



**UNIVERSITA' POLITECNICA DELLE MARCHE  
FACOLTA' DI INGEGNERIA**

---

**Corso di Laurea in Ingegneria Elettronica**

**ANALISI DI FORMATI MULTIMEDIALI AUDIOVIDEO E DEL  
PROCESSO DI POST-PRODUZIONE**

**ANALYSIS OF MULTIMEDIA AUDIOVIDEO FORMATS AND  
THE POST-PRODUCTION PROCESS**

Relatore:

**Prof. Ennio Gambi**

Tesi di Laurea di:

**Fabrizio Acciavatti**

**A.A. 2019 / 2020**



# Indice

<b>Sommario</b>	<b>4</b>
<b>1. Formati multimediali audiovisivi</b>	<b>5</b>
<b>2. Codifiche video</b>	<b>7</b>
2.1. Cos'è una codifica video . . . . .	7
2.2. Acquisizione di una sequenza video . . . . .	8
2.3. Funzionamento di una generica codifica video . . . . .	9
2.4. Alcune codifiche video . . . . .	17
2.5. Accenno agli algoritmi di compressione audio . . . . .	22
<b>3. Fasi operative di un processo di post-produzione di un'emittenza televisiva</b>	<b>25</b>
3.1. Utilizzo del formato MXF . . . . .	25
3.2. Trasmissione DVB-T e DVB-T2. . . . .	33
<b>4. Conclusioni</b>	<b>34</b>
<b>Bibliografia</b>	<b>36</b>

# Sommario

L'utilizzo e la condivisione di contenuti multimediali digitali è divenuto un aspetto quotidiano della nostra vita. Tra di essi i video svolgono un ruolo molto importante. La maggiore richiesta di qualità di visualizzazione e una distribuzione dei contenuti sempre più efficiente portano ad un aumento dei requisiti in termini di prestazioni. Pertanto, le tecniche di compressione video risultano in continua evoluzione. Questa tesi ha l'obiettivo di presentare la teoria di base su cui si fonda la codifica video, per poi descrivere i principali standard sviluppati fino ad oggi.

# Capitolo 1

## Formati multimediali audiovisivi

### 1.1 L'essenza ed i metadata

Qualsiasi dato multimediale è composto da due parti fondamentali: l'essenza ed i metadata.

La prima costituisce l'elemento multimediale stesso. Un'essenza può essere quindi un video, una foto oppure un brano audio. Ci sono molti formati di essenze come:

- wav, mp3, per l'audio;
- png, jpeg per le immagini;
- mpeg, avi per i filmati.

Bisogna unire all'essenza l'altra componente fondamentale dei dati multimediali, ovvero i metadata, che forniscono informazioni sull'essenza stessa. I metadata sono dei dati descrittivi riguardanti un elemento multimediale come un video o un brano audio. Svolgono la funzione di identificare il materiale come se fossero delle etichette. I metadata forniscono informazioni supplementari che descrivono ulteriormente l'essenza come ad esempio l'autore del brano o del video, i luoghi della ripresa e così via. Possiamo suddividere i metadata in due categorie fondamentali: i metadata descrittivi ed i metadata strutturali.

I metadata descrittivi si suddividono in:

- **Identificativi e di locazione** racchiudono tutto ciò che possa identificare l'essenza.
- **Di processo** descrivono come l'essenza viene assemblata, come viene composta o editata.
- **Amministrativi** comprendono diritti, modalità di accesso protetto dell'utente e tutto ciò che concerne la parte amministrativa.
- **Interpretativi** Includono la descrizione del linguaggio, il numero della scena o del brano, il nome dell'artista o degli attori, o dei personaggi.
- **Parametrici** comprendono le codifiche ed i parametri di streaming.

I metadata strutturali invece possono essere:

- **Di relazione** contengono i collegamenti, le relazioni e le sincronizzazioni tra i vari media.
- **Spaziali e temporali** hanno la funzione di creare dei locatori spaziali e temporali attraverso la creazione di date ed indicazioni sulla durata dell'elemento multimediale.

In passato archiviare i dispositivi di registrazione ed i metadata avveniva su due supporti diversi e questo poteva creare lo spiacevole inconveniente di perdere uno dei due supporti, perdendo anche così la possibilità di ritrovare ciò che stavamo cercando nell'archivio multimediale.

Per risolvere questo tipo di problema alla radice basta registrare il materiale ed i dati descrittivi, quindi i metadata, in un unico file, utilizzando un formato standard che come tale sia riconosciuto da tutti i soggetti interessati, ossia dai produttori audio e video e da coloro che le utilizzano.

# Capitolo 2

## Codifiche video

### 2.1 Cos'è una codifica video

La codifica è definita come il processo con cui un particolare insieme di dati viene rappresentato utilizzando una quantità ridotta di informazioni. La codifica video consiste quindi nel processo di conversione di un video digitale in un flusso di dati numerici di dimensioni più adatte alla trasmissione e all'immagazzinamento su memorie di massa. Questo permette di trattare quantità di dati sempre superiori, mantenendo un alto grado di qualità. Il principio si basa sull'eliminazione delle ridondanze, delle informazioni che vengono ripetute per un certo tempo, siano esse frequenze, parti di immagine o colori simili.

