Característiques tècniques del vídeo digital

PID_00236708

Aniol Marín Atarés

Temps mínim de dedicació recomanat: 2 hores

Universitat Oberta de Catalunya

Índex

1. La naturalesa tècnica del vídeo

En general, quan estudiem el vídeo, ens centrem en els aspectes més aplicats: el llenguatge, la fotografia, el so, els procediments... tots ells, evidentment, necessaris per a crear bons clips. En aquest mòdul en canvi analitzarem la base tecnològica sobre la qual es poden desenvolupar tots aquests altres conceptes. Potser, en el nostre cas, no és estrictament necessari conèixer la tècnica per crear els nostres clips, però ens pot ajudar a entendre molt millor com treballar les imatges de les quals disposem. En especial, pot ser molt interessant per a entendre la gran diferència entre fitxers de gran mida i fitxers de gran qualitat.

Per començar, cal entendre què és comú en tots els vídeos. Algunes variants, com ara els videojocs, tenen característiques molt particulars, com ara la generació en temps real de la imatge com a resposta a la interacció de l'usuari, però en el fons es basen en la mateixa tecnologia. Així, doncs, per començar, buscarem un marc comú que després puguem adaptar a totes les altres modalitats de vídeo.

Les característiques tècniques que comparteixen gairebé tots els clips ens permeten dir que el vídeo:

- És una **representació en dues dimensions** d'una realitat que en origen sol ser tridimensional.

- Ens arriba a través d'una pantalla rectangular que **té una relació d'aspecte** concreta (relació entre l'alçada i l'amplada) que pot variar de vídeo a vídeo però **que es manté fixa en cada clip**.

- **Es compon de fotogrames estàtics** que se succeeixen ràpidament i creen una il·lusió de moviment.

- Un cop generat, **consisteix en un flux de dades codificat digitalment**, que diversos aparells s'encarreguen de convertir de nou en imatges i so.

Evidentment, com ja sabem, hi ha excepcions a la majoria d'aquestes característiques, algunes d'elles prou conegudes. En mencionarem unes quantes a tall d'exemple.

1.1. Vídeo en 3D

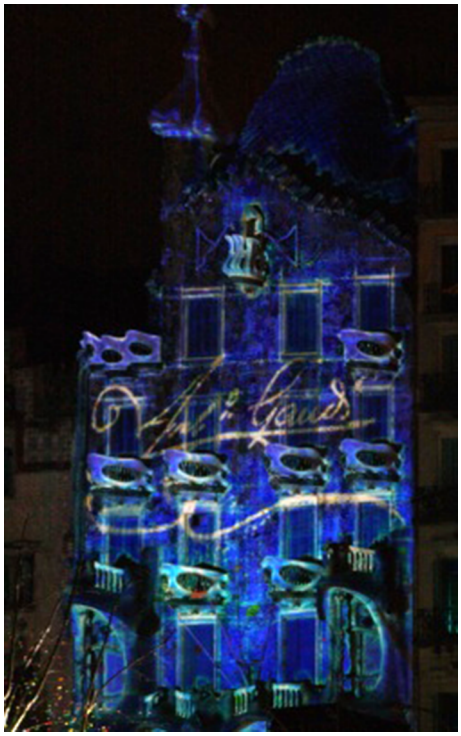
Tot i que les experiències de vídeo estereoscòpic són gairebé tan antigues com el cinema mateix, el vídeo 3D va experimentar un creixement molt important l'any 2009, moment en el qual molts cinemes generalistes van incorporar projectors 3D per a l'estrena d'*Avatar* de David Cameron. Actualment el vídeo en 3D és en ple creixement i s'està introduint amb força en el sector domèstic.

Càmera estereoscòpica. Font: JVCAmerica - Treball propi, CC BY 2.0, https://www.flickr.com/photos/jvcamerica/5320816138

Els mètodes per a aconseguir-ho difereixen entre si, però tots consisteixen a fer arribar a cadascun dels ulls humans una imatge plana lleugerament diferent, que el cervell interpreta com a tridimensional. Tot i que aquest efecte es pot aconseguir en les fases de postproducció o generant directament les imatges per ordinador, les autèntiques gravacions en 3D s'aconsegueixen gravant amb una càmera estereoscòpica.

1.2. *Video mapping*

Les instal·lacions de *video mapping* consisteixen en projeccions que es fan sobre volums en comptes de fer-ho sobre pantalles planes. Tot i la complexitat que pot suposar preparar la imatge perquè s'adapti al volum en particular, no deixen de ser vídeos ordinaris que es presenten en un o diversos projectors. S'acostuma a fer sobre façanes d'edificis, però es poden fer virtualment a tot arreu. Un bon exemple és la reconstrucció visual de l'interior de l'església de Sant Climent de Taüll, en la qual s'utilitzen sis projectors per projectar sobre les parets deteriorades les imatges originals.



Exemple de *videomapping* sobre una façana. Font: amadalvarez - Treball propi, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=22314721

1.3. Cinema tradicional i vídeo analògic

Durant dècades, tant el cinema com el vídeo van ser analògics. El cinema clàssic utilitzava una pel·lícula fotosensible, que s'havia d'exposar i revelar abans de poder-se editar i projectar. El vídeo analògic, d'altra banda, utilitzava principalment cintes magnètiques que codificaven els fotogrames en línies i feia servir, per tant, un sistema molt més semblant al del vídeo actual. En la resta, però, es diferencien poc del vídeo digital.

La pel·lícula tradicional de cinema, poc utilitzada actualment, és un exemple analògic. Font: Runner1616 - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=19318310

Actualment encara es poden trobar algunes excepcions a aquesta tendència del cinema digital, com la pel·lícula *The Hateful Eight* de Quentin Tarantino, presentada l'any 2016 i rodada exclusivament amb film analògic. De totes maneres, aquests mètodes tenen una difusió molt reduïda. La gran majoria dels cinemes actuals, per exemple, utilitzen només projectors digitals. D'altra

banda, els formats de vídeo analògic, com ara el VHS, el Laserdisc o el Hi8, van deixar de tenir importància amb l'arribada al mercat dels primers sistemes DV i DVD, a finals de la dècada de 1990.

