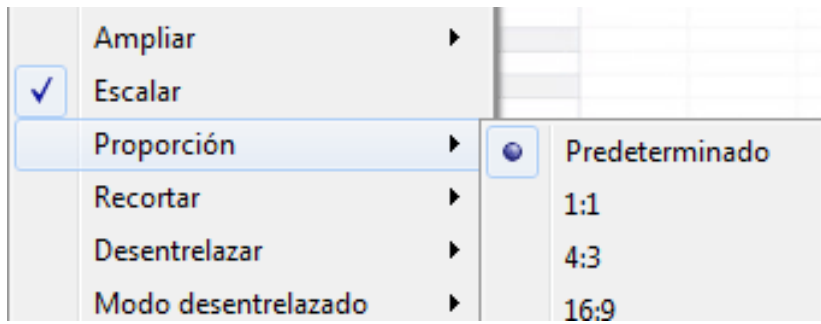
$$\frac{720}{576} = \frac{5}{4} = 1,25$$

Sovint treballar amb senyals de resolució estàndard en entorns informàtics suposa haver d'encongir horitzontalment la imatge per a veure-la en la relació correcta, cosa que els programes de captura ja fan internament. Aquest problema no hi és amb els senyals d'alta definició, ja que la majoria dels seus for-

mats són de píxel quadrat: 1.280×720 (720p) i 1.920×1.080 (1080i i 1080p) (16:9). El format 1.440×1.080 (1080i) és l'excepció, amb un píxel representat en pantalla lleugerament rectangular.

Activitat 9

Amb el programa VLC es poden veure els dos formats habituals de treball en vídeo digital. El menú és Vídeo/Ràtio de l'aspecte i els valors van des de Per defecte (el programa tria la relació segons les dades del mateix fitxer de vídeo) fins a relacions d'aspecte més rectangulars. Comproveu que els vídeos en alta definició sempre són en relació 16:9.



Selecció de relacions en VLC

2. Característiques del so

La distribució d'imatge en gairebé qualsevol aspecte va associada a la distribució d'àudio, per això cal tractar-lo en aquest apartat. Com que l'àudio té característiques diferents de la imatge, també es tracta diferent, en general de manera més senzilla. Començarem amb l'àudio analògic i després passarem al digital, i veurem els estàndards de compressió en l'apartat següent.

2.1. Àudio analògic

Convertit l'àudio en un senyal elèctric, aquest pot alimentar directament l'equip, altaveu, cascos, etc. en què s'hagi de reproduir. Salvant la potència necessària, aquest sistema d'àudio abraça des de micròfons fins a sistemes de bafles de concerts amb èxit des de fa dècades, i sobreviu en l'era digital amb les sortides RCA dels equips CD, DVD, descodificadors de cable i satèl·lit, o les sortides d'auriculars de 3,5 mm de tot reproductor MP3, etc. No obstant això, el soroll, les interferències, la qualitat dels connectors i del cable afecten el seu resultat final; a més no inclou cap senyalització que faciliti la seva sincronia amb el vídeo que pugui acompanyar, per la qual cosa se circumscriu actualment a l'entorn de consum.



Connectors d'àudio estèreo analògic. Minijack (a dalt) i parella RCA (a baix)

2.2. Àudio digital

El so sempre serà analògic, però la transmissió i emmagatzematge s'han digitalitzat, i el protegeixen així del soroll i el fan immune a les interferències, alhora que el compatibilitzen amb els sistemes digitals. A més, des del CD fins a l'MP3, l'àudio digital ha passat de ser no comprimit a utilitzar tècniques de compressió que veurem més endavant. Necessita connectors Canon XLS, encara que el format AES també s'utilitza per a comprimir l'àudio digital en el senyal de vídeo SDI, HD-SDI o 3G, i transmetre tot el conjunt per un únic cable coaxial.

El senyal d'àudio digital segueix el format AES/EBU, amb cada mostra de 16 bits i a velocitats de 32.000, 44.100 i 48.000 mostres segons l'amplada de banda analògica que es vulgui representar (fins a 16 kHz, qualitat FM, fins a 22 kHz, qualitat CD, i fins a 24 kHz, qualitat d'estudi professional, respectivament). Sens dubte segueix el criteri de Nyquist. Així, un senyal estèreo digital AES suposa un flux de dades de:

Flux d'àudio = 44.100 mostres × 16 bits/mostra × 2 canals = 1,41 Mbps

Si a més sabem que l'estàndard AES reserva 32 bits per a transportar cada mostra i altres dades associades (com el codi de temps), resulta:

Vegeu també

Les tècniques de compressió es tracten en el mòdul "Digitalització, emmagatzematge i transmissió d'àudio i vídeo" d'aquesta assignatura.



Connectors Canon d'àudio digital (AES)

Flux de transport d'àudio = flux d'àudio \times 2 = 2.82 Mbps

Això suposaria gairebé una connexió de 3 Mbps només per a reproduir l'àudio estèreo per una xarxa, el cas d'àudio multicanal o envoltant multiplica per 3 o 6 aquesta xifra. Evidentment s'ha de poder comprimir sense perdre gran qualitat per a poder distribuir àudio per xarxa, cosa que veurem més endavant.

Activitat 10

Amb el VLC, obriu un fitxer d'àudio i mentre està en reproducció accediu a Eines/Informació dels còdecs. A la pestanya Detalls del còdec indica el format en el qual està comprimit l'àudio, i a la pestanya Estadístiques hi ha una sèrie de dades en què destaca especialment Taxa de bits d'entrada, és a dir, els kilobits per segon que el programa va llegint del fitxer d'àudio, Taxa de bits del contingut, que indica quants dels kilobits per segon llegits realment són per a àudio (els altres són de control, qualitat, sincronia, etc.).

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

Dades que proporciona VLC d'un fitxer d'àudio

Compareu la taxa de bits de contingut amb el flux d'àudio (no el flux de transport d'àudio) que hauria de tenir en format AES per a donar una idea de la compressió aconseguida amb el format d'àudio del fitxer perdent la qualitat de so que es pugui detectar.

Proveu-ho amb formats diferents (MP3, WMA, WAV, etc.)

3. La revolució en la gestió dels senyals

Una vegada hem caracteritzat els senyals principals, ens queden altres informacions addicionals que tenen definit el seu lloc en l'entorn domèstic i professional, com l'audiodescripció per a persones amb discapacitat visual, el teletext d'informació general, els subtítols per al col·lectiu amb diferents graus de discapacitat auditiva, etc. Aquests senyals haurien de poder subsistir o tenir equivalents funcionals en qualsevol plataforma de distribució digital, sempre a costa d'un consum més alt d'amplada de banda, però els beneficis pel que fa a l'audiència són evidents. Fins i tot a certs països són obligats.

La gestió d'aquest cúmul d'informació lligada (vídeo, àudio, serveis addicionals) en una productora audiovisual es fa mitjançant equips denominats **matrius** de vídeo i àudio, que seleccionen els senyals adequats del conjunt de senyals que hi ha per a aconseguir un producte resultant coherent. Sense entrar en els detalls, hi ha hagut matrius analògiques i actualment totes són digitals.



Connexions d'entrada i sortida d'una matriu de vídeo i àudio

En haver de treballar amb senyals totalment digitals, les condicions són les mateixes que les de treball d'un encaminador *router* IP, aparell concentrador de dades que encamina els paquets digitals de dades des d'un origen fins a una destinació. La fusió entre les matrius digitals d'àudio/vídeo i els encaminadors IP no ha tardat a arribar, tenint en compte les funcions avançades dels encaminadors IP d'última generació en entorns de molt alta velocitat. Així, les instal·lacions tradicionals d'àudio i vídeo amb cables coaxials dedicats des de cada equip fins a la matriu va desapareixent, i són substituïdes per una xarxa d'alta velocitat en la qual es connecten tots els equips a l'encaminador IP central. Una xarxa escalable, és a dir, una xarxa la grandària de la qual és fàcilment ampliable si calen equips nous per a la producció.

Aquest canvi substancial en el nucli de producció a més unifica la gestió de la resta de senyals, incloent la telefonia interna, la gestió remota dels equips, la integració de la producció a la xarxa de distribució al client, a altres centres de producció o directament al mercat Internet, etc. Sens dubte hi ha un altre tipus de problemes nou, com la lluita contra el retard de les dades i la millora de la robustesa del sistema davant la congestió de la xarxa i la pèrdua conseqüent de paquets de dades.

Si en el camp de la distribució audiovisual, la Xarxa ja tenia guanyada la batalla, en el camp de la producció audiovisual la victòria és clara a mitjà termini. Molts equips de difusió ja inclouen connexions de xarxa (fibra o coure) d'alta velocitat (4 Gbps d'ara endavant) preparats per a aquest canvi de procés. El transport de vídeo i àudio per IP augmentarà vertiginosament.