Ogni dato acquisito o documento occupa un determinato spazio in memoria o su disco, misurato in bit. I contenuti multimediali quali immagini e video richiedono uno spazio di archiviazione più elevato rispetto ad altri tipi di informazione. Questo è sempre stato oggetto di studio e, nel corso degli ultimi anni, sono state date diverse risposte. Negli ultimi anni possiamo notare come sia sempre crescente la quantità di informazione che è possibile immagazzinare su supporti digitali. Non si deve dimenticare il fatto che questi contenuti, oltre ad essere memorizzati, devono anche essere trasmessi. La quantità di dati che possono essere inviati e ricevuti è oggi molto più elevata rispetto ad anni fa. Però i supporti di memorizzazione e le reti risultano essere comunque un collo di bottiglia in molteplici situazioni. Per esempio, un video 4K non compresso richiede circa 30 GB di spazio di archiviazione; questo potrà sembrare una quantità accettabile date le caratteristiche (risoluzione 4K, profondità di pixel HDR, 60 fps), ma parliamo di un video che ha una durata di soli 20 secondi. Se consideriamo un film con le stesse caratteristiche ma della durata di 2 ore, esso richiederebbe ben 10,8 TB. Inoltre, dopo alcuni semplici calcoli si ottiene che il bit-rate è di circa 12 Gbit/s. Attualmente i dispositivi non sono in grado di trasferire dati ad una velocità simile, perciò il video non può essere riprodotto se non frame per frame. In aggiunta,

è facile stimare che la quantità di dati da trasmettere non consente lo streaming della sequenza video attraverso una rete, locale o remota che sia.

La soluzione a tutti questi problemi è una sola: la compressione (o codifica). Essa consente di ridurre enormemente la quantità di dati necessaria per rappresentare lo stesso contenuto. In base all'algoritmo di compressione utilizzato i codec si distinguono in due grosse famiglie. Ci sono quelli **lossy** che sono capaci di comprimere maggiormente i file sacrificando nel processo parte dell'informazione. Nel caso dell'audio, delle immagini e dei filmati, un certo livello di degradazione è un compromesso accettabile per ridurre l'occupazione o la banda richiesta dal file.

Ci sono poi quelli **lossless** invece che mantengono tutte le informazioni contenute in ogni singolo fotogramma. E' ovvio che nel comprimere informazioni come testi, documenti o programmi non ci si possa permettere la perdita di nessun bit di informazione, per cui dovremo utilizzare necessariamente tecniche lossless.

## **2.2 Acquisizione di una sequenza video**

Qualunque scena vista dal nostro occhio si può considerare, matematicamente parlando, una proiezione di valori di luminosità rappresentabili da numeri reali su di un piano cartesiano in diversi istanti. Modellando il segnale come una funzione, possiamo pertanto dire che sia il dominio (dal punto di vista spaziale e temporale) sia il codominio sono continui. Per acquisire un'immagine 2D dal mondo reale lenti circolari focalizzano la radiazione luminosa proveniente dagli oggetti su sensori costituiti da un array di CCD (Charge Coupled Device) o CMOS (Complementary Metal-Oxide Semiconductor). Essi convertono quindi i segnali luminosi in segnali elettrici, che vengono poi elaborati. La rappresentazione in forma digitale comporta una discretizzazione dei dati.

### **Discretizzazione spaziale**

Considerando la proiezione della scena su di un'immagine (es. un fotogramma di un video), essa viene partizionata in blocchi regolari, applicando cioè una griglia quadrata o rettangolare all'immagine acquisita. Tale griglia corrisponde alla matrice di sensori utilizzata dal dispositivo e viene misurato il valore di tensione. Ogni valore costituisce un pixel. Il numero di campioni che viene prelevato dall'immagine influenza pesantemente il livello di dettaglio che si otterrà alla fine del procedimento.



### **Discretizzazione temporale**

Il campionamento dal punto di vista del tempo consiste nel catturare un'immagine ad intervalli regolari nel tempo. Una frequenza di campionamento elevata porta ad avere un video in cui i movimenti sono fluidi e molto simili a quelli naturali. Al contrario, una frequenza bassa porta ad avere una visione "a scatti". Valori tipici sono dell'ordine dei 25- 30 frame per secondo per video a definizione standard, 50-60 frame per secondo quando si considera l'alta definizione.

Il valore di ogni pixel è rappresentato da un numero intero a precisione finita rappresentato da un certo numero di bit. La profondità di colore adottata è individuata dal numero di bit di cui è costituita ogni rappresentazione. Tipicamente si utilizzano 8 bit, ma sempre più diffuso è l'uso di 10 bit. In questo caso si parla di immagine HDR (High Dynamic Range).

### **Interlacciato o progressivo**

Quello che è stato descritto finora è chiamato campionamento di tipo progressivo. Grazie ad esso ogni frame risulta costituito da tutti i pixel di cui è formato. Un altro tipo di campionamento molto diffuso è invece quello interlacciato. Esso richiede la suddivisione di ogni frame in linee orizzontali. In una sequenza video interlacciata ogni fotogramma è costituito dalle sole linee con indice pari o dispari, con la conseguenza che la risoluzione in altezza risulta dimezzata. Il vantaggio di questo approccio consiste nella possibilità di trasmettere una quantità doppia di frame rispetto al progressivo, mantenendo però lo stesso bit-rate. Il ricevente ha così una visione più fluida del filmato senza una perdita di qualità rilevante. Il passaggio da una tecnica di campionamento ad un'altra comporta un'operazione di conversione che richiede un'elaborazione aggiuntiva e la possibile presenza di artefatti. Poiché il campionamento interlacciato è strettamente legato al funzionamento degli schermi analogici del passato, esso sta rapidamente lasciando il passo a quello progressivo.