1.4. Vídeo en 360º i realitat augmentada

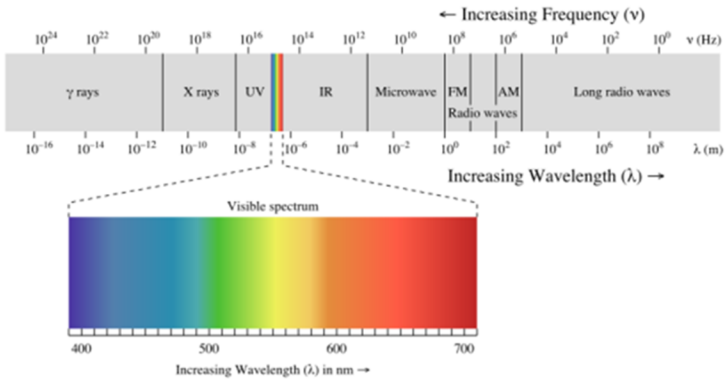
Tot i que ja existia anteriorment, durant la dècada de 2010 s'experimenta una gran crescuda de les experiències immersives de realitat virtual i de realitat augmentada, especialment a partir del llançament de *Google Cardboard* l'any 2014. Aquestes experiències, entre d'altres, inclouen el vídeo en 360º (o vídeo immersiu), que permet a l'espectador decidir cap a quina part de l'escena vol mirar. La diferència més important d'aquesta forma de vídeo és que, tot i que tècnicament es continua mostrant en una pantalla, l'experiència no està limitada per la «finestra» inamovible que suposa la pantalla clàssica. Quan utilitzen metratge real aquests vídeos s'aconsegueixen amb omnicàmeres o bé amb una matriu de càmeres col·locades en cercle. Tot i que l'efecte aconseguit és espectacular, i que és difícil aplicar-hi la majoria de convencions audiovisuals, continua sent una forma de vídeo.

Model simple de Google Cardboard
Font: Evan-Amos - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=45580283

Cal observar que tots els casos de vídeo, fins i tot els de realitat augmentada, comparteixen un aspecte: es componen de fotogrames estàtics. Així doncs, qualsevol vídeo crea la *il·lusió* de moviment, però sempre a partir de l'estaticisme. És la nostra percepció la que l'interpreta com un element dinàmic. Aquesta és una qüestió important, que seguidament analitzarem a fons.

2. Fonaments de la visió

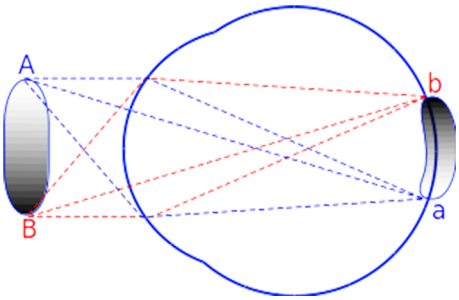
El vídeo és possible gràcies a les característiques de la visió humana. Aquesta, com ja sabem, ens serveix per a captar la llum que ens arriba del nostre entorn. Aquesta llum és energia en forma d'ona o vibració que viatja a una gran velocitat. En concret, és un tipus de radiació electromagnètica, com ara els rajos X, les microones o les ones de ràdio, i només es diferencia d'aquestes altres per la freqüència de la vibració. És emesa per alguns cossos, com ara el sol o una bombeta, i absorbida o reflectida en diferent mesura per altres cossos (com per exemple per un mirall o un tros de fusta). Quan diem que un objecte és, per exemple, vermell, és perquè absorbeix totes les freqüències visibles menys les que percebem com a vermelles, que són per tant les que ens arriben a nosaltres. Els objectes que percebem com a negres reflecteixen molt poc totes les ones, mentre que els blancs les reflecteixen gairebé totes. Val a dir que aquest és el motiu pel qual la tonalitat exacta del blanc depèn del conjunt concret de freqüències d'ona que estigui il·luminant l'objecte, i com que aquest depèn de la font lumínica sovint ens cal fer balanç de blancs.



En color, espectre visible dins de les freqüències de la radiació electromagnètica
Font: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2521356

2.1. Òptica i fisiologia de l'ull humà

La llum viatja a una velocitat constant en línia recta. A efectes pràctics, podem considerar que només és desviada quan canvia de medi (segons el tipus de material i l'angle d'incidència sobre aquest). Aquest canvi de medi és el que aprofiten tant les òptiques de la càmera com els nostres ulls per a enfocar la imatge. En concret, l'ull humà aprofita la petita obertura de l'iris per fer passar un sol feix de llum, i la còrnia i el cristal·lí per fer que aquest reprodueixi una còpia en miniatura i invertida de la imatge exterior sobre la retina.



Esquema simplificat del funcionament òptic de l'ull humà
Font: Javalenok - By Inkskape, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=25728443

A cada punt de la retina, doncs, hi arriben concentrats només els rajos de llum d'un punt concret de la imatge, i que per tant són de certes freqüències (que percebrem com a un cert color). Hi ha dos tipus de cèl·lules encarregades de captar aquests rajos:

- Els **bastons** estan especialitzats a captar-ne la quantitat. No són sensibles a cap color en particular sinó a tots més o menys per igual. Són més sensibles, per la qual cosa en condicions de poca llum ens semblarà que les imatges tenen colors més atenuats.

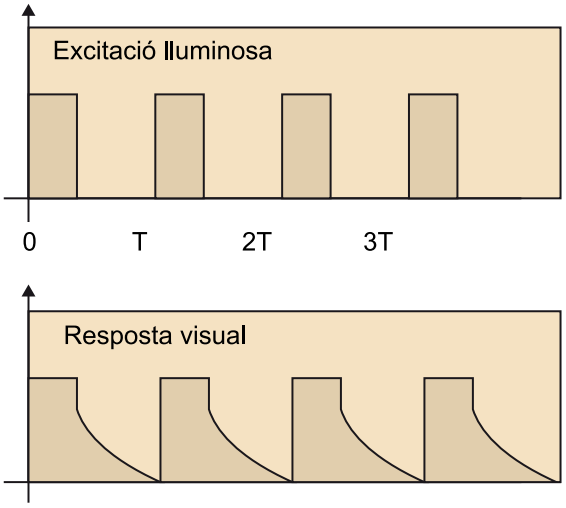
- Els **cons** són menys sensibles, però estan especialitzats a captar certes freqüències i, per tant, ens donen informació sobre el color. En concret, tenim tres tipus de cons, que reaccionen al màxim amb les freqüències que percebem respectivament com a vermell, verd i blau, i menys amb les freqüències més distants a la seva.

Així doncs, captem tota la informació de la llum reduint-la només a quatre paràmetres: un d'intensitat (equiparable a la luminància) i tres de color (equiparable a la crominància). Quan percebem el color blanc és perquè ens arriben diverses freqüències que activen tant els bastons com tots els cons. Quan veiem quelcom de color groc és perquè estimula els bastons, els cons sensibles al vermell i els sensibles al verd, però no estimula els cons sensibles al blau. Això serà important per a entendre com funciona el color en el vídeo.

2.2. Persistència retinal

A diferència del que passa amb una càmera, on podem controlar la velocitat d'exposició, l'ull humà rep llum constantment i no està preparat per a discriminar exposicions molt curtes. Els cons i bastons, un cop estimulats, tarden unes fraccions de segon a tornar al seu estat de repòs. Això fa que, mentre la intensitat total sigui la mateixa, una ràfega d'estímuls sigui indistingible d'un estímul continu. Les antigues bombetes incandescents, per exemple, emeten un flux continu, mentre que les de baix consum i les LED emeten llum de forma discontínua, intermitentment. Aquest fet, però, passa desapercebut i totes

sembla que il·luminin contínuament. D'aquest fenomen se'n diu persistència retinal, i com es pot intuir és clau per a entendre el vídeo tal i com el coneixem. Ens falta, però, descriure un tercer fenomen.