4. Codificació d'imatge fixa

4.1. La necessitat de comprimir

Les imatges requereixen considerablement més espai que el text, i l'àudio i el vídeo encara en demanen més en termes d'emmagatzematge i velocitat de transmissió. Il·lustrarem aquest aspecte amb unes quantes xifres i introduïrem així alguns conceptes bàsics que utilitzarem al llarg de tot el curs a partir d'aquest punt:

- Una pàgina d'aquest llibre conté unes 35 línies de 80 caràcters; si cada caràcter pot representar-se, emmagatzemar-se i transmetre's digitalment amb 1 byte (8 bits), cada pàgina ocupa 22.400 bits.
- Una imatge de vectors (és a dir, creada amb línies, com els dissenys d'Autocad o una font escalable) conté uns 500 traços. Si cada traç es representa per les seves coordenades d'inici i final (quatre nombres en total) i cada nombre per 2 bytes, la imatge ocupa 32.000 bits.
- La música en qualitat CD suposa 44.100 mostres/segon, si les mostres són de 16 bits i és estèreo, es generen 1.411.200 bits cada segon.
- Una imatge fotogràfica de 720×576 píxels conté 414.720 píxels en total, si la imatge és en blanc i negre cada píxel es representa per un byte, amb la qual cosa ocupa 3.317.760 bits.
- Si la imatge anterior fos en color, cada píxel es representa per 3 bytes (un per al vermell, un altre per al verd i un altre per al blau), i ocuparia el triple, 9.953.280 bits.
- El so digital envoltant pot arribar a suposar gairebé 9 Mbps sense cap tractament.
- En el cas de vídeo, un senyal PAL de 625 línies a 25 quadres/segon es descompon en luminància i dos senyals de crominància que necessiten uns 216 Mbps.
- Si és HD a 1080i, augmenta a 1,15 Gbps.

Els requeriments de memòria i velocitat de transmissió per a aquestes informacions s'han de reduir de manera notable perquè els puguin gestionar els equips actuals, amb capacitats d'emmagatzematge de l'ordre de Gbytes i memòries intermèdies de l'ordre de Mbytes, i de sistemes de transmissió que rondin alguns Mbps. Necessitem comprimir, i en alguns casos de manera dràstica.

4.1.1. Codificació i compressió. Tipus de compressió

És usual que s'utilitzin els termes *codificació* i *compressió* de manera equivalent, però **codificació** senzillament expressa una informació d'una manera diferent amb un codi nou, mentre que **compressió** és reduir la mida total en bits. La compressió és un cas particular de codificació, en el qual el codi nou utilitza menys elements per a representar la mateixa informació.

Així, la taula ASCII codifica els caràcters i els expressa en format binari, i no hi ha cap compressió en el procés. En canvi, una codificació que recorri a una eina informàtica que representi un fitxer de dades amb un altre d'una mida més petita en bits (com ZIP o RAR en Windows, SIT en OSX o TG en Linux) és un cas de compressió.

No obstant això, encara podem perfilar més la definició de compressió: quan indiquem que ha de representar la mateixa informació, no significa en tots els casos que hagi de ser exactament la mateixa informació final que l'original, és a dir, la compressió pot ser **sense pèrdues** o **amb pèrdues**. En el primer cas podem fer el procés invers per a tornar a tenir la mateixa informació exacta original: de tot fitxer ZIP o RAR se'n pot recuperar el fitxer original. En el segon cas acceptem una degradació del missatge per a aconseguir una compressió efectiva, sempre que no faci inservible la informació que porta.

4.1.2. Redundància i entropia

A què ens referim amb informació inservible? En molts casos la informació que es vol transmetre té **dades redundants**, informació previsible o repetitiva que es pot comprimir de manera intensa, com per exemple diversos espais en blanc consecutius en un text, o diversos píxels consecutius d'un mateix color en un fax. La seva reconstrucció en el moment de descomprimir és exacta. En la compressió sense pèrdues s'elimina la redundància de la informació, sense tocar res de la informació fonamental. És la ideal per a fitxers de dades (ZIP, RAR, SIT, TG, etc.), alguns formats d'imatge com GIF, PIC, un mode de treball específic del format de compressió d'imatge JPEG 2000, etc.

No obstant això, aquestes compressions sense pèrdues no rendeixen gaire. Aconseguir un 50% de reducció és generalment un èxit, per la qual cosa la compressió amb pèrdues és la que domina en el món de la imatge i el so, on calen taxes de reducció molt més altes.

La base de la compressió amb pèrdues és l'eliminació de les **dades irrellevants**, com per exemple les petites variacions de color entre píxels propers (ja que l'ull humà té poca resolució per al color), o els sons de freqüències superiors als 16.000 Hz (perquè pocs éssers humans adults són capaços de sentir-los). La informació irrellevant se suprimeix completament, i es modifica permanentment la informació que es vol transmetre, encara que d'una manera tolerable per al destinatari. Alguns exemples de formats estàndard de compressió amb pèrdues són els següents:

- Formats de **compressió d'imatge** com JPEG, EZW, SPIHT i gairebé tots els modes de treball de JPEG 2000.
- Formats de **compressió d'àudio** com AAC o MP3.
- Formats de **compressió de vídeo** com la família MPEG.

Evidentment l'èxit de qualsevol compressió dependrà de la informació redundant o irrellevant que hi hagi en el missatge, per la qual cosa és habitual que el rendiment de la compressió variï gairebé constantment. L'entropia és una mesura de la informació essencial d'un missatge (la diferència entre la quantitat total de dades que conté el missatge i la seva redundància). Com més entropia, més dades no es poden eliminar, per la qual cosa la codificació amb pèrdues rendirà poc. Com menys entropia, la codificació amb pèrdues serà molt efectiva.

4.1.3. Mesura de la compressió d'imatge

Per a mesurar la compressió aconseguida en una imatge i poder comparar diferents sistemes de compressió com farem en el curs, utilitzem el *factor de compressió* que ja hem vist, un valor numèric que augmenta amb la compressió aconseguida. Així, donada una imatge original i el seu equivalent comprimit, el factor de compressió és:

$$\text{Factor de compressió} = \frac{\text{Pes de la imatge original}}{\text{Pes de la imatge comprimida}}$$

Una altra manera de representar la compressió és indicant el nombre de bits que necessita cada píxel de la imatge, el bit per píxel (bpp):

$$\text{bpp} = \frac{\text{Pes de la imatge (expressat en bits)}}{\text{Nombre total de píxels de la imatge}}$$

Nota

Recordeu indicar tots dos pesos sempre en la mateixa unitat (bits, bytes, Kbytes, etc.).

De manera senzilla podem deduir que el bpp d'una imatge original i el bpp de la seva imatge comprimida es relacionen amb el factor de compressió entre si:

$$\text{Factor de compressió} = \frac{\text{bpp de la imatge original}}{\text{bpp de la imatge comprimida}}$$

Activitat 11

Si una imatge és de 1.600×1.200 píxels amb 24 bits de color/píxel, i es comprimeix en un format tal que el fitxer resultant és de 230 Kbytes, quin és el bpp de la imatge original? I el de la comprimida? Quin serà el factor de compressió aconseguit? Atenció amb les unitats, no barregeu bits amb bytes.

4.1.4. Mesura objectiva i subjectiva de la qualitat d'imatge

Actualment se segueixen investigant tècniques quantitatives que puguin assignar un valor numèric a la qualitat visual de qualsevol imatge, que ens donen una idea de la degradació que hi ha hagut després d'aplicar certa compressió amb pèrdues i que ens ajuden a triar l'estàndard de compressió més adequat.

En tot cas, presentem una mesura, l'**error quadràtic mitjà**, una manera de calcular les diferències que hi ha entre una imatge i la mateixa imatge comprimida. Bàsicament és una aproximació a veure en què són diferents, cosa que l'ull humà fa de manera natural en instants però que matemàticament és molt difícil de concretar, ja que no se saben amb detall els mecanismes del cervell humà que ho fan possible. La idea és restar píxel a píxel l'una de l'altra i sumar les diferències; amb la qual cosa, com més s'apropa a zero el resultat, més semblants són les imatges.

No obstant això, és una mesura no del tot fiable, ja que els errors visuals són percebuts per l'ull humà sobretot per l'entorn en què apareixen, cosa que aquesta eina no valora. Després de veure el format JPEG veurem una manera molt més visual i entretinguda de detectar els errors amb l'aplicació Photoshop, capaç de calcular la imatge diferència, és a dir, una resta visual píxel a píxel entre, per exemple, una imatge sense comprimir i el seu equivalent comprimit, que posa en evidència la informació perduda en la compressió.

4.2. JPEG

Comencem veient el primer **sistema de codificació d'imatge fixa** i introduint una sèrie de conceptes que utilitzarem al llarg de tot el curs. JPEG és un comitè (Joint Photographic Experts Group), no un estàndard de codificació, que a la fi dels anys vuitanta va idear un procés pel qual una imatge digital pot comprimir-se mantenint un compromís amb la qualitat visual. En aquells anys ja començava a ser necessari algun sistema perquè les imatges no ocupessin els espais reduïts d'emmagatzematge digital que hi havia, i per a poder ser transmeses per línies de telefonia digital de baixa velocitat. En general, el procés és prou robust i té uns resultats acceptables, especialment amb fotografies na-

turals i imatges realístiques, i permet arribar a una mitjana de compressió de 10:1 amb pèrdues imperceptibles visualment, per la qual cosa encara és un format molt utilitzat.

Per primer cop en aquest curs hem de distingir entre el sistema de codificació (còdec) de la imatge, que és JPEG, i el **contenidor** o format de fitxer on es guarden les dades resultants de la codificació, del qual hi ha dos tipus, JFIF i EXIF. Així, JPEG/JFIF és el format clàssic, mentre que JPEG/EXIF permet afegir **metadades**. Avui alguns mòbils amb càmera i GPS integrats afegeixen la longitud i latitud en què va ser feta. El nostre interès se centra en el còdec.