## **2.3 Funzionamento di una generica codifica video**

I diversi standard di codifica video che sono stati sviluppati negli ultimi vent'anni condividono lo stesso schema di base. Ciò che li distingue è infatti principalmente l'efficienza, oltre all'applicazione di destinazione per cui sono stati pensati.

Per realizzare un sistema di codifica sono necessari due componenti tra loro complementari, ovvero un encoder (codificatore, compressore) ed un decoder (decodificatore, decompressore). L'encoder si occupa di convertire i dati della sorgente in un bit stream, richiedendo la minima quantità possibile di bit per la sua rappresentazione. Questo va effettuato prima della memorizzazione e dell'eventuale trasmissione. Per far ciò si rimuove la ridondanza statistica insita nei dati, applicando una cosiddetta codifica di sorgente. La codifica, come già detto, può essere senza perdita di informazione (codifica lossless) ma per ottenere risultati più efficienti dal punto di vista della dimensione del bit stream, si accetta di avere una perdita più o meno significativa (codifica lossy).

Il decoder invece è necessario per riconvertire i dati compressi nella loro forma originaria, o approssimarla il più possibile se si è optato per una codifica lossy.

La coppia enCOder/DECOder è indicata con la parola codec.

## **Rappresentazione dei colori**

Per la rappresentazione di un'immagine è necessario codificare per ogni pixel informazioni quali la luminosità e i colori. Esistono varie rappresentazioni chiamate spazi colore.

- **RGB**

Nello spazio colore RGB, ogni pixel è rappresentato da 3 numeri interi che indicano la quantità di rosso, verde e blu. Combinando questi tre colori si può creare ogni altro colore.

Ogni immagine è quindi rappresentata da 3 matrici, una per ogni componente di colore.

Come già accennato, il numero di bit utilizzato per rappresentare una singola componente determina la profondità di colore dell'immagine e quindi l'effettivo numero di tonalità che possono essere rappresentate. L'utilizzo di 8 bit per componente consente di avere 256 possibili tonalità di rosso, verde e blu. Combinandole insieme il numero di colori rappresentabili è 16,8 milioni circa. Grazie all'utilizzo di 10 bit per componente il numero di colori diventa 1,07 miliardi.

- **YCbCr**

Lo spazio colore YCbCr deve la sua origine ad una considerazione riguardo il sistema visivo dell'uomo (HVS = Human Visual System). Infatti è noto come l'occhio umano sia meno

sensibile ai colori rispetto alla luminosità. RGB non tiene conto di questa osservazione e quindi dà la stessa importanza ad ognuna delle 3 componenti. YCbCr invece separa le informazioni relative alla luminosità (luminanza, Y) da quelle riguardanti i colori (crominanza blu Cb, crominanza rossa Cr e crominanza verde Cg). Si definisce quindi la luminanza Y come media pesata tra le componenti dei colori:

$$Y = k_r * R + k_g * G + k_b * B$$

Le crominanze rossa Cr, blu Cb e verde Cg sono invece la differenza tra l'intensità della componente colore e Y:

$$Cr = R - Y \qquad Cb = B - Y \qquad Cg = G - Y$$

Si potrebbe pensare che questa rappresentazione sia svantaggiosa dato che prevede 4 componenti contro le 3 di RGB. Tuttavia si può notare che la somma tra le 3 crominanze è una costante, ovvero proprio la luminanza Y. Perciò, per esempio, il valore di Cg si può ricavare come differenza tra Y e le altre 2 crominanze debitamente pesate. Non è quindi necessario che venga trasmessa.

Inoltre, vista l'osservazione riguardante l'HVS, è possibile rappresentare le componenti Cb e Cr con una risoluzione inferiore rispetto ad Y. Gli schemi di decimazione delle crominanze (chroma subsampling) più frequentemente usati sono:

- 4:4:4

Ognuna delle 3 componenti ha la stessa risoluzione e quindi ad un campione di Y corrisponde un campione Cb ed uno Cr.

- 4:2:2

Le componenti di crominanza hanno la stessa risoluzione verticale della componente di luminanza, ma la risoluzione orizzontale è invece dimezzata. Ogni 4 campioni di Y ce ne sono 2 di Cb e 2 di Cr.

- 4:2:0

La risoluzione delle componenti di crominanza è dimezzata sia verticalmente che orizzontalmente e quindi a 4 campioni Y corrispondono un solo campione Cb ed uno Cr.

Come per RGB ogni immagine è rappresentata da 3 matrici, sebbene abbiano un significato diverso. La differenza fondamentale sta nel fatto che esse possono non avere la stessa dimensione. L'utilizzo dello spazio colore YCbCr può quindi essere considerato una prima forma di compressione.

## **Predizione**

Una sequenza video è formata da immagini che vengono visualizzate una dopo l'altra in modo da rappresentare, almeno nel caso di filmati dinamici, soggetti in movimento. Tra un fotogramma e l'altro però solitamente non tutta l'immagine varia. Si può perciò sfruttare la correlazione tra frame successivi per ridurre la quantità di informazioni da memorizzare.

Anche all'interno di un singolo fotogramma è possibile trovare una correlazione tra pixel adiacenti.