A la part superior, intensitat de l'estímul. A la part inferior, la intensitat de la resposta a aquest estímul.

2.3. Percepció del moviment

Encara que els objectes es moguin, nosaltres l'únic que percebem és que la seva llum ens estimula una zona diferent de la retina. El cervell és el que s'encarrega de reconèixer que les informacions que rep de diversos cons i bastons corresponen al mateix objecte. Hi ha una malaltia anomenada acinetòpsia que consisteix precisament a no ser capaç de processar aquesta informació; les persones que la pateixen veuen el món com si fos una successió de fotografies. En la resta, aquest procés és automàtic i ens pot portar a veure moviment fins i tot allà on no n'hi ha, com per exemple en les «ones d'estadi» que es generen a les grades dels estadis esportius, que tot i que semblen avançar es deuen només al moviment vertical sincronitzat dels espectadors.

Fotografia d'una ona d'estadi
Font: Eva Rinaldi - Big Day Out, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=24788458

La naturalesa de la llum, la persistència retinal i la percepció de moviment són els tres factors principals que permeten que la tecnologia del vídeo funcioni tal com l'entenem.

3. El senyal de vídeo digital

Després d'haver entès com funciona la visió, veurem com funciona en canvi el vídeo, que com ja hem anticipat aprofita les característiques de la visió humana que hem vist abans.

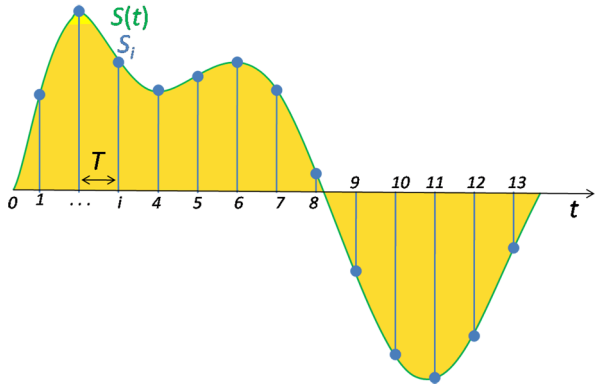
Descomposició d'una imatge en un senyal binari
Font: Jose Leal AV - Treball propi, domini públic, https://commons.wikimedia.org/w/index.php?curid=3199633

En primer lloc, cal entendre que el vídeo en essència no és més que un flux digital de dades. Aquest flux consisteix en una línia de zeros i uns (unidimensional) que codifiquen una imatge que, en origen, era analògica, multidimensional i contínua. Per analitzar aquest flux és més fàcil pensar en l'exemple del so, ja que en poder-se reduir en una sola dimensió (pressió) és una mica més fàcil d'imaginar.

Per a passar una informació analògica, com el so o la imatge, a una informació digital, cal que la reduïm a dígits (zeros i uns); això ho fem a través de dos processos: el mostreig i la quantificació.

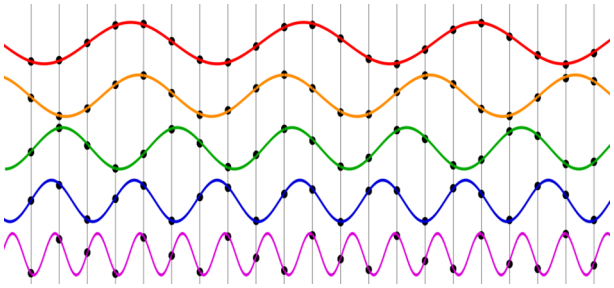
3.1. Mostreig

El mostreig consisteix a mesurar una variable que canvia de valor contínuament i assumir que, entre mostra A i mostra B, els valors s'hauran mogut entre a i b. Com veiem en el cas del so, les freqüències més baixes (amb ones més grans) es representen sense problemes, mentre que les ones més petites només es capten bé amb mostrejos més freqüents. Com més freqüent sigui el mostreig, doncs, més acurat serà el flux resultant. Al mateix temps, un mostreig més freqüent comporta més quantitat de dades.



En groc, una funció matemàtica. En blau, el seu mostreig.
Font: Email4mobile (talk) - http://en.wikipedia.org/wiki/Arxiu:Signal_Sampling.png. Domini públic, https://commons.wikimedia.org/w/index.php?curid=8693098

El cas del so és fàcil d'entendre perquè és unidimensional, i per tant només ens cal prendre una mostra a intervals regulars. En el cas de les imatges, en canvi, ens cal prendre diverses mostres simultàniament en dues dimensions: alçada i amplada. Cadascuna d'aquestes mostres és un **píxel**. Un píxel és, doncs, la mostra d'un punt de la imatge en un moment determinat. Tots els píxels que es prenen en un moment determinat formen un **fotograma**, que és la mínima unitat temporal del vídeo.

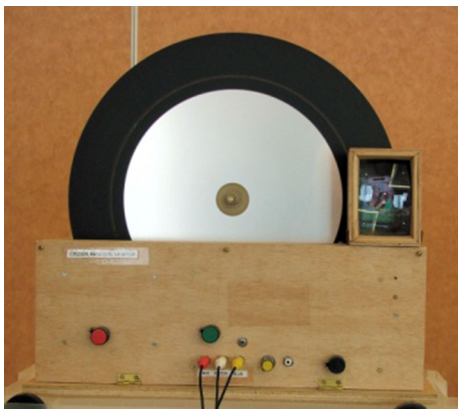


Diferents freqüències i, en negre, els seus mostrejos. Es pot observar la poca precisió dels mostrejos de les ones blava i violeta.
Font: No machine-readable author provided. LucasVB assumed (based on copyright claims). - No machine-readable source provided. Own work assumed (based on copyright claims). Domini públic, https://commons.wikimedia.org/w/index.php?curid=1536518

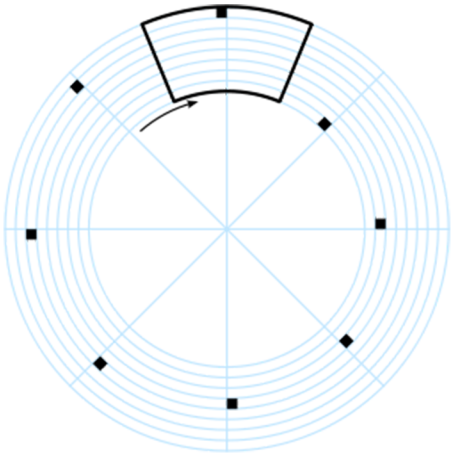
Com que les imatges tenen dues dimensions espacials, per linealitzar tots els píxels d'un fotograma, es treballa amb línies. Per exemple, en un vídeo de 640 x 480, hi ha 480 línies horitzontals de 640 píxels cadascuna. El flux de vídeo descriu una línia sencera píxel a píxel abans de passar a la següent.