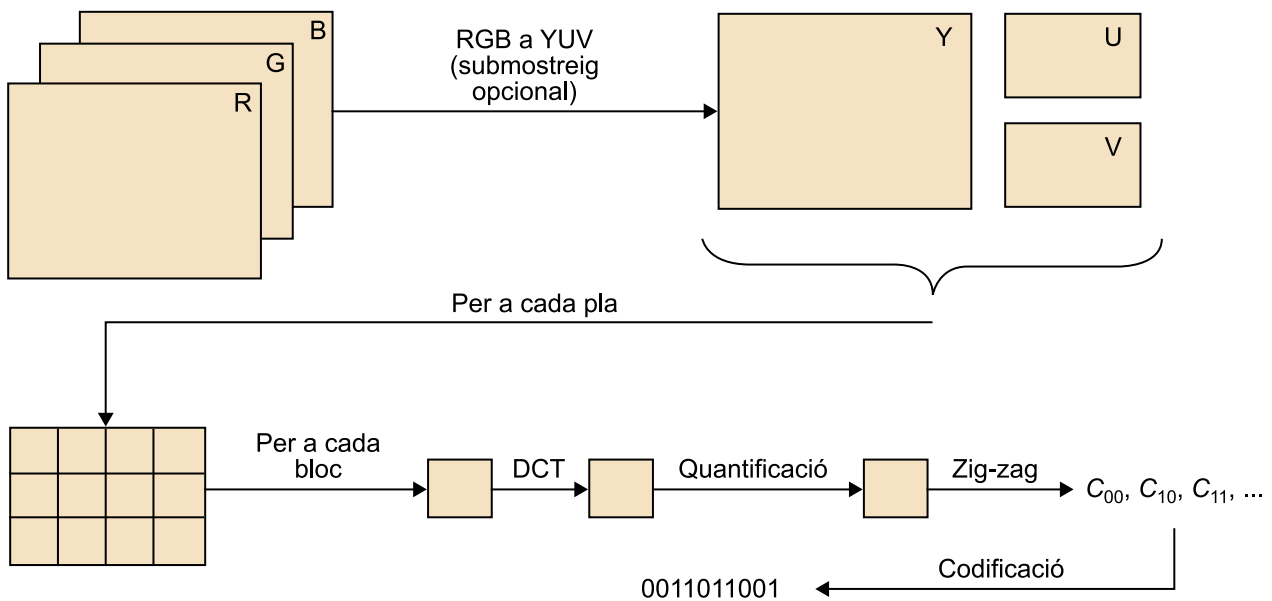
Metadades

Són les dades addicionals associades a la imatge, com per exemple, la data i hora de la captura, velocitat de l'obturador, nom de la càmera, dades de configuració, etc.

JPEG pot treballar en diversos modes diferents, encara que només detallarem dos modes **amb pèrdues**: el més utilitzat, el **mode seqüencial**, amb el qual introduïrem diversos conceptes de codificació rellevants en cada pas del procés de codificació, i el **mode progressiu**.

4.2.1. Mode seqüencial

Un esquema general dels diferents passos que es fan en aquest mode seqüencial és el següent:



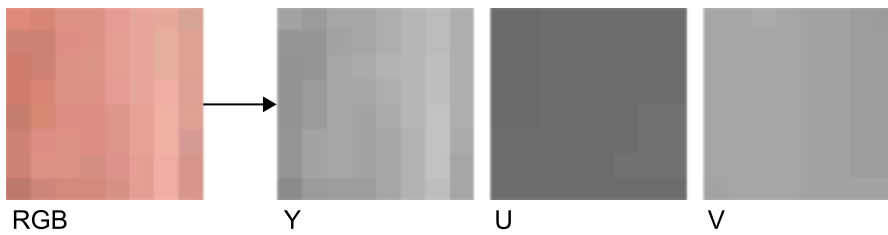
Esquema general dels passos en el mode seqüencial

1) Pas 1: canvi d'espai de color i submostreig del color

Si la imatge és RGB, en primer lloc es transforma a YUV. Com ja hem vist al principi d'aquest mòdul, com que l'ull humà té menys resolució de color, les imatges U i V se submostregen i redueixen el seu nombre de píxels. La imatge Y (blanc i negre) es manté inalterada.

Reduir el nombre de píxels

La reducció es fa en relació 4:4:4, 4:2:2 o 4:2:0 segons la qualitat triada.



2) Pas 2: divisió en blocs i seqüenciació

Cadascun dels tres components es comprimeix per separat, i es divideix en blocs de 8×8 píxels. Així, cada component es converteix en una seqüència de blocs, que comença pel de la seva cantonada superior esquerra i acaba amb el de la seva cantonada inferior dreta. A partir d'aquest punt el procés es fa individualment per cada bloc.

3) Pas 3: codificació de cada bloc

Cada bloc d'imatge de 8×8 píxels es transforma matemàticament en un altre bloc de 8×8 valors (també sencers) que ja no representen una imatge petita sinó les **frequències espacials** d'aquesta. El bloc, doncs, conté la mateixa informació que tenia, però amb altres valors: un pur exemple de codificació. I sense pèrdues fins a aquest punt, ja que podríem fer el procés invers i recuperar exactament la imatge petita original.

Blocs de píxels

Els píxels són de valor sencer de 0 a 255, però en aquest punt es resta 128 a cadascun perquè estiguin al marge de -128 a 127. Aquesta operació tan curiosa prèvia a la codificació s'explica en el pas següent.

La transformació utilitzada es denomina DCT ('Discrete Cosinus Transform') i és la base de la compressió d'imatge i vídeo avui dia.



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	<i>-4</i>	<i>-21</i>	<i>4</i>	<i>2</i>	<i>0</i>	<i>-1</i>	<i>1</i>
23	23	35	43	55	70	76	47		<i>-68</i>	<i>4</i>	<i>9</i>	<i>21</i>	<i>9</i>	<i>5</i>	<i>-1</i>	<i>-1</i>
23	30	36	34	43	68	76	47		<i>4</i>	<i>0</i>	<i>-7</i>	<i>12</i>	<i>-2</i>	<i>0</i>	<i>-1</i>	<i>1</i>
22	39	35	31	41	67	80	47		<i>-18</i>	<i>-3</i>	<i>6</i>	<i>6</i>	<i>3</i>	<i>-2</i>	<i>0</i>	<i>0</i>
25	33	37	31	36	68	80	50		<i>1</i>	<i>3</i>	<i>2</i>	<i>2</i>	<i>1</i>	<i>-1</i>	<i>0</i>	<i>0</i>
32	33	35	30	35	68	80	35		<i>-2</i>	<i>0</i>	<i>0</i>	<i>-1</i>	<i>0</i>	<i>-1</i>	<i>0</i>	<i>0</i>
10	31	28	28	35	69	73	35		<i>1</i>	<i>-1</i>	<i>3</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>

4) Pas 4: quantificació de cada bloc

Si tots dos blocs contenen la mateixa informació, per què bona part dels valors o coeficients del segon bloc són de valor molt reduït, fins i tot nul, respecte als restants? Aquest és el quid del procés: la transformació DCT compacta la informació associada a la imatge en els primers coeficients (situats prop de la cantonada superior esquerra), cosa que ens permetria eliminar bona part de la resta amb valors nuls o de baix valor (indicats en vermell i cursiva en la imatge adjunta), i reduir el nombre total de coeficients que representen el bloc.

Però quin criteri triem per a eliminar els coeficients petits? I què vol dir *petit*? Per a automatitzar la selecció dels coeficients que s'eliminaran en cada bloc d'una manera raonable s'utilitza una taula de referència denominada de **quantificació**, que conté 8 × 8 valors de pes o ponderació, definits després de moltes proves de qualitat subjectiva fetes amb grups d'observadors. Aquesta taula conté pesos petits en les posicions properes a la cantonada superior esquerra i pesos creixents com més lluny n'estan. Cada coeficient de la imatge transformada es divideix pel pes que li correspon per posició, així els allunyats de la cantonada superior esquerra queden dividits per pesos elevats, i donen resultat zero (no es tenen en compte els decimals).

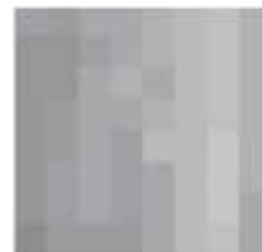
Exemple

Vegem un exemple de taula de quantificació utilitzada 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 féssim la transformació inversa, denominada IDCT, per a recuperar el bloc d'imatge original a partir del bloc transformat i quantificat, l'efecte visual de la pèrdua de coeficients és mínim. En l'exemple següent, descartar els indicats en vermell no afecta sensiblement el resultat "destransformat".



Valors de coeficient del bloc transformat i degudament quantificat								→	Valors de píxel del bloc visual destransformat resultant							
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	

Sens dubte, emmagatzemar en un fitxer els coeficients transformats que ens interessin en comptes de tots els píxels originals donaria una bona taxa de compressió. Si a més descartem els que tenen valor negligible, la taxa millora sensiblement i es minimitza l'efecte visual de les pèrdues introduïdes quan s'inverteix la transformació i es recupera la imatge. A més, la TCD no necessita cap processador de càlcul potent, i es pot executar amb maquinari senzill. No obstant això, el procés de compressió JPEG encara no ha acabat.