La predizione di un frame basata sui pixel vicini tra loro (Intra-frame) e sui fotogrammi temporalmente adiacenti (Inter-frame) è un elemento fondante di tutte le codifiche video e le tecniche di predizione sono state continuamente migliorate nel passaggio da uno standard al successivo.

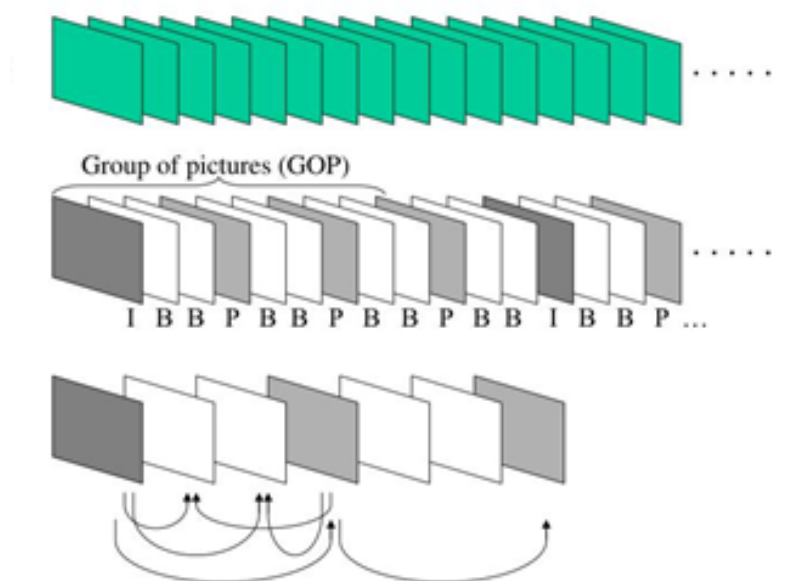
## **Predizione spaziale Intra-Frame**

Il fotogramma viene suddiviso in unità fondamentali chiamate blocchi. I blocchi sono di forma quadrata e sono costituiti da più pixel. Dimensioni tipiche sono 4x4 pixel, 8x8 o 16x16 (nella codifica H.264 / AVC) e la dimensione può essere differente per componenti distinte dell'immagine (per esempio con lo spazio colore YCbCr si possono avere blocchi di dimensione 16x16 pixel per la luminanza e 8x8 pixel per le componenti di cromaticità). Più la dimensione è piccola, maggiore è l'accuratezza della predizione e minore è il numero di bit richiesti per codificare il blocco. Tuttavia è necessario un certo numero di bit per segnalare al decodificatore che un certo blocco non è rappresentativo dell'immagine originale ma è invece il risultato di una predizione e per specificare tutte le informazioni necessarie a riprodurre la predizione. Perciò potrebbe accadere che la quantità di bit risparmiati non trasmettendo le informazioni sui pixel risulti inferiore a quella necessaria per le segnalazioni. Ciò porta quindi a considerare l'idea di utilizzare blocchi più grandi. In sintesi, la predizione intra-frame utilizza i campioni dei blocchi adiacenti già codificati per predire i valori delle componenti di colore o di luminanza e cromaticità del blocco considerato.

Le tecniche di compressione più avanzate come la recente H.265 / HEVC fa uso anche di macro-blocchi con dimensioni di 64x64 pixel, ma allo stesso tempo può anche utilizzare partizioni più complesse rispetto a quelle quadrate.

### **Predizione temporale Inter-Frame**

La predizione inter-frame è nota anche come stima del moto. Così come la predizione intra-frame, essa si basa sulla suddivisione di ogni immagine in blocchi o macro-blocchi formati da pixel. Una volta considerato un fotogramma, se ne prende un altro già codificato e quest'ultimo diventa il frame di riferimento (reference frame). Per ogni blocco del frame considerato, si cerca nel reference frame il blocco che più lo approssima. Per ottimizzare la ricerca ci si può limitare a cercare il blocco candidato in un'area limitata attorno al blocco considerato. Quando il candidato migliore è stato trovato, si crea un vettore di movimento (motion vector) che esprime l'offset orizzontale e verticale tra i blocchi dei due fotogrammi. L'operazione viene quindi eseguita per ogni blocco e si ottiene infine un insieme di vettori che indicano l'eventuale spostamento dei blocchi. Dalla collezione di vettori si ricavano i blocchi che vengono predetti. Infine ogni blocco predetto viene sottratto al blocco di riferimento per ottenere un blocco residuo, quello che effettivamente verrà memorizzato o trasmesso. Anche in questo caso minore è la dimensione dei blocchi maggiore è l'attendibilità della predizione, a discapito della capacità di compressione.



**Figura 2.1.:** Esempio di Group of Pictures

Una sequenza video può essere considerata formata da gruppi di immagini detti GOP (Group Of Pictures). La struttura di un GOP (fig.2.1) determina quale o quali sono i reference frame per un determinato fotogramma. Un GOP può contenere frame di diversi tipi:

- **I-frame** (Intra-coded frame): consiste di un'immagine che viene compressa mediante la sola predizione spaziale, senza fare riferimento a nessun altro frame del GOP. Fa anzi da riferimento per tutti gli altri tipi di fotogrammi, essendo quello che approssima certamente con più fedeltà l'immagine originale. Un Group Of Pictures contiene sempre almeno un I-frame.
- **P-frame** (Predictive frame): utilizza come reference frame un fotogramma precedente. Richiede meno bit rispetto ad un I-frame dato che vengono codificate solo le differenze tra il frame considerato e quello di riferimento (che può essere sia un I, un B o un altro P).
- **B-frame** (Bi-predictive frame): utilizza come reference frame sia un fotogramma precedente che uno successivo. La compressione che si ottiene è superiore rispetto a quella dei frame I e P, ma soffre di una latenza maggiore dato che per la codifica è necessario attendere che sia stato compresso il frame temporalmente successivo. Ciò provoca quindi una codifica out-of-order.