Presentació progressiva d'un fotograma interlineat. En la segona meitat de les línies, captades moments més tard, el cotxe ja s'havia desplaçat.
Font: Mikus (talk) - Treball propi (text original: I created this work entirely by myself). Domini públic, https://commons.wikimedia.org/w/index.php?curid=12136645

Fins fa relativament poc han coexistit dos sistemes d'emmagatzematge de les línies: l'escaneig interlineat (p. ex. 720i) i l'escaneig progressiu (p. ex. 720p). L'interlineat s'utilitzava sobretot per a la televisió analògica, i en cada fotograma descrivia primer totes les línies senars i després totes les línies parells, que per tant es gravaven i projectaven en moments diferents. L'escaneig progressiu és l'únic que s'utilitza àmpliament en l'actualitat, i descriu les línies en ordre, una a una. Tot i que cada cop és més poc freqüent trobar vídeos interlineats, cal tenir en compte aquesta diferència perquè els programes d'edició encara permeten exportar en interlineat; en cas de confondre'ls els contorns dels objectes en moviment es veuran desdibuixats, en presentar les línies en un moment inadequat.



Model casolà de televisió mecànica. Es pot observar la imatge a la petita pantalla de la dreta.
Font: User:G1MFG - Own work. Domini públic. https://commons.wikimedia.org/w/index.php?curid=18009704

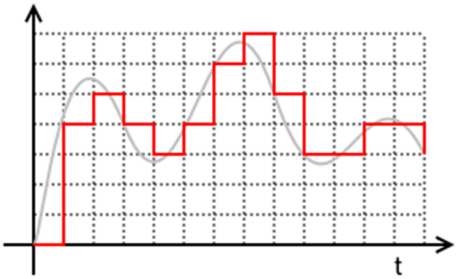


Disc de Nipkow
Font: Hzeller, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3583147

El mostreig ja és suficient per ell mateix per a crear vídeo. A principis del segle XX, per exemple, es va començar a experimentar amb la televisió mecànica, que es pot entendre perfectament només amb aquest. El receptor consistia en una sola bombeta i un disc de Nipkow que girava al davant. El disc de Nipkow té forats a alçades diferents i es col·loca de manera que quan fa un gir complet, cada forat dibuixa una línia de llum a la pantalla. La bombeta rebia el senyal analògic que li arribava d'una càmera, que també funcionava amb un sol disc de Nipkow sincronitzat i una cèl·lula fotosensible. Així, la bombeta s'encenia només quan la cèl·lula rebia llum i mostrava a la pantalla una imatge de vídeo (amb poca definició, però vídeo al cap i a la fi).

3.2. Quantificació

D'altra banda, quan prenem una mostra digital, li hem d'assignar un valor concret. Aquest valor té una precisió limitada al nombre de bits, i per tant els valors reals s'aproximen per excés o per defecte. En el cas del vídeo en blanc i negre, el valor determina la lluminositat del punt entre un mínim (negre) i un màxim (blanc), passant per una gamma de grisos més o menys àmplia segons el cas.



En gris, la senyal analògica. En vermell, la seva quantificació digital.
Font: Domini públic. https://commons.wikimedia.org/w/index.php?curid=384721

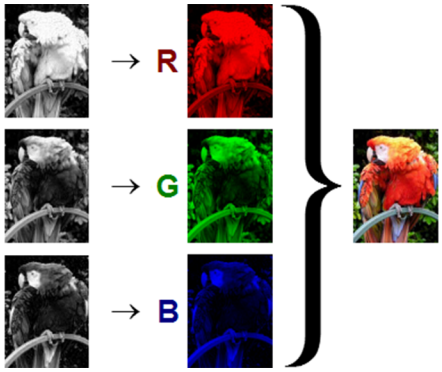
Com que la informació s'emmagatzema en sistema binari (zeros i uns) la precisió que tinguem dependrà del nombre de dígits que utilitzem per descriure-la (altrament anomenats bits). El mínim, evident, és un sol bit per píxel. En aquest cas cada píxel seria o blanc o negre, sense altres valors intermedis. La precisió augmentarà exponencialment segons el nombre de bits (n) que dediquem a cada mostra: per n = 2bits tenim 2^n = 4 possibilitats (00, 01, 10, 11), per n = 3bit tindríem n^3 = 8 possibilitats (000, 001, 010, 011, 100, 101, 110, 111) etc.

A l'hora de quantificar intensitats no hi ha problema: els valors s'ordenen de menys intens (negre) a més intens (blanc) i generen una imatge en escala de grisos. El problema ve a l'hora de representar el color, que com hem dit depèn de les longituds d'ona de la llum que rebem i que acostumen a ser una barreja de freqüències diverses.

Retrat de Charles Chaplin en diverses precisions d'escala de grisos
Font: Strauss-Peyton Studio - National Portrait Gallery. Domini públic. https://commons.wikimedia.org/w/index.php?curid=24576869

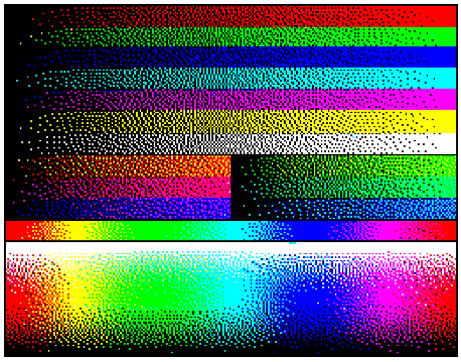
Hi ha diverses maneres de resoldre-ho; la més simple d'entendre és el model **RGB** (acrònim de l'anglès, *Red Green Blue*). Aquest sistema descompon el color real de la imatge en la suma de només tres colors: vermell, verd i blau, que

com ja hem vist corresponen als colors més eficients a l'hora d'estimular els nostres cons i bastons. Si els estimulem tots tres alhora, aconseguirem percebre el color com a blanc.



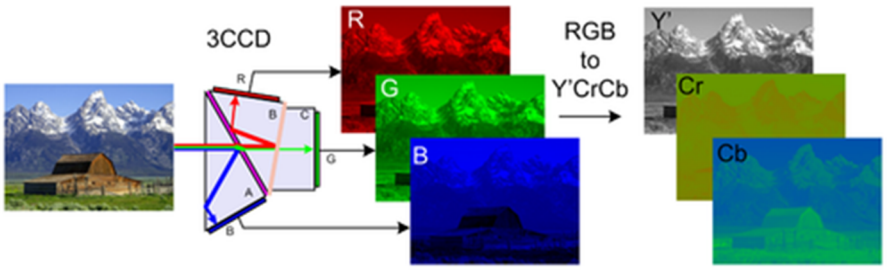
Tres imatges en blanc i negre que representen cada canal RGB. En sumar-les s'obté la imatge completa.
Font: derivat de Ricardo Cancho Niemietz (talk) - self-made adaptation of a common Wikipedia image (Image:Parrot.red.macaw.1.arp.750pix.jpg). Domini públic, https://commons.wikimedia.org/w/index.php?curid=4435272

El més interessant d'aquesta descomposició és que cada canal de color determina només la intensitat d'ell mateix, de manera que és com si es tractés de tres imatges en blanc i negre, una associada amb cada canal. Normalment s'assignen al color verd més bits que als altres dos colors, ja que el percebem amb més precisió. Així, doncs, utilitzant només 3 bits per cada píxel ja podríem representar 8 colors en el nostre flux de vídeo.