Fem una pausa en el procés d'explicació i raonem les qüestions següents:

a) Com és possible que la DCT faci insensible visualment l'eliminació de bona part dels 64 valors? Després d'anys de proves i comparatives amb totes i cadascuna de les transformacions conegudes, la DCT és la que condensa més bé l'important en menys dades sense necessitat de gaire potència de càlcul (factor que en els anys vuitanta era crític). Hi ha transformacions millors, però totes necessiten un càlcul intens superior, com per exemple JPEG2000, que veurem més endavant en aquest mateix mòdul.

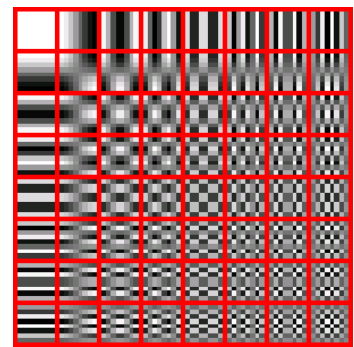
b) Què significa visualment cadascun dels 64 coeficients? Tal com s'ha comentat, aquestes dades generades per la DCT representen freqüències espacials, és a dir, cada coeficient, tot sol, generaria un bloc d'imatge simple corresponent a una oscil·lació visual.

La imatge del marge és un resum del tipus de bloc d'imatge generat per cada coeficient segons la seva posició en el bloc transformat. Per exemple, el coeficient que se situa en la cantonada superior esquerra genera tot sol un bloc d'imatge de 8x8 píxels com l'indicat en aquesta mateixa cantonada, un bloc visualment homogeni. El segon coeficient genera un bloc d'imatge de 8x8 píxels similar a una ona horitzontal de mig període; el tercer, una ona horitzontal d'un període, etc.

Doncs bé: qualsevol fragment de fotografia en blanc i negre de 8x8 píxels que es triï pot representar-se com una suma d'aquests 64 blocs, degudament ponderats (multiplicats) per un valor. Si es tingués temps (anys) i paciència, es podrien anar provant sumes d'aquests blocs degudament amplificats cadascun per un valor triat per nosaltres. El dia que aconseguíssim la suma perfecta que reproduís exactament el fragment de fotografia, descobriríem que els 64 valors seleccionats coincideixen amb els que genera la DCT.

Potència de càlcul en els vuitanta

El 80386 d'Intel era el processador del moment fins al 1989.



Exemple complet de la imatge generada per cada coeficient per separat

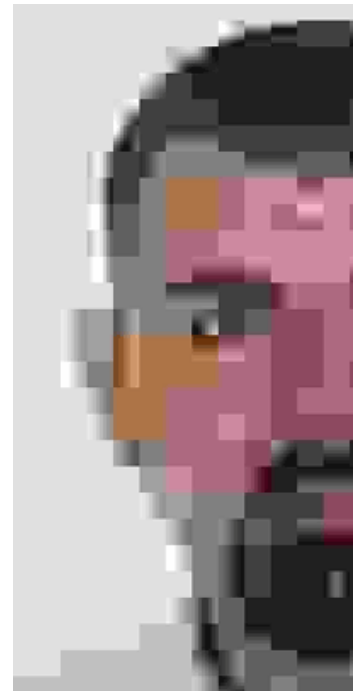
El primer coeficient del bloc és el pes o valor de la freqüència zero que es denomina **component de freqüència contínua o baixa (visual)**, que com es pot veure recrea el valor mitjà de lluentor de tot el bloc, i genera tot sol un bloc visual de 8×8 homogeni. Així, si esborrèssim tots els components del bloc excepte aquest i apliquéssim la transformada inversa per a tornar a tenir un bloc d'imatge, tindríem el que denominem un *efecte mosaic*.

L'ull humà és molt sensible a aquest valor mitjà i s'adona de qualsevol error en aquest valor, de tal manera de qualsevol variació que hi hagi entre dos quadres consecutius genera un efecte mosaic. Per això JPEG cuida el valor d'aquest coeficient com cap altre, com veurem en el pas següent.

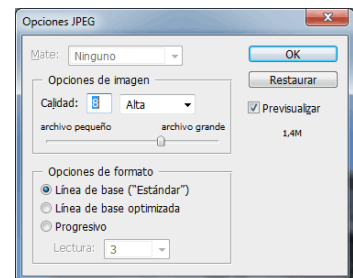
Els 63 coeficients restants es denominen de **freqüència alterna o alta** i són responsables de les variacions de lluentor a partir d'aquest nivell de base. Els blocs d'imatge amb molt detall visual, quan se'ls aplica la DCT, tenen molts coeficients d'aquest tipus amb gran valor, mentre que als blocs de poc detall els passa el contrari.

c) Qualitat d'imatge i les taules de quantificació

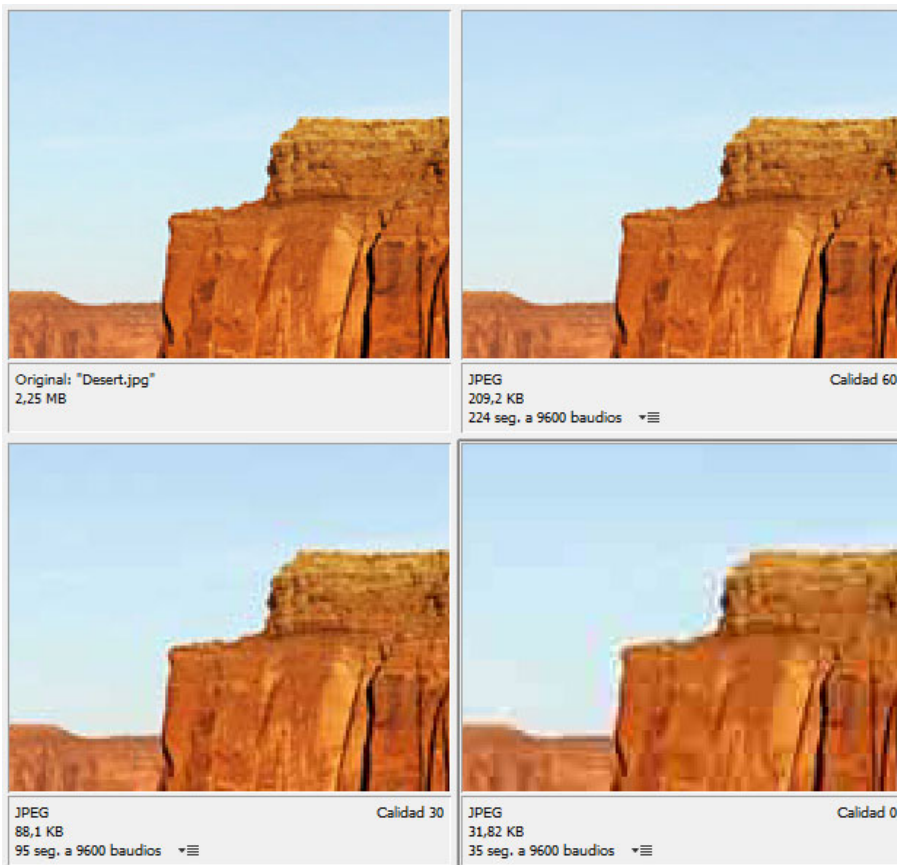
Quan es vol codificar una imatge en format JPEG, el programari (per exemple, Photoshop) pregunta el nivell de qualitat que volem mantenir de la imatge original en la nova imatge JPEG. Usualment és un paràmetre quantitatiu (com un valor de 0 a 10, o de 0 a 12, o un percentatge de 0 a 100%, etc.) en què el primer és la qualitat més baixa i la compressió més alta, i l'últim la qualitat més alta i la compressió més baixa. Altres cops permet seleccionar la qualitat o compressió amb valors qualitatiu (de "menys detall" a "més"), com per exemple en les configuracions de les càmeres digitals.



Imatge codificada només amb els tres primers components de cada bloc



Opcions de compressió JPEG de Photoshop



Quadre comparatiu de compressions JPEG sobre una mateixa imatge

Aquest paràmetre de qualitat seleccionat per l'usuari està íntimament lligat a la taula de quantificació que s'aplica per a eliminar els coeficients d'alta freqüència de tots els blocs transformats de la imatge que es vol codificar. Així, com més alt és el valor de qualitat, més petits són els valors de la taula de quantificació, per la qual cosa la divisió dona un resultat més alt i "sobreviu" més coeficients d'alta freqüència. Quan la qualitat és màxima, els valors de la taula de quantificació són tots 1, cosa que no afecta gens.

També el valor de qualitat pot fer seleccionar al codificador tècniques de submostreig de color 4:2:2 o 4:2:0 per a reduir sensiblement el nombre de blocs de color, i millorar la compressió sense necessitat que l'usuari faci res.

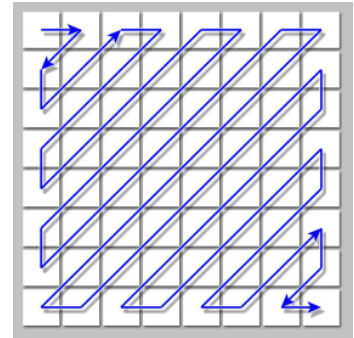
5) Pas 5: codificació dels coeficients

Si bé els 64 coeficients són sencers, en aquest pas es converteixen a binari i s'emmagatzemen en fitxer amb una codificació que redueix la necessitat de bits per al seu emmagatzematge. Com que el coeficient de contínua és tan important, es tracta amb una codificació específica.