Sebbene sia ancora largamente utilizzata la predizione a blocchi e macro-blocchi, numerosi studi hanno dimostrato che il partizionamento orientato agli oggetti (ovvero adattando le regioni alle dimensioni degli oggetti rappresentati in una sequenza video) porta a risultati sensibilmente migliori, a discapito però della crescente complessità computazionale che ciò comporta. HEVC è il primo standard di codifica ad adottare un approccio ibrido tra quello a blocchi e quello ad oggetti.

### **Codifica a trasformate**

Ora che sono stati definiti quali sono e come sono strutturati i frame che andranno a comporre la sequenza video, si passa alla fase di vera e propria codifica lossy dei singoli fotogrammi.

Essa è composta da numerose fasi, ovvero:

- Preparazione dei blocchi
- Passaggio al dominio della trasformata
- Quantizzazione
- Codifica entropica

## Preparazione dei blocchi

Operando la stessa suddivisione in blocchi e macro-blocchi utilizzata per la predizione, ogni valore dei pixel di cui è composto viene centrato in 0 sottraendo  $2^{(n-1)}$ , dove  $n$  è il numero di bit che viene usato per la rappresentazione. Quindi nel caso di profondità di colore a 8 bit si sottrae 128, nel caso di 10 bit (HDR) 512. Questa operazione va effettuata su ogni componente colore se lo spazio colore utilizzato è RGB, solo su Y se è YCbCr.

## Passaggio al dominio della trasformata

Questa operazione permette di ridurre la correlazione, così da concentrare in un numero minore possibile di valori l'energia del segnale considerato. La trasformata deve essere reversibile e implementabile, senza richiedere un utilizzo di risorse eccessivo. Come per la compressione delle immagini JPEG, anche in MPEG-1 e successivi si predilige l'uso della trasformata DCT (Discrete Cosine Transform) invece della classica FFT (Fast Fourier Transform) poiché raggiunge risultati sensibilmente migliori nella riduzione della ridondanza, a discapito però di una complessità computazionale maggiore. La trasformata discreta del coseno (DCT) opera su  $\mathbf{X}$ , un blocco di  $N \times N$  campioni visto come matrice, fatti da pixel ottenuti direttamente dall'immagine originale o tramite predizione. Ciò che calcola è  $\mathbf{Y}$ , ovvero un blocco delle stesse dimensioni di  $\mathbf{X}$  ma fatto da coefficienti.  $\mathbf{X}$  ed  $\mathbf{Y}$  sono legati dalle seguenti relazioni matriciali:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^T \quad \mathbf{X} = \mathbf{A}^T\mathbf{Y}\mathbf{A}$$

$\mathbf{A}$  è la matrice  $N \times N$  per la trasformata. Ogni coefficiente dato in output dalla trasformata DCT è una combinazione lineare tra un set predeterminato di  $N \times N$  pattern. Quello che si ottiene infine è una nuova matrice nella quale è possibile identificare alcune sezioni significative.

- Il primo valore in alto a sinistra ( $\mathbf{Y}(0,0)$ ) è il coefficiente DC. Esso rappresenta la componente continua (a frequenza nulla), ovvero il colore di base dell'intero blocco.
- La prima riga (escluso il primo valore) è data dai coefficienti AC orizzontali.
- La prima colonna (escluso il primo valore) è data dai coefficienti AC verticali.
- Tutti gli altri valori sono anch'essi coefficienti AC e rappresentano la variazione della componente R, G, B o Y rispetto al coefficiente DC.

Allontanandosi dalla componente DC la componente AC considerata rappresenta una frequenza sempre maggiore.

Nota la differenza tra i coefficienti AC e DC, si può concludere che se il blocco considerato è costituito prevalentemente da un singolo colore allora la matrice avrà pochi coefficienti AC non nulli, mentre accadrà esattamente l'opposto nel caso in cui l'immagine contenga molte sfumature di colore.

## **Quantizzazione**

La quantizzazione consiste nel mappare valori presi da un intervallo continuo in altri valori distribuiti su di un intervallo limitato e discreto. Questo intervallo viene diviso in livelli di ampiezza uguale (quantizzazione uniforme) oppure differente (quantizzazione non uniforme). Di fondamentale importanza è il numero di livelli in cui questo intervallo viene suddiviso e di conseguenza la loro dimensione. La quantizzazione è infatti un processo non invertibile che porta sempre ad una perdita di informazione (compressione lossy), la cui entità dipende esclusivamente dal numero di livelli che vengono considerati. Ciò infatti più di qualsiasi altro parametro determina la qualità visiva che si ottiene come output della codifica.