Paleta de només 8 colors (amb tramat)
Font: RGB_24bits_palette_color_test_chart.png: Ricardo Cancho Niemietz (talk)The original uploader was Ricardo Cancho Niemietz at English Wikipediaderivative work: DMahalko (talk), Dale Mahalko, Gilman, WI, USA -- Email: dmahalko@gmail.com - RGB_24bits_palette_color_test_chart.png. Domini públic, https://commons.wikimedia.org/w/index.php?curid=11802688

Normalment, però, el vídeo no es codifica directament en sistema RGB. El sistema que normalment s'utilitza es coneix com a Y'UB (tot i que hi ha variacions amb nom semblant, com per exemple Y'CbCr). Aquest sistema ve de l'època en què van coexistir la televisió en blanc i negre i la televisió en color; hauria sigut impossible construir una imatge en blanc i negre a partir dels tres canals RGB, per tant la imatge es descomponia en tres canals: un d'intensitat (Y', luma), que és suficient per a generar una imatge en blanc i negre, i dos de color (U i B, crominància) que no són suficients per a formar una imatge reconeixible per ella mateixa. Els tres canals de color RGB es dedueixen a partir de les diferències entre els valors de la luminància i els de la crominància. És un sistema menys intuïtiu però molt més eficient pel que fa al flux de dades.



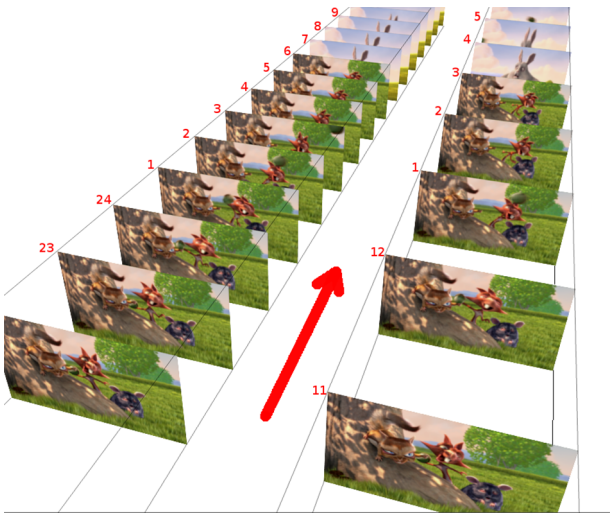
Procés de transformació de la imatge, primer a RGB i després a Y'CrCb
Font: LionDoc - Own work. Domini públic. https://commons.wikimedia.org/w/index.php?curid=19224869

La quantificació del color s'anomena profunditat de color o *color depth*. Actualment en vídeo s'acostuma a treballar amb profunditats de color d'entre 16 bit (65.536 colors) i 48 bit ($2{,}81 \cdot 10^{14}$ colors). Les profunditats de més de 10 milions de colors (que corresponen aproximadament a 24 bit) no són perceptibles a l'ull humà, però ens serveixen per a fer correccions de color més precises en la postproducció.

3.3. Fotogrames, definició i ràtio d'aspecte

Fins ara hem entès com s'estructura linealment el senyal de vídeo. Però com bé sabem, el vídeo el percebem almenys en dues dimensions espacials (alt i ample) i una de temporal. Com ja hem comentat, el flux de vídeo forma imatges gràcies als píxels. Cadascuna d'aquestes imatges s'anomena fotograma.

El **fotograma** és la unitat temporal bàsica de qualsevol vídeo. Consisteix essencialment en una fotografia que es mostra en pantalla només per uns instants. La successió de fotogrames és la que, gràcies a la persistència retinal i a la percepció de moviment, ens fa que veiem el vídeo com una imatge en moviment. Perquè això sigui d'aquesta manera, però, cal que els fotogrames se succeeixin a una certa velocitat mínima.



Fotogrames ordenats en el temps. El flux de l'esquerra conté 24fps, mentre que el de la dreta conté només 12fps.

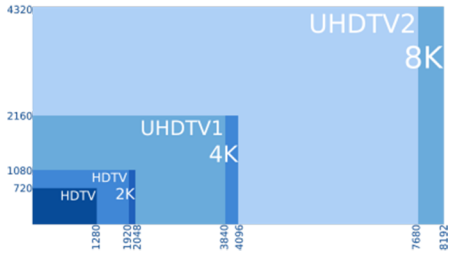
La velocitat dels fotogrames es mesura en **fotogrames per segon** (**fps**, de l'anglès *frames per second*). Els fps mínims per a percebre un vídeo com a continu depenen de diversos factors, però se situen més o menys a 12 fps. Velocitats de fps superiors fan que la imatge sigui més nítida, però també augmenten la quantitat d'informació que cal processar.

Per diversos motius, les velocitats més habituals de reproducció són **24 fps** per al cinema, **25 fps** a Europa i altres països on s'utilitzaven sistemes PAL/SECAM, i **30 fps** per als EUA i altres països on s'utilitzaven sistemes NTSC. També comença a ser freqüent l'ús de velocitats més altes, normalment en múltiples de la velocitat base (48 fps, 50 fps i 60 fps respectivament), tot i que els 90 fps o més són bastant habituals en videojocs i experiències de realitat virtual.

La resolució del vídeo normalment es mesura comptant els píxels verticals. És una herència del vídeo analògic, ja que aquest s'estructurava en línies horitzontals, com les línies d'un llibre, i per tant només tenia sentit comptar la resolució verticalment. Algunes de les resolucions verticals més habituals són les següents:

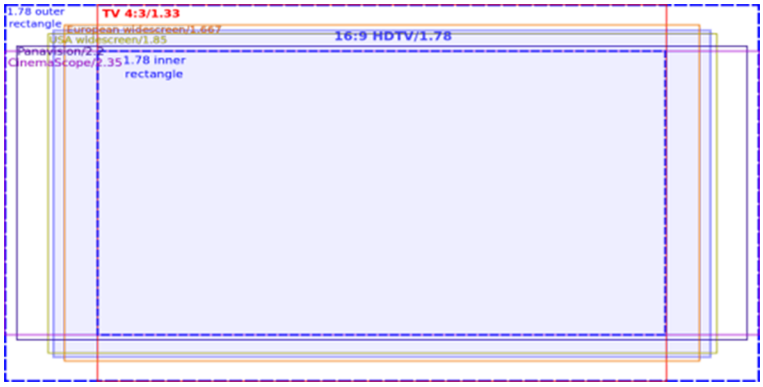
Sistema	Definició vertical (px)	Altres usos (o noms)
NTSC	480	DVD
PAL	576	DVD
HDTV	720	(HD ready)
	1.080	(Full HD, 2K*) Blu-ray
UHDTV	2.160	(4K *)
	4.320	(8K *)

(*) Els noms 2K, 4K i 8K provenen de la nomenclatura del cinema, que es basa en la definició horitzontal, i per tant poden generar confusió. Per exemple, tot i que la definició vertical de l'UHDTV és de només 2.160 píxels, se l'anomena 4K perquè la seva resolució horitzontal és aproximadament aquesta (típicament el fotograma és de 3.840 x 2.160 px), mentre que al seu torn el 8K té només 4.320 píxels de definició vertical i el 2K només 1.080.