En primer lloc s'ordenen seguint un ordre en zig-zag establert. A continuació es codifiquen de la manera següent:

- **Coefficient de baixa freqüència:** guardar la diferència entre dos valors ocupa gairebé sempre menys espai que guardar els dos valors, sempre que aquests valors siguin semblants. És el que es denomina **codificació diferencial (DPCM)**, i s'aplica al coeficient de contínua de cada bloc: ja que en la majoria d'imatges el valor mitjà d'un bloc és usualment similar al del bloc immediatament anterior, s'emmagatzema el valor del primer i després les diferències de cadascun amb el coeficient de contínua següent. Si a més s'emmagatzema en el fitxer JPEG abans que la resta de coeficients, tindrem la imatge mosaic de manera immediata com a vista preliminar.
- **Coefficients d'alta freqüència:** en la seqüència de valors generada per l'ordenació en zig-zag usualment hi ha molts valors repetits, la majoria zeros, per la qual cosa s'apliquen tècniques de compressió que aprofitin aquesta característica (**Huffman**).

Segons com s'emmagatzemin en el fitxer els components de color, parlem de **components entrelaçats** (en què es van emmagatzemant els blocs de luminància i de color alternativament) o de **no entrelaçats** (primer s'emmagatzemen els de luminància i després els de color).



El camí en zig-zag d'ordenació dels coeficients

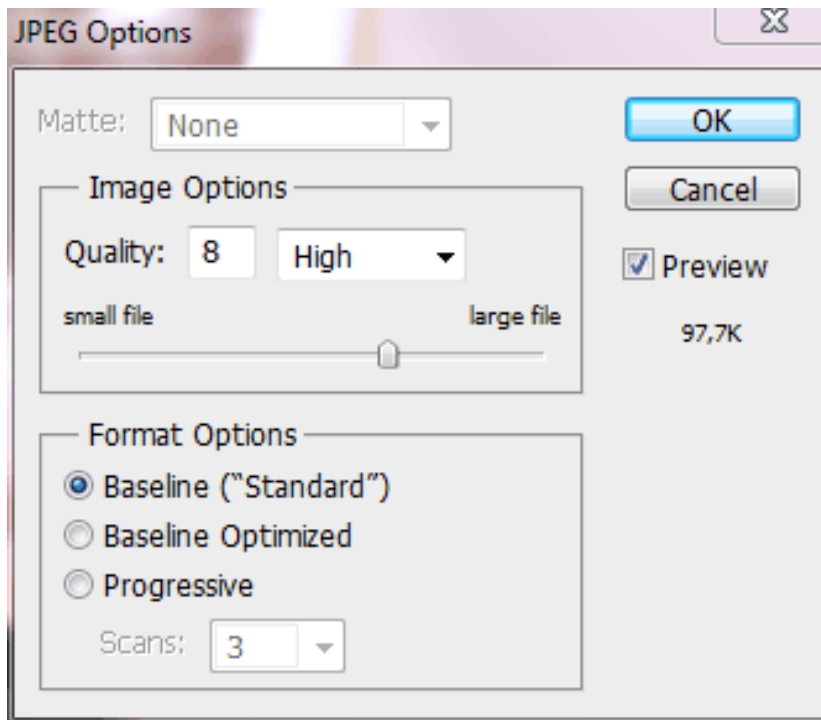
Descodificació d'una imatge JPEG.

El fitxer JPEG té tota la informació necessària per a recuperar la imatge comprimida segons el valor de qualitat amb el qual es va generar, ja que no solament inclou els coeficients transformats de tots els blocs, sinó la informació necessària per a recuperar-los a partir de les versions codificades, el valor de qualitat seleccionat i la taula de quantificació seleccionada per a eliminar elements.

Activitats

Activitat 12

Es tracta de comprimir una imatge en format JPEG amb el Photoshop per a modificar la qualitat de la imatge i la grandària del fitxer resultant. Per a això obriu l'aplicació i llegiu una imatge, a continuació amb *Archivo/Guardar como* seleccioneu un nom nou i com a format d'enregistrament trieu el JPEG (*.jpg, *.jpeg, *.jpe, les tres extensions habituals que s'afegeixen al nom del fitxer). Després d'acceptar, apareix la finestra següent, que demana la qualitat visual a la qual es vol comprimir (un valor de 0 a 12). De moment, mantingueu com a opció el mode estàndard (*baseline*), del qual parlarem en l'apartat següent.

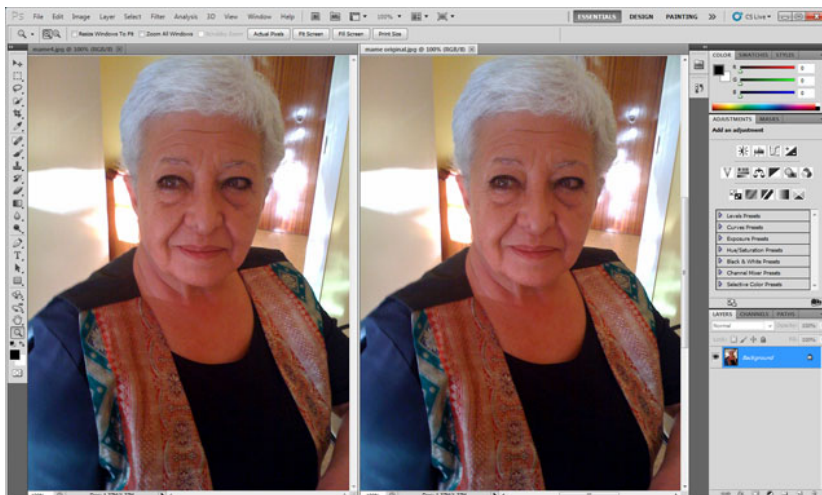


Després d'acceptar, el Photoshop emmagatzemarà el fitxer comprimit. Visualitzeu el resultat i amb el sistema operatiu que tingueu (Windows, OSX, etc.) anoteu la mida en Kbytes del fitxer resultant. Es pot calcular senzillament el factor de compressió.

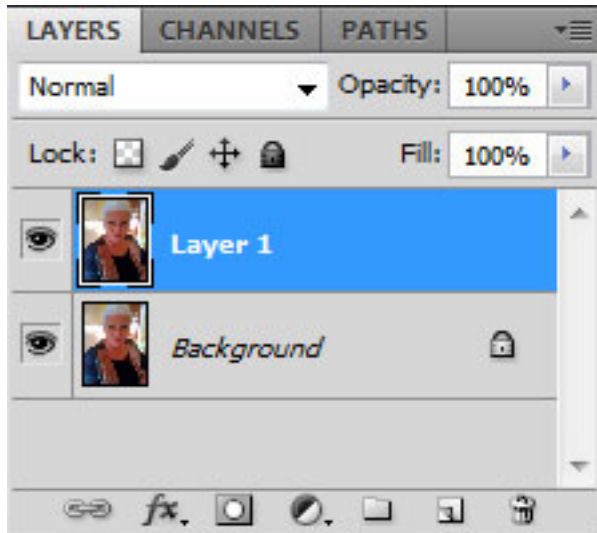
Activitat 13

El factor de compressió pot donar una idea aproximada del resultat, però no és útil si es vol descobrir realment on s'està generant la pèrdua d'informació visual. Per a això es pot crear una imatge diferència que sigui la resta píxel a píxel entre la imatge sense comprimir i la comprimida, en aquest cas, en format JPEG.

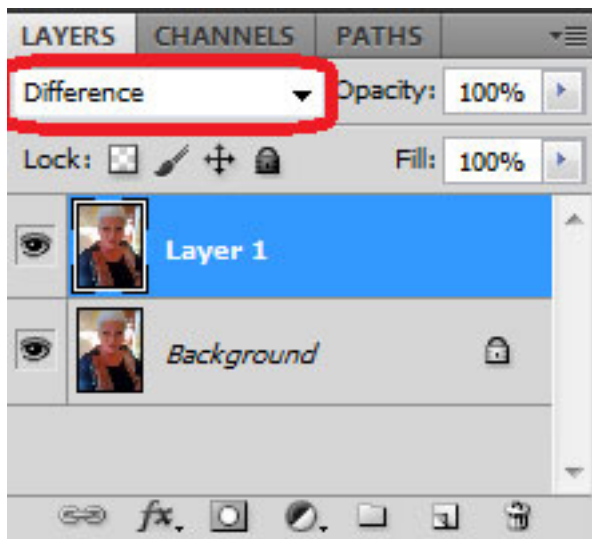
Obriu el Photoshop, llegiu la imatge original i llegiu també la imatge comprimida (us aconsellem que en trieu una de comprimida en un factor de 4 a 6, així els errors seran més evidents).