Nel caso in esame la trasformata DCT fornisce già coefficienti interi, tuttavia è possibile ugualmente ridurre il numero di valori utilizzati, facendo per esempio tendere a zero i coefficienti con un valore poco significativo. Come fatto per YCbCr, per ottenere una buona quantizzazione si possono fare osservazioni riguardanti l'HVS (Human Visual System). L'occhio umano difatti è più sensibile alle basse frequenze rispetto a quelle alte. Motivo per cui si adotta una quantizzazione fine (si usano cioè più bit per campione) per la componente DC e le componenti AC a bassa frequenza, una quantizzazione più grossolana (meno bit per campione) per le componenti AC ad alta frequenza. Questa differenza è indicata nella matrice di quantizzazione con il valore di soglia che aumenta allontanandosi dal valore in alto a sinistra. Dividendo ogni coefficiente della DCT per il corrispondente valore di soglia e arrotondando il risultato si ottiene la nuova matrice dei coefficienti quantizzati.

## **La compattazione**

Quest'ultimo passo porta al passaggio dalla struttura matriciale dei frame ad ottenere una sequenza di bit più corta possibile che li rappresenti. La compattazione (detta anche codifica entropica) è una forma di codifica che permette di rappresentare i messaggi prodotti dalla



sorgente in modo più efficiente, ma senza alcuna perdita di informazione, in modo da poter risalire senza errori dai dati codificati a quelli originari.

## 2.4 Alcune codifiche video

Per la codifica di contenuti video negli ultimi decenni si sono susseguiti numerosi standard.

Essi sono stati proposti da 2 organizzazioni riconosciute a livello mondiale:

- L'ISO/IEC MPEG (Moving Picture Expert Group) è un comitato tecnico congiunto delle organizzazioni internazionali ISO (International Standardization Organization) e IEC (International Electrotechnical Commission) incaricato di definire standard per la rappresentazione e compressione dei segnali audio e video.
- L'ITU-T (international Telecommunication Union-Telecommunication Standardization Sector) è l'agenzia specializzata nelle tecnologie per l'informazione e le comunicazioni.

Parliamo di H.262 / MPEG-2 Video e H.264 / MPEG-4 Advanced Video Coding (AVC), oltre al nuovo H.265 / HEVC.

### MPEG-1

MPEG-1 è uno standard sviluppato da MPEG con lo scopo di comprimere in modo lossy sequenze video e audio. Benchè tecnicamente in grado di codificare video con formati anche superiori all'odierno 4K, esso venne ottimizzato per applicazioni video con basse risoluzioni ed un bit-rate di 1.5 Mbit/s, in modo da ottenere una significativa compressione delle immagini ed una qualità simile a quella delle cassette VHS (Video Home System). Il codec MPEG-1 venne largamente utilizzato nei video CD e inizialmente nei DVD, anche se presto fu sostituito dal più performante MPEG-2.

Dal punto di vista video, il funzionamento della codifica rispecchia quasi fedelmente la procedura descritta precedentemente. L'unica differenza degna di nota è la presenza, scomparsa negli standard video successivi, dei D-frame, oltre ai già conosciuti I, P e B-frame [3]. Essi consistono in immagini indipendenti dalle altre, la cui codifica è di tipo intra-frame e fa uso dei soli coefficienti DC della trasformata, rimuovendo quindi completamente gli AC. Di conseguenza questi frame, tranne in casi molto rari come un'immagine consistente in una superficie di un solo colore, hanno una qualità nettamente inferiore rispetto agli altri.

Vengono perciò utilizzati solo quando viene visualizzata un'anteprima ad alta velocità del video, per esempio durante la ricerca di una determinata scena. Tuttavia nel caso in cui il decoder sia abbastanza veloce, i D-frame possono essere tranquillamente sostituiti dagli I-frame.

Alcuni limiti che hanno portato alla nascita degli standard successivi sono stati certamente la qualità modesta delle immagini, la compressione migliorabile quando si aumenta la risoluzione del video da codificare, la disponibilità del solo profilo colore 4:2:0 e il supporto a due soli canali audio. Non è da meno però un altro fatto, ovvero il mancato supporto per i video interlacciati, oggi in rapido abbandono ma allora molto diffusi. Ciò infatti richiedeva un doppio processo di codifica, con un costo in termini di tempo di computazione non trascurabile.

## **H.262 / MPEG-2**

H.262, conosciuto anche come MPEG-2, è uno standard sviluppato congiuntamente da ITU-T e MPEG all'inizio degli anni '90 ed introdotto ufficialmente nel 1994. Condivide gran parte delle caratteristiche di MPEG-1, migliorandolo però nei suoi punti critici. La qualità video aumenta drasticamente grazie anche all'ottimizzazione per bit-rate maggiori del conservativo 1.5 Mbit/s: si parla in questo caso di 4 - 9 Mbit/s. Aggiunge inoltre il supporto per i video interlacciati e per l'audio multicanale. Date le numerose applicazioni possibili, quali l'impiego nei DVD e nella TV digitale (DVB - Digital Video Broadcasting), ed essendo adattabile sia a video a definizione standard che a quelli in alta definizione, di fondamentale importanza è stata l'introduzione dei profili e dei livelli. I profili consentono di definire un insieme limitato di caratteristiche che l'applicazione può utilizzare. I livelli invece permettono di definire parametri come il bit-rate massimo dei video codificati, il frame rate e la loro dimensione, in modo da limitare la quantità di risorse di sistema richieste durante l'utilizzo del codec. La combinazione tra un profilo ed un livello consente all'utilizzatore del codec di avvalersi di una configurazione già ottimizzata per una determinata applicazione, senza dover impostare tutti i parametri manualmente.