Comparació de la mida del fotograma de les resolucions més habituals
Font: Libron - Arxiu: 8K UHD, 4K SHD, FHD i SD.svg, CC0, https://commons.wikimedia.org/w/index.php?curid=42210549

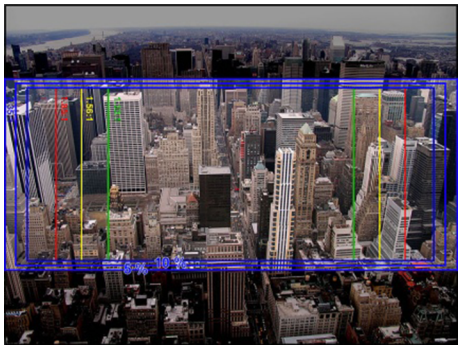
El vídeo, evidentment, també té una definició horitzontal, però és una mica més complexa. En general es pot deduir de la **ràtio d'aspecte**, que és el que determina les proporcions de la imatge. Per exemple una ràtio 1:1 indica que la imatge és quadrada. Una imatge 720p amb ràtio 1:1, doncs, tindria una resolució de 720 x 720 píxels. De totes maneres cal tenir en compte que no sempre s'utilitzen píxels quadrats; tot i ser cada vegada més rars, a vegades s'utilitzen píxels rectangulars, i per tant la resolució horitzontal no sempre es pot calcular directament amb la ràtio d'aspecte.



Comparació de diverses ràtios habituals
Font: MarkWarren - Treball propi, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=8255615

Les ràtios d'aspecte són molt diverses, i per tant cal tenir-les molt en compte. Es defineixen sempre en una equació matemàtica, dividint l'amplada per l'alçada. Això dóna nom a dues nomenclatures, que tot i que sovint generen confusió representen exactament el mateix:

Ràtio entera	Ràtio unitària	Exemples
1:1	1:1	Plataformes de *videosharing* per *smartphone* (Periscope, Facebook, Instragram)
4:3	1.33:1	Televisió analògica
16:10	1.6:1	Monitors d'ordinador, dècada de 2000
16:9	1.77:1	Monitors d'ordinador, a partir de l'any 2008 HDTV
64:27	2.370:1	CinemaScope



Diversos retalls possibles d'una imatge 1:1. La zona central, en verd, es consideraria zona segura per a la majoria de ràtios.

Font: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=69807

La ràtio d'aspecte és quelcom que cal tenir molt en compte a l'hora de produir vídeo. Tot i que es pot retallar, a l'hora de gravar ja determina la composició fotogràfica. Durant la producció pot crear conflicte si s'ha de muntar material d'origen amb ràtios diferents. Per últim, cal tenir en compte en quines pantalles s'exhibirà el material, ja que cada pantalla té una ràtio d'aspecte pròpia que, per tant, és òptima per a aprofitar al màxim la seva superfície. És per això que a l'hora de gravar a vegades s'utilitza el que s'anomena àrea segura (*Safe Area*), que marca la zona del fotograma que després es veurà sense problemes en la majoria de ràtios d'aspecte més habituals.

3.4. Quantitat d'informació

La ràtio, els fps, la resolució i la profunditat de color del material que gravem ens determinen en tot moment la quantitat màxima d'informació de la qual disposem. La qualitat de la imatge final dependrà de com es manipula aquesta informació, que si no es vigila es pot anar reduint o degradant durant el procés. És important tenir-les en compte a l'hora de captar i editar les imatges, perquè la pèrdua d'informació és un procés irreversible. La informació sempre es pot reduir, però mai augmentar.

Exemple de degradació de la informació. La primera i la tercera imatge tenen la mateixa resolució. La tercera, però, està recalculada a partir de la segona. Tot i tenir la mateixa quantitat de píxels que la primera, la tercera imatge té la mateixa quantitat d'informació que la segona.
Font: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=164041

La quantitat d'informació és independent de la resolució. Si gravem quelcom amb una ràtio 4:3, 15 fps, 480 p i 8 bit, per més conversions que li fem, no aconseguirem generar una imatge ni millor ni més gran. Podem retallar el fotograma per canviar la ràtio a 16:9 o afegir bandes negres, però no recuperar el que ens ha quedat fora de camp. Podem duplicar cada fotograma o fins i tot calcular els que manquen per obtenir 30 fps, però no obtindrem més informació en aquest fotograma nou. De la mateixa manera podem augmentar el nombre de píxels a 1080 p i la seva profunditat de color a 48 bit, però per més que el fitxer resultant sigui més gran no recuperarem la informació que s'ha perdut pel camí. El fitxer resultant, encara que sigui d'alta resolució, contindrà una imatge de baixa qualitat.

Gravar a major resolució de la que s'utilitzarà al clip final, en canvi, no presenta aquest problema. La quantitat d'informació sempre es pot reduir. A més, fins a un cert punt, ens permet jugar amb aquesta informació extra. Així, si hem de treballar en un projecte a 720 p i disposem d'una càmera 4K té sentit gravar amb el màxim de resolució disponible. En la fase de postproducció es pot utilitzar la imatge 4K sencera sense problemes, els programes d'edició

n'ajustaran la resolució a l'hora d'exportar. Al mateix temps podem utilitzar-ne només una part, ja que la informació en excés ens permet ampliar una part de la imatge a mida de fotograma sense que la resolució final se'n ressenti. De la mateixa manera es podrà fer zoom digital, o estabilitzar la imatge eliminant tremolors sense que apareguin zones negres als marges.

4. Formats de vídeo digital

Després d'haver analitzat totes les característiques intrínseques del vídeo digital, només ens falta entendre com s'emmagatzema i processa. Actualment, des de que és captat pel sensor de la càmera i fins que es reprodueix, el vídeo s'emmagatzema en arxius d'ordinador. Aquests arxius han de contenir informació suficient per a poder entendre el flux de dades, cosa que passa per especificar fps, resolució, sistema i profunditat de color, ràtio d'aspecte, pistes de so, còdec, etc. Per això, els arxius de vídeo utilitzen diferents formats que permeten als programes d'edició i de reproducció reconèixer i tractar adequadament el flux de vídeo.

És important no confondre el format d'arxiu amb el còdec. Tal com veurem, el format no diu res sobre com funcionen els mètodes de codificació i compressió (que veurem més endavant). L'única cosa que fa el format és indicar de quina manera està estructurada internament la informació.

4.1. Formats

Els formats són fitxers estandarditzats. Normalment es distingeixen fàcilment entre ells per l'extensió de l'arxiu. Cada format estructura la informació interna a la seva manera, i poden variar molt entre ells. A l'hora de triar un format cal tenir en compte el següent:

- N'hi ha una gran varietat, cadascun d'ells dissenyat per a un ús específic. Alguns estan pensats per a la producció, com el DPX, i d'altres per a fer arribar al consumidor final, com el 3GP.