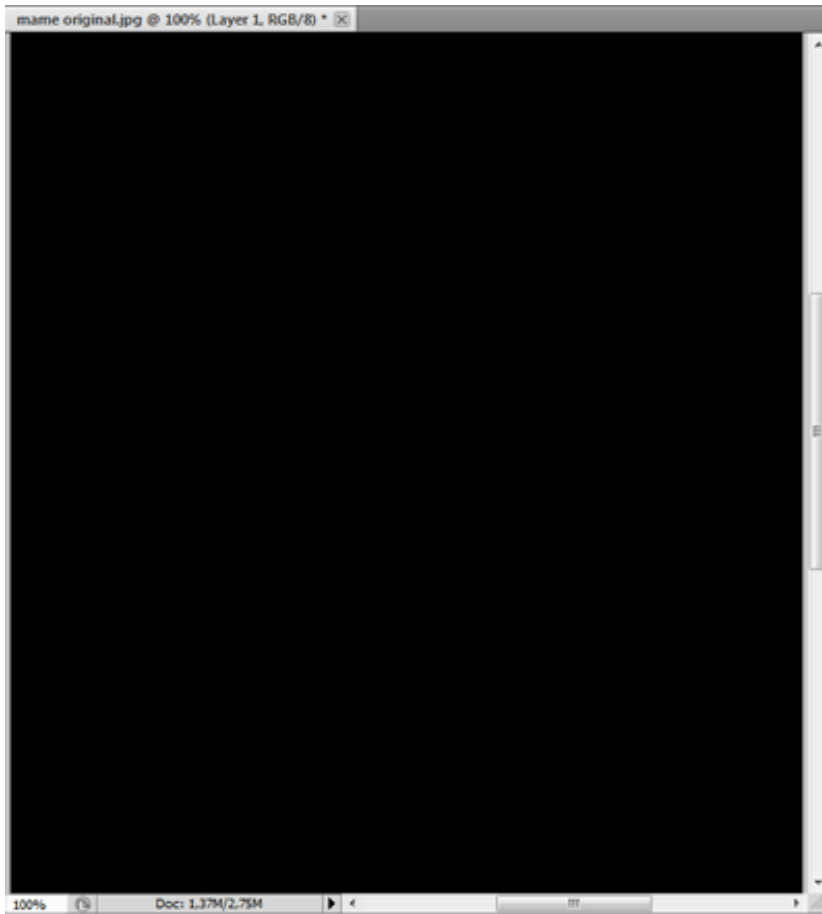
A continuació, seleccioneu amb un clic la imatge original sense comprimir i amb *Ctrl* + *J* (Windows) o *Command* + *J* (Mac) dupliqueu la capa de fons a la paleta de capes. El resultat són dues capes: la capa de fons original a baix i una còpia que el Photoshop denomina *Capa 1*.



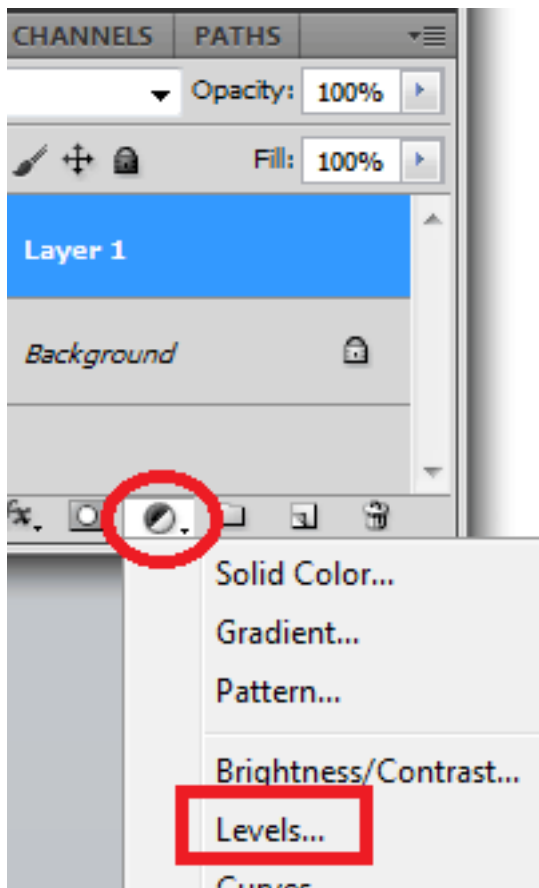
Les dues capes són idèntiques; per a demostrar-ho presentarem la imatge diferència: es pot avançar que, si són iguals, la diferència entre píxels corresponents serà 0, i per tant tots els píxels de la imatge diferència seran de valor 0, és a dir, totalment negres. En el desplegable de la finestra de capes, trieu *Diferència*:



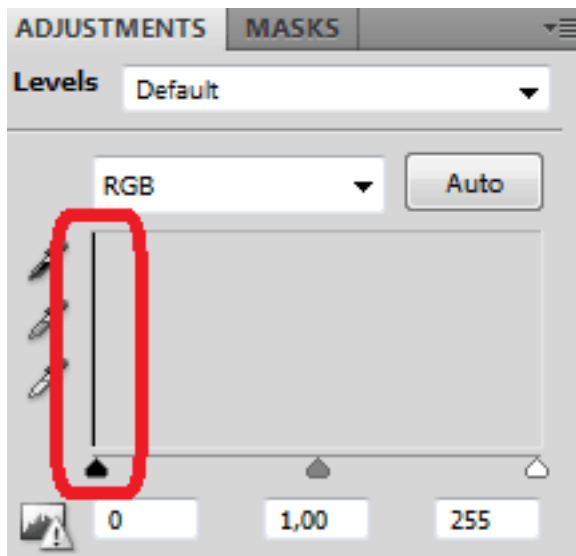
Tal com era previsible, la imatge diferència apareix totalment negra.



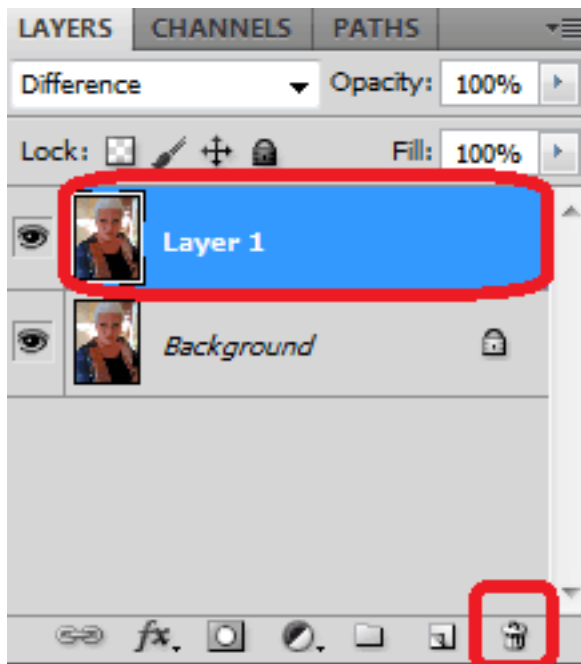
Siguem escèptics i confirmeu que és així d'una manera més explícita; seleccioneu l'eina d'ajust i, de les opcions existents, trieu *Niveles*.



Activarà una finestra de diàleg que reflecteix l'histograma de la imatge diferència. És estret perquè tots els píxels són del mateix valor, no es distribueixen en altres valors que no siguin el zero.



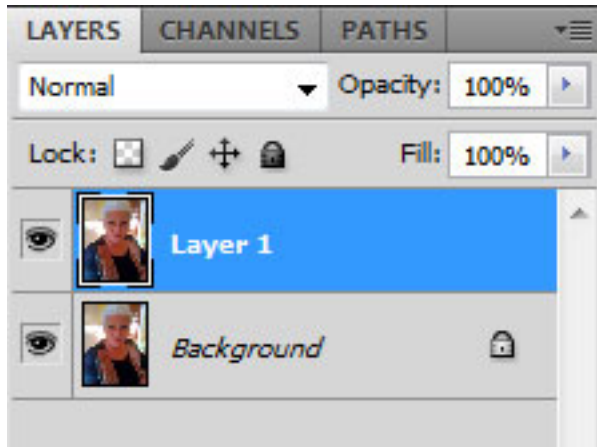
Una vegada demostrada que la imatge diferència entre dues imatges iguals és negra, esborreu la capa 1 seleccionant-la amb un clic a la finestra de capes i prement la paperera situada en aquesta mateixa finestra.



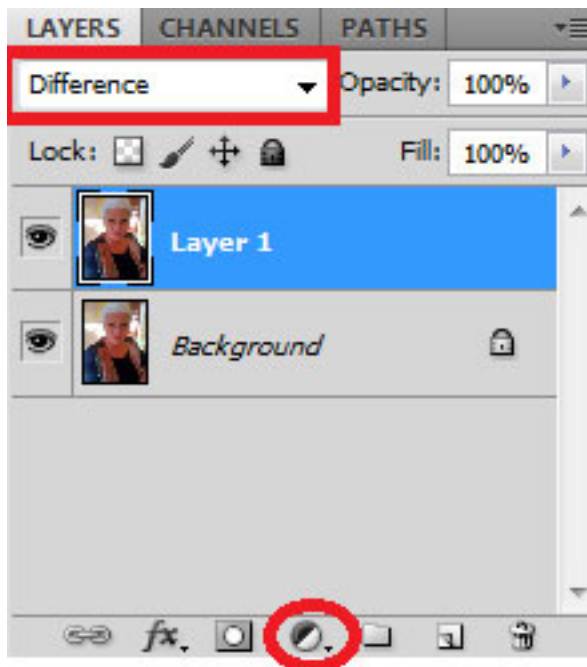
Ara que sabeu fer una imatge diferència i que hem tornat al punt de partida, torneu a fer l'operació però amb les dues imatges que encara teniu obertes al Photoshop. Seleccioneu l'eina *Mover* de la barra d'eines, premeu el ratolí sobre la **imatge comprimida** sense deixar-lo anar, arrossegueu el ratolí fins a la imatge sense comprimir, premeu i mantingueu la tecla de majúscules, i finalment deixeu anar el ratolí.