Vista la somiglianza con il predecessore MPEG-1, MPEG-2 ha potuto diffondersi rapidamente in quanto il relativo codec, e di conseguenza il software e l'hardware in cui è implementato, è retrocompatibile con il vecchio standard.

## **H.264 / AVC**

H.264 / Advanced Video Coding (AVC) [1] (noto anche come MPEG-4 Recommendation 10) è uno dei più recenti standard di codifica video sviluppato grazie alla collaborazione tra ITU-T VCEG (Video Coding Experts Group) e ISO/IEC MPEG. Nato alla fine degli anni '90 ma pubblicato ufficialmente nel 2003, è stato esteso negli anni successivi per supportare nuovi formati video, tra cui la risoluzione 4K e HDR. Alcune soluzioni definite in H.264 / AVC sono di implementazione facoltativa, lasciando agli sviluppatori la scelta circa quali utilizzare e quali no. Ciò però può causare problemi di incompatibilità tra diversi encoder e decoder. La soluzione arriva dalla definizione, come già anticipato in MPEG-2, di numerosi profili e livelli, i quali determinano quali soluzioni di codifica sono utilizzate e regolano di conseguenza le applicazioni di destinazione e la complessità computazionale.

Rispetto agli standard precedenti le prestazioni in termini di efficienza sono notevolmente migliorate, dando la possibilità di comprimere una sequenza video con un numero minore di bit a parità di risoluzione e qualità dell'immagine. Questo è dovuto in particolar modo alle modifiche che coinvolgono il sistema di predizione. Se prima un frame poteva essere codificato come di tipo I, P o B, ora esso può essere suddiviso in regioni o gruppi di macro-blocchi 16x16 chiamate slice le quali possono avere codifiche diverse. Un singolo fotogramma può essere quindi composto da slice di tipo I, P, B, SP e SI (gli ultimi due sono switching slices).

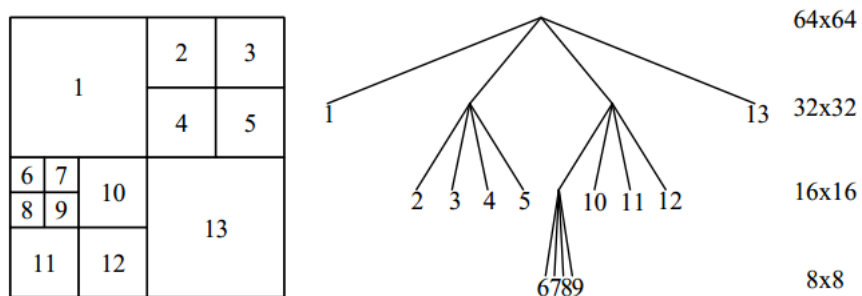
Oltre ad aver sostituito in molte applicazioni il suo predecessore MPEG-2, esso ha avuto molte applicazioni aggiuntive. È stato utilizzato infatti per le trasmissioni televisive principalmente in alta definizione veicolate tramite satellite, cavo e antenna. Inoltre viene usato in sistemi di acquisizione quali videocamere, smartphone e impianti di videosorveglianza.

## **H.265 / HEVC**

H.265 / HEVC [2] è il nuovo standard di codifica video risultato del lavoro congiunto tra ITU-T VCEG e ISO/IEC MPEG, i quali hanno lavorato a questo progetto formando il JCT-VC (Joint Collaborative Team on Video Coding). La prima versione dello standard è datata gennaio 2013, mentre le successive due versioni sono rispettivamente nate nel 2014 e 2015 come estensione del codec originale. La nascita di HEVC non è dovuta solo alla volontà di aumentare la capacità di compressione (il proposito è quello di ottenere una compressione doppia rispetto ad AVC a parità di condizioni), ma anche a quella di consentire lo sviluppo di

nuovi servizi, rendendo possibile la fruizione di contenuti in altissima risoluzione quali 4K e 8K, con una profondità di colore maggiore di quella classica (HDR) e con un livello di dettaglio mai visto prima.

HEVC richiede una complessità computazionale maggiore rispetto ad AVC, tuttavia è stato studiato per sfruttare l'hardware odierno permettendo la parallelizzazione di encoding e decoding. Ciò quindi porta ad annullare lo svantaggio nella maggioranza dei casi (oggi i processori multi-core sono diffusi in ogni tipo di piattaforma, sia essa un computer, uno smartphone o un tablet). HEVC differisce da AVC sotto numerosi aspetti tecnici. Essendo studiato per codificare immagini ad altissima risoluzione, il limite di 16 x 16 pixel come grandezza massima dei blocchi risulta poco conveniente. È molto probabile trovare scene in cui la compressione potrebbe agire su aree più vaste, per esempio 64 x 64 pixel. Ogni immagine viene partizionata in CTB (Coding Tree Block) di forma quadrata, in numero uguale per le componenti di luminanza e crominanza se lo schema di decimazione di quest'ultima è 4:4:4. Se, come è più probabile, il chroma subsampling è 4:2:0, le CTB di crominanza hanno un'area che è 1/4 di quella della componente di luminanza. Ogni CTB ha dimensione L x L dove L viene scelto tra 16, 32 o 64 campioni. L'insieme fatto dalle tre CTB e dagli elementi di sintassi forma l'unità fondamentale CTU (Coding Tree Unit). Ogni CTU è però a sua volta divisibile in più CU (Coding Units) di dimensione variabile,



**Figura 2.2.:** Esempio di partizionamento di una CTU 64x64 in CU con dimensione da 8x8 a 32x32

consistenti in blocchi più piccoli per ogni componente chiamati CB (Coding Blocks). Da ciò deriva la possibilità di vedere un CTU sottoforma di albero quaternario (quad-tree), dove la radice rappresenta il blocco più grande, mentre ogni sottoalbero un CU, il quale può essere a sua volta suddiviso in ulteriori CU. È evidente quindi come la scelta di usare una struttura gerarchica sia l'ideale. Un esempio di questo partizionamento è visibile nella Figura 2.2. La

scelta riguardante il tipo di predizione da effettuare (intra-frame o con compensazione del moto) viene fatta sulle CU.