- Poden estar més o menys suportats. Un format ben difós, com l'MP4, és normalment una garantia per a assegurar la compatibilitat amb tots els programes i dispositius. En canvi alguns formats, com el Matroshka (MKV), tenen un suport menor i només els reconeixen alguns programes i reproductors.

- Alguns són de pagament. Molts formats, com ara el FLV, estan protegits per patents. En la majoria dels casos l'usuari final no ha de pagar res, però sí que cal adquirir-ne una llicència per a incorporar-lo als programes d'edició i de reproducció. D'altres, com l'OGG, són completament lliures i gratuïts.

- No tots els formats tenen les mateixes opcions. Alguns, per exemple, ofereixen la possibilitat de fer-los servir per *streamings*, de gestionar contingut DRM (control digital de contingut, de l'anglès *Digital Right Management*), d'incloure subtítols, d'utilitzar un còdec específic, etc.

Logotip del format MKV, un format lliure
Font: https://www.matroska.org/info/trademarks/index.html. Domini públic, https://commons.wikimedia.org/w/index.php?curid=41309227

Alguns dels formats més habituals són els següents:

Nom	Extensions	Característiques més importants
RAW	.cr2, .nef, .rw2, .sr2, .yuv,...	Sense format. Informació directa del sensor de la càmera. Poc suport.
3GP	.3gp, .3g2	Pensat per a reproducció, en *streamings* o aplicacions 3G.
AVI	.avi	Pensat per a vídeo poc comprimit, i per tant serveix tant per a edició com per a reproducció. Antiquat, però molt utilitzat.
FLV	.flv, .f4v,...	Pensat per a reproducció en web.
Matroshka	.mkv, mk3d,...	Pensat per a reproducció, però pot contenir de tot.
MP4	.mp4, .m4p	Pensat per a reproducció, especialment amb grans compressions.
MXF	.mxf	Pensat per a edició.
OGG	.ogg, .ogv	Pensat per a reproducció.
QuickTime	.mov, .qt	Pensat per a edició.

Tot i això, el format no és més que la punta de l'iceberg. És habitual confondre el format d'arxiu amb el seu contingut. El format és simplement un contenidor, una manera de distingir uns fitxers d'uns altres, però no especifica com interpretar el flux de dades. D'això, se n'encarrega el còdec. Un mateix format pot utilitzar molts còdecs diferents, de la mateixa manera que sovint presenta fps o resolucions diferents. A més, arxius en diversos formats poden utilitzar el mateix còdec. A l'hora d'editar i de reproduir, doncs, hem de parar atenció a quin còdec estem fent servir.

4.2. Còdecs

Els còdecs (contracció de l'anglès, ***co**der* ***dec**oder*) són les fórmules que s'utilitzen per a emmagatzemar el flux de dades. Bàsicament són algoritmes, que determinen com s'ha de codificar i descodificar la informació.

Logotip del còdec lliure FFmpeg
Font: Hervé Flores - movieconverter-studio.com/ PUBLIC/ffmpeg/logo-new/ffmpeg-logo-src/ - ffmpeg-logo.svg, ffmpeg-logo.png. Domini públic https://commons.wikimedia.org/w/index.php?curid=25543554

És important codificar la informació perquè, si s'utilitzessin sempre els fluxos de dades sense codificar, els fitxers serien immensos i en resultaria impossible l'emmagatzematge i la transmissió. De la mateixa manera, és important que aquest procés estigui ben controlat, de manera que l'editor o el reproductor puguin descodificar la informació.

A part del grau de compatibilitat amb editors i reproductors, les diferències més importants entre còdecs són en com gestionen la compressió i la pèrdua. Aquestes són les característiques que fan que siguin més o menys adequats per a l'edició o per a la reproducció.

Compressió

Comprimir significa, estrictament, fer que la informació ocupi menys espai. Per exemple, codificant només amb compressió podem fer que un arxiu de vídeo que ocupa 1 GB passi a ocupar 700 MB. Això s'aconsegueix en primer lloc buscant les similituds entre diferents parts de la informació (redundància) i descrivint-la com a tal.

Per exemple, si en el nostre vídeo tenim diversos fotogrames completament negres, ens podríem estalviar de descriure'n els píxels un per un si especifiquem que el primer píxel és negre i que des del segons i fins al final de la seqüència són identics al primer. A l'hora de descodificar ens seria molt fàcil reconstruir aquests fotogrames negres tal i com eren a l'origen, encara que en el fitxer comprimit no trobem els píxels descrits un per un.

Normalment sempre es comprimeix amb pèrdua, però, si no n'hi hagués gens, després ens seria possible tornar a descomprimir exactament la mateixa imatge, tal i com era originalment.

Exemple de *morphing*. La imatge central es pot reconstruir perfectament només amb les dues imatges laterals.
Font: Iainf 18:46, 18 Juny 2006 (UTC) - Imatges originals: George-W-Bush.jpeg and w:Image:Arnold_Schwarzenegger_bio.jpg.
Domini públic, https://commons.wikimedia.org/w/index.php?curid=875906

Pel que fa al vídeo bàsicament hi ha dos mètodes de compressió diferenciats: els que es basen en comprimir la informació de dins del fotograma (compressió *intraframe*), i els que comprimeixen un fotograma d'acord amb els fotogrames adjacents (compressió *interframe*). Aquests segons són els que aconsegueixen un grau més gran de compressió. Com que molt sovint els fotogrames s'assemblen entre ells, se'n sol descriure sencer només un a intervals regulars, l'anomenat *I-frame*, mentre que la resta es reconstrueixen a partir d'aquest. Els canvis entre un *I-frame* i el següent per mitjà d'informacions més simples, com ara el desplaçament o el *morphing*, aconsegueixen una gran reducció de la quantitat d'informació. Aquestes tècniques sovint són molt fàcils de detectar quan hi ha errors en els *I-frames*, com passa en la imatge inferior, ja que el reproductor intenta reconstruir la imatge basant-se en una informació errònia i per tant omple certes zones de la pantalla amb formes i colors que no es corresponen a res en concret, però que sovint es poden reconèixer perquè recorden el moviment dels objectes que haurien de representar.

Els algoritmes de compressió, doncs, busquen totes les similituds matemàtiques del nostre fitxer i les descriuen amb el mínim d'informació possible. Quan aquest procés es fa dins del mateix fotograma ens permet guanyar només una mica d'espai, però és relativament fàcil i ràpid de revertir a la fase de descodificació, i per tant els fitxers resultants es poden editar sense gaires problemes. Quan la compressió es fa entre fotogrames, però, l'edició es fa més difícil, ja que el programa ha de reconstruir el fotograma, a la qual cosa destina molts recursos i baixa el rendiment. És així que s'explica que editar fitxers de vídeo més grans consumeixi menys recursos de l'ordinador que no pas utilitzar arxius molt més petits.