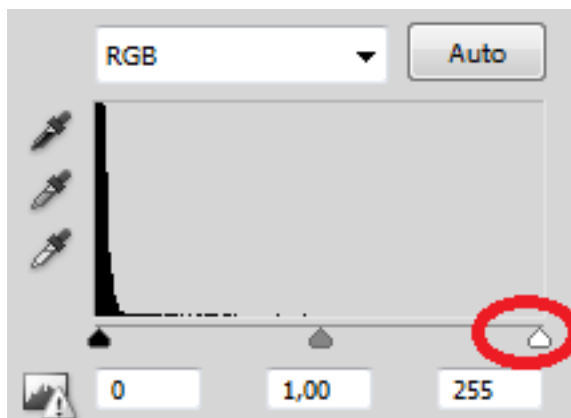
Si bé visualment no sembla que hi hagi hagut cap efecte, a la finestra de capes s'hi haurà creat una segona capa, anomenada *Capa 1*, que és la imatge comprimida, mentre que el fons serà la imatge sense comprimir.



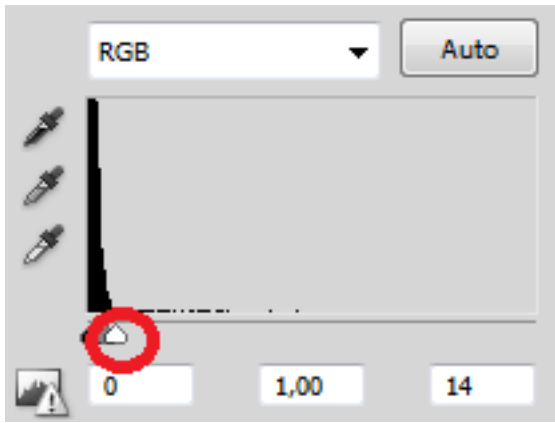
Ja tenim les dues imatges en capes, per la qual cosa podem tornar a fer l'operació *Diferència* (apareixerà aproximadament negra) i a continuació l'eina de capes, triant *Niveles*:



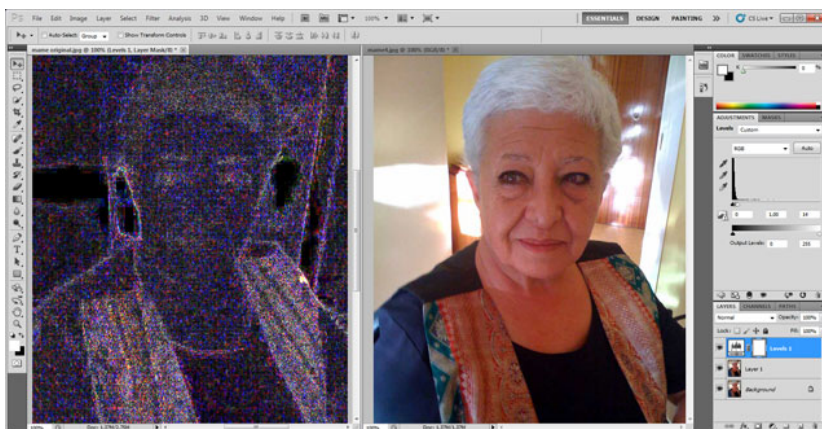
Gairebé segur que l'histograma ja no serà tan uniforme:



Per a poder destriar clarament els píxels diferència, desplaceu el *gadget* blanc cap a l'esquerra i així realçareu visualment els seus valors.



Així, tots els errors generats per la codificació amb pèrdues de JPEG es reflecteixen clarament situats en zones específiques de la imatge, com les vores dels objectes i en general les zones amb molt detall visual.



Fent el procés amb nivells de compressió més elevats apareixen igualment errors en zones de color homogeni i un clar efecte de blocs en general, intrínsecs a la manera en què JPEG treballa la imatge.

4.2.2. Altres modes de treball de JPEG

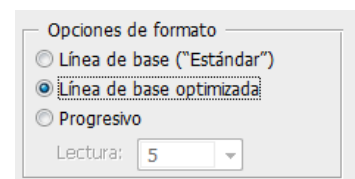
El mode seqüencial base explicat és vàlid per a la majoria d'aplicacions de tractament d'imatge, en què aquestes tenen 8 bits per cadascun dels components (RGB) i fins i tot una quarta capa de transparència.

No obstant això hi ha altres modes (fins a 44) per a aplicacions específiques. Per exemple, per a imatges de qualitat molt alta amb 12 bits per component, o per a aplicar altres sistemes de codificació de coeficients més efectius (codificació aritmètica en comptes de Huffman).

Una versió millorada del mode seqüencial base, el **mode seqüencial base optimitzat**, permet millorar del 2% al 8% la mida del fitxer final. No obstant això, algunes aplicacions poden no ser compatibles amb aquest mode.

Activitat 14

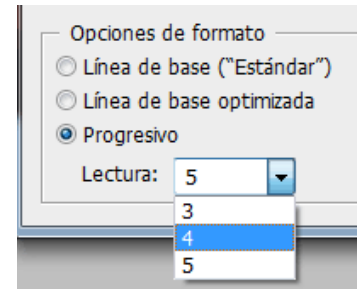
Visualitzeu la imatge diferència en el cas de compressió JPEG en mode *Línea de base optimitzado* i valoreu la diferència amb JPEG estàndard utilitzant la mateixa imatge original



Mode seqüencial base optimitzat

de l'exercici anterior i comprimint-la a la mateixa qualitat. Valoreu la qualitat visual i factor de compressió de la resultant.

El tercer mode interessant, el **mode progressiu**, molt utilitzat per a imatges de llocs web amb servidors amb connexions de baixa velocitat o molt saturats, permet veure una imatge mosaic (és a dir, la imatge amb només el coeficient de contínua de cada bloc) i les seves millores successives de qualitat segons l'ordre de recepció de les dades; així l'espectador pot tenir vistes parcials cada vegada més detallades de la imatge que va rebent. Hi ha diverses maneres d'implementar aquesta solució, encara que la més generalitzada és emmagatzemar en el fitxer JPEG en primer lloc els coeficients de contínua i els primers coeficients d'alterna de tots els blocs, després coeficients d'alterna de menys rellevància visual (més detall), després de rellevància encara més petita (i més detall), etc. El procés usualment fa tres, quatre o cinc fases de separació dels coeficients. Un fitxer JPEG en mode progressiu té normalment una mida en bytes superior que el seu equivalent en mode seqüencial o seqüencial optimitzat.



Mode progressiu

Activitat 15

Feu l'exercici anterior triant en aquest cas el mode progressiu amb 4 lectures. Amb un navegador o programa de visualització, obriu el fitxer d'imatge generat i comproveu que es veu amb una millora gradual.

Finalment, el **mode JPEG-LS** (*lossless JPEG*, JPEG sense pèrdues) és un mode creat uns anys més tard (1993) que codifica la imatge **sense pèrdues**. Utilitza una DCT per a codificar, però no fa servir taules de quantificació, sinó que fa previsions dels píxels a partir dels píxels veïns. Adreçat a aplicacions mèdiques d'imatge, aconsegueix uns factors de compressió de 2:1, i no té rellevància fora d'aquest camp.

4.3. JPEG 2000

Una dècada més tard de l'aparició del format JPEG l'escenari havia canviat sensiblement: s'havien investigat amb èxit transformacions noves que superaven la clàssica DCT, encara que obligaven a fer un càlcul intensiu, però la capacitat de càlcul dels equips domèstics i professionals quintuplicava el de deu anys abans, i feia factible un format nou de codificació que donés resposta a les exigències noves del mercat, entre d'altres:

- Codificació amb pèrdues o sense pèrdues.
- Alt factor de compressió amb menys artefactes visibles.
- Robustesa contra errors per a poder transmetre a baixa velocitat, etc.
- Capacitat d'afegir dades addicionals o metadades.
- Millor codificació d'imatges sintètiques o bitò.

Lamentablement, la situació legal del format JPEG2000 va ser sempre controvertida, ja que hi ha patents sobre parts del procés, la qual cosa n'ha frenat des del principi la popularització.

El format JPEG dóna qualitats subjectives d'imatge molt baixes en compressions més baixes de 0,25 bpp, en què el JPEG2000 encara pot donar qualitat fins a 0,1 bpp. I en els rangs de compressió habituals, aquest sistema de codificació nou supera el JPEG.

El secret és la transformació Wavelet, que **tracta la imatge com un tot** (no la descompon en blocs) i li aplica diversos filtres espacials i aprofita de manera molt eficient la codificació dels resultats. Segons els filtres que s'utilitzin, aquesta transformació pot ser **sense pèrdues** o **amb pèrdues**, i en aquest últim cas millora la compressió a costa d'algun efecte visual en els contorns (l'alta freqüència espacial).

Així, si el JPEG2000 és capaç de comprimir amb qualitat a taxes més altes, és millor per a transmetre imatge a baixa velocitat. Com que a més el **tipus de descodificació és progressiu**, la imatge rebuda es descodifica a partir d'una primera versió bàsica (baixa freqüència espacial) que va millorant detalladament, perfecte per a transmissió: si s'interromp en cert moment, almenys tindrem una imatge completa parcialment detallada.

L'estàndard permet incloure dades addicionals (metadades) en el fitxer d'imatge. El JPEG ja ho permetia en el format EXIF, però el JPEG2000 ho fa de manera més estandarditzada i segueix el format XML.

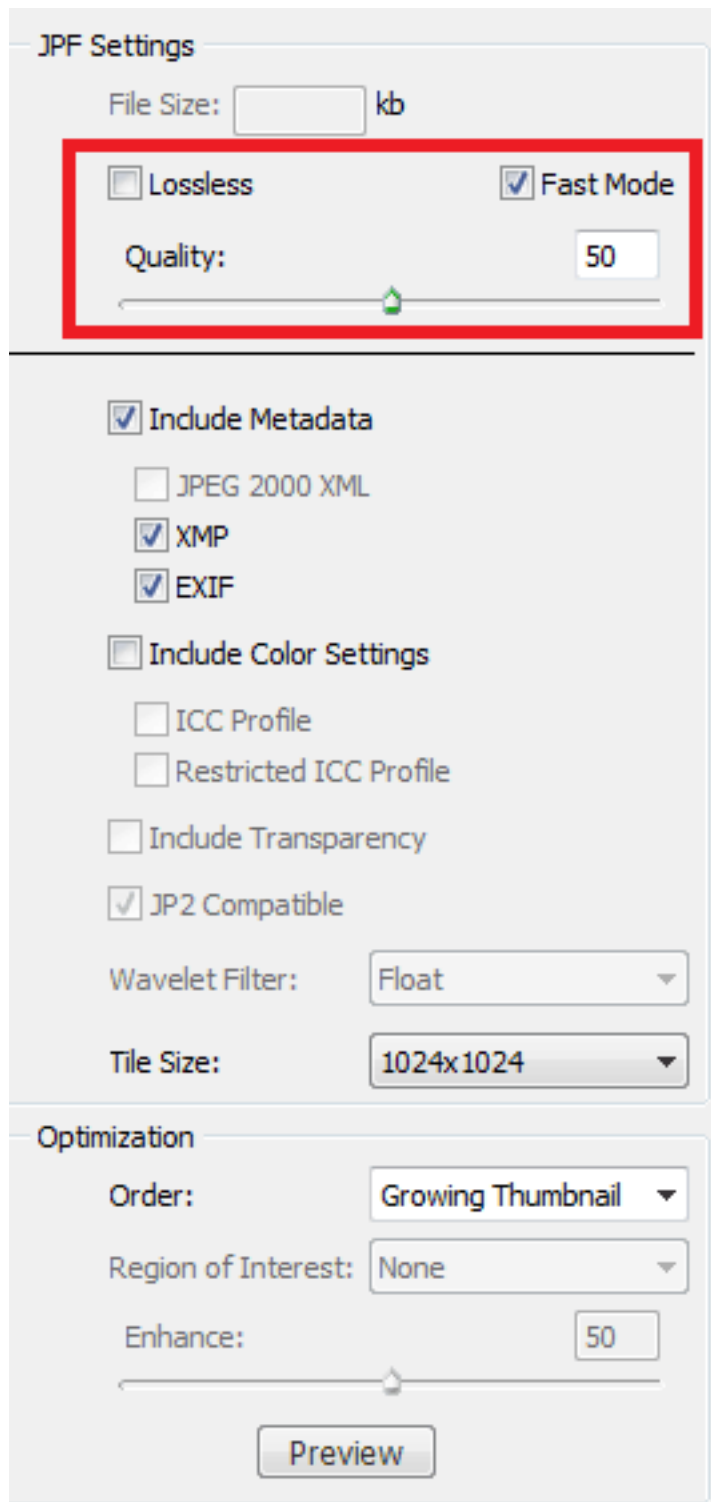
Activitat 16

El Photoshop llegeix i grava en format JPEG2000 (extensions *.jpf, *.jpx, *.jp2, *.j2c i *.jpc) i accepta una configuració personalitzada en cada enregistrament, com, per exemple, si es vol una compressió sense pèrdues en una qualitat indicada per un valor de 0 a 100.

Feu algunes compressions amb aquest format en mode amb pèrdues, proveu el format sense pèrdues i valoreu el nivell de compressió aconseguit pel que fa a la mida del fitxer.

Problemes legals amb JPEG2000

Bona part del programari de tractament d'imatges no l'inclouen o ho fan opcionalment, com l'Adobe Photoshop. En el cas de formats de compressió sense pèrdua, actualment el PNG és el més estès.



The image shows a 'JPF Settings' dialog box with several sections. At the top, there is a 'File Size' field with a text input and 'kb' label. Below this, a red rectangular box highlights the 'Lossless' checkbox (unchecked) and the 'Fast Mode' checkbox (checked). Underneath, a 'Quality' slider is set to 50. The next section contains several checkboxes: 'Include Metadata' (checked), 'JPEG 2000 XML' (unchecked), 'XMP' (checked), 'EXIF' (checked), 'Include Color Settings' (unchecked), 'ICC Profile' (unchecked), 'Restricted ICC Profile' (unchecked), 'Include Transparency' (unchecked), and 'JP2 Compatible' (checked). Below these are two dropdown menus: 'Wavelet Filter' set to 'Float' and 'Tile Size' set to '1024x1024'. The 'Optimization' section at the bottom includes 'Order' set to 'Growing Thumbnail', 'Region of Interest' set to 'None', and an 'Enhance' slider set to 50. A 'Preview' button is located at the bottom center.

Opcions de codificació amb pèrdues o sense del format JPEG2000

4.4. JPEG XR

El 2006 Microsoft va introduir un còdec propietari d'imatge inicialment denominat *HD Photo* adreçat exclusivament al sector de la fotografia digital que des de l'any 2009, quan es va convertir en estàndard ISO, és conegut com a *JPEG XR (extended range)*. L'extensió dels fitxers d'imatge és .wpd.

Les seves característiques principals són les següents:

- Codificació amb pèrdues o sense pèrdues.
- Descodificació progressiva.
- Millor compressió que el JPEG.
- Un ventall de formats ampli: 8, 16 o 32 bits per canal de color, els píxels poden tenir valor enter o fins i tot flotant per a adaptar-se a càmeres fotogràfiques de rang dinàmic molt alt, i un format en blanc i negre de 16 bits.
- El format contenidor pot incloure metadades i un canal alfa.
- No s'adreça a ser utilitzat en imatges sintètiques o amb bitons (text o gràfics).

Si bé aquest format només el suporta Windows des de la versió Vista i hi ha alguns connectors comercials (per exemple, per a l'Adobe Photoshop oferts per Microsoft), avança lentament enfront de l'estàndard obert d'imatge DNG (Adobe, 2004). L'alliberament del JPEG XR el 2010 per part de Microsoft podria donar-li una possibilitat d'èxit en facilitar-ne l'ús en aplicacions de codi obert.

4.5. Tendències futures de la codificació d'imatge

A mitjan anys vuitanta el matemàtic Michael Barnsley va presentar la base d'un sistema de compressió d'imatges amb pèrdues basat en fractals que va generar gran expectació.

Un **fractal** és un element senzill que, iterat sobre si mateix diverses vegades, genera un element complex nou. La idea de Barnsley és generar una imatge original a partir d'un element visual bàsic amb iteracions. Així, només s'hauria d'emmagatzemar o transmetre el fractal o grup de fractals, no pas la imatge. Fins i tot és possible reconstruir la imatge original a resolucions superiors de les de partida amb molt poca capacitat de càlcul.

El problema principal, no obstant això, és com trobar de manera automàtica l'element visual bàsic (o el conjunt d'elements bàsics) que pogués aconseguir-ho. La necessitat d'una gran capacitat de càlcul per a aconseguir deduir-los i la falta d'un algorisme eficient per a tot tipus d'imatges ha frenat fins avui aquest sistema de compressió, encara que la investigació segueix oberta fins i tot per a aplicar-lo en vídeo.

Resum

S'han descrit els diferents mecanismes de transport d'imatge i so per a la seva captura, i s'han diferenciat els analògics dels digitals i, entre aquests, els d'entorn purament informàtic, valorant la qualitat que permet aconseguir cadascun. En aquesta visió s'han introduït els conceptes de luminància i crominància, base dels sistemes de compressió d'imatge i vídeo actuals.

Hem establert un primer contacte amb la problemàtica de la transmissió de vídeo i àudio digital per Internet, hem valorat els problemes de retard i pèrdues que hi pot haver i hem demostrat l'existència de compressió en tot sistema d'emmagatzematge en arxius multimèdia.

A continuació hem descrit el sistema de codificació d'imatge estàtica JPEG, les seves fases de treball, els paràmetres que la configuren i els modes que hi ha. Aquest cas ens ha permès diferenciar els conceptes de codificació i compressió, i ens ha il·lustrat algunes maneres de valorar tant la compressió aconseguida com els errors introduïts en el cas de compressió amb pèrdues. Finalment, hem fet un cop d'ull superficial als estàndards JPEG2000 i JPEG XR, i els hem comparat amb l'estàndard original.