Fotograma reconstruït a partir d'un *I-frame* corromput
Font: LikakiPhotos - Treball propi, CC BY-SA 3.0, **https://commons.wikimedia.org/w/index.php?curid=18308443**

Pèrdua

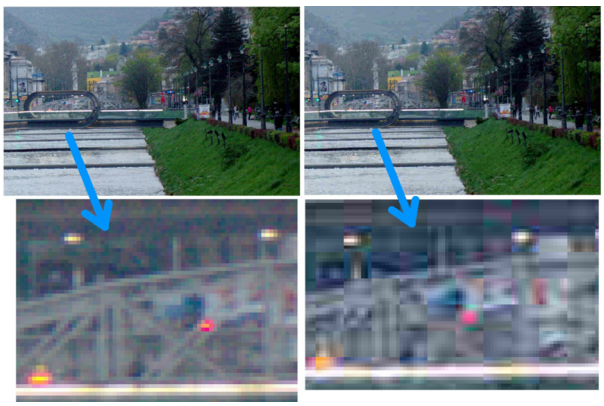
La pèrdua és el mètode de codificació que ens permet estalviar més espai. Consisteix a eliminar o simplificar part de la informació, de manera que es redueix dràsticament la mida del fitxer. Cal distingir-la de la compressió perquè, com ja hem dit abans, la pèrdua és un procés irreversible.

La pèrdua és independent de la compressió. Un fitxer pot estar molt comprimit i no tenir gens de pèrdua. Un altre, en canvi, pot tenir molta pèrdua i no estar gens comprimit. De totes maneres, com hem dit, se solen aplicar sempre tots dos en major o menor mesura.

Un televisor antic mostrant neu analògica.
Font: Mysid - Self-taken photograph. Domini públic, https://commons.wikimedia.org/w/index.php?curid=848644

Aplicar mètodes amb pèrdua significa reduir la quantitat d'informació encara que no es pugui recuperar. Per exemple, si en una seqüència de vídeo hi apareix un televisor analògic antic no sintonitzat, la «neu» que apareix a la pantalla consisteix en píxels aleatoris en blanc i negre, sense cap lògica, i per tant una compressió sense pèrdua hauria de descriure-la píxel per píxel per a poder-la reconstruir exactament igual. En canvi, aplicant un bon mètode amb pèrdua es podria descriure simplement com el que són: píxels aleatoris, cosa que es pot descriure amb una petita expressió matemàtica. Tot i que la neu no serà mai exactament la mateixa, serà impossible notar la pèrdua, mentre que la mida del fitxer s'haurà reduït enormement. Al mateix temps, però, també ens serà impossible recuperar els píxels exactes de la neu original a partir del fitxer comprimit.

A la pràctica, la majoria de mètodes amb pèrdua es dediquen a simplificar o a descartar informació. Les mateixes tècniques de *morphing* i de desplaçament que comentàvem abans solen provocar pèrdues d'informació.



Ampliacions d'una mateixa imatge amb dos nivells de compressió diferents. A la dreta es poden observar diversos artefactes, tot i que la majoria són difícils de distingir en la imatge sencera.

Alguns dels altres mètodes més utilitzats són el **Chroma Subsampling** i la transformada cosinus discreta. Quan tots aquests mètodes s'utilitzen en excés es diu que la imatge presenta artefactes de compressió, és a dir, distorsions recognoscibles que no eren a la imatge original.

És molt important entendre que aquests artefactes no depenen de la resolució del vídeo, que pot ser perfectament de 4K i presentar artefactes, sinó de la pèrdua d'informació del còdec.

En els fitxers destinats a reproducció, la pèrdua pot tenir efectes molt positius. Com ja hem dit, la mida del fitxer es redueix dràsticament. A més, fins a un cert punt, la pèrdua passa totalment inadvertida. Si per exemple el fitxer d'origen té una profunditat de color de 48 bit, es pot reduir perfectament a 24 bit, que és aproximadament la màxima profunditat de color que l'ull humà pot distingir i per tant és indistingible de la profunditat de color original. El Chroma Subsampling també es basa a reduir la informació del color sense que l'ull humà se n'adoni. Molts dels formats de reproducció utilitzen còdecs que eliminen aquestes informacions supèrflues.

Pel que fa a l'edició, en canvi, la pèrdua d'aquesta informació pot significar una gran limitació. D'una banda, dificulten moltíssim el procés d'edició, ja que la reconstrucció dels fotogrames consumeix una gran quantitat de recursos de l'ordinador. De l'altra, la pèrdua d'informació fa que qualsevol modificació de la imatge creï fàcilment nous artefactes. Especialment, els artefactes poden aparèixer quan s'exporta, ja que el nou còdec no està preparat per a comprimir de nou la informació que ja contenia pèrdues importants.

Cal tenir en compte que la pèrdua depèn del còdec, no del format. Com hem dit abans, el format només descriu el contingut, però accepten una gran varietat de còdecs. Així, doncs, a l'hora d'importar material, si el format que tenim no està clarament destinat a la producció de vídeo cal tenir molt clar quin tipus de còdec conté, ja que segons quin sigui i del que haguem de fer el material pot ser de gran qualitat o gairebé inservible.

Còdecs més habituals

Hi ha principalment dos tipus de còdecs: els que estan pensats per a editar i els que estan pensats per a ser reproduïts. Sempre que sigui possible cal utilitzar formats d'edició intermediaris, de gran mida però amb poca pèrdua.

Alguns dels còdecs més habituals en petites produccions són els següents:

Nom	Ús principal recomanat
AppleVideo	Reproducció
H.264	Edició i reproducció
H.265	Reproducció
MPEG-1	Reproducció. Antiquat
MPEG-2	Reproducció
ProRes	Edició
Theora	Reproducció
VP9	Reproducció

També cal tenir en compte que no tots els formats poden incloure tots els còdecs, alguns no són compatibles entre ells. Per exemple, l'AVI no és compatible amb el còdec VP9. De totes maneres, molt sovint arxius en formats diferents utilitzen el mateix còdec.

Array de discos durs utilitzats en producció de vídeo
Font: Open Grid Scheduler / Grid Engine - Treball propi, CC0. https://commons.wikimedia.org/w/index.php?curid=38267117

La tria del còdec d'edició depèn sobretot de la mida de la producció. Com menys volum de dades tingui el projecte, més fàcil serà utilitzar compressions baixes. Els formats amb menys compressió i sense pèrdua serien sempre ideals per a editar, però suposen un problema d'emmagatzematge. Els grans estudis generalment sí que treballen amb formats intermediaris, amb una mida que obliga a emmagatzemar la informació en raids com el de la fotografia, que poden contenir desenes de discos durs. Fins i tot en aquests casos els formats

que s'utilitzen tenen pèrdua respecte al *raw*. Al mateix temps, els formats finals en els quals s'exporten els projectes tenen molta més pèrdua que no pas els fitxers intermediaris.

Tal com es pot deduir, intentar treballar sense compressió ni pèrdua en produccions de baix pressupost pot ser virtualment impossible. Moltes vegades, doncs, l'única opció és utilitzar formats amb pèrdua moderada. Si es té la possibilitat, però, és millor aplicar els mètodes de pèrdua com més tard millor.