- Se la predizione scelta è intra-frame, sono ben 35 le modalità tra cui è possibile scegliere: DC, planare oppure confrontando con un blocco raggiungibile spostandosi lungo una delle 33 direzioni possibili (per fare un confronto, in H.264 esse erano solo 8).
- Se invece si è optato per la predizione inter-frame, è possibile un'ulteriore suddivisione delle CB in PB (Prediction Blocks) con dimensione che può arrivare ad un minimo di 4 x 4. Esse costituiscono le PU (Prediction Units). Ogni PB ha uno o più MV (Motion Vectors) associati ed è consentita l'ereditarietà tra essi, ovvero ogni PB può ereditare un MV appartenente ad un altro PB adiacente dal punto di vista spaziale o temporale.

Indipendentemente dalla scelta, le CB vengono ancora suddivise in TB (Transform Blocks) costituenti TU (Transform Units). Su di esse viene effettuata la trasformata DCT (o una derivata dalla DST (Discrete Sine Transform)) ottenendo infine i coefficienti da codificare. Dopo la fase di quantizzazione, per la codifica entropica si usa una versione migliorata di CABAC (Context Adaptive Binary Arithmetic Coding). Essa condivide molto con quella usata in H.264 (Huffman infatti è stato via via abbandonato nelle codifiche video odierne), sebbene le prestazioni siano state migliorate. Per rendere più fedele possibile all'originale l'immagine in fase di ricostruzione diminuendo la quantità di artefatti dovuti alla blocchettizzazione vengono effettuati filtraggi chiamati DBLK (Deblocking, presente anche in H.264) e SAO (Sample Adaptive Offset).

Precedentemente è stato accennato come vantaggio di HEVC lo sfruttamento delle architetture multiprocessore. La parallelizzazione necessaria è data in primo luogo dall'uso delle cosiddette tiles. Una tile è una regione di un'immagine codificabile e decodificabile in modo indipendente dalle altre. Un frame può essere diviso in tiles segmentandola in regioni rettangolari aventi tutte lo stesso numero di CTU. La mancanza di dipendenza tra esse fa in modo che possano essere processate in diversi thread. In alternativa è possibile utilizzare WPP (Wavefront Parallel Processing). In questo caso si fa riferimento alla suddivisione dei frame in slice. Ogni slice viene divisa in righe di CTU ed ognuna di esse viene processata in modo indipendente.

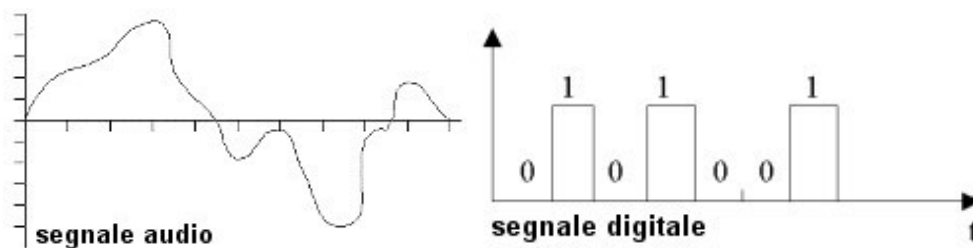
Le principali novità di HEVC che sono state finora sinteticamente presentate danno un'idea del livello di complessità ma allo stesso tempo di flessibilità che questo codec è in grado di offrire.

La complessità dell'encoder è di circa 10 volte superiore a quella di un codificatore AVC mentre la complessità del decoder HEVC è di circa 3 volte superiore a quella AVC.

In conclusione se dal punto di vista della qualità oggettiva il cambiamento sembrerebbe non molto marcato, anche se non trascurabile, notevoli sono invece i progressi in termini di bit-rate richiesto e di dimensione del file di output che può arrivare ad essere fino al 50% inferiore.

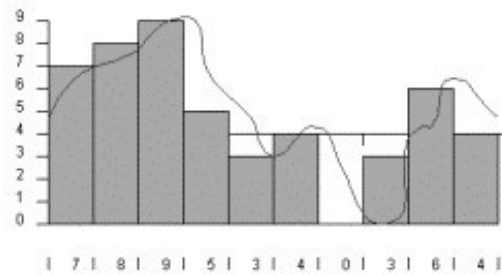
## 2.5 Accenno agli algoritmi di compressione audio

Il segnale audio è per sua natura analogico: varia in modo continuo nel tempo (Fig. 2.3). Il segnale digitale, invece, è un segnale discreto: varia in modo netto nell'unità di tempo, prendendo come valore 1 o 0.



**Figura 2.3:** – La rappresentazione grafica dei due segnali analogico e digitale

Per rappresentare un segnale audio, per sua natura analogico e dunque con un andamento continuo, con un segnale digitale, mantenendo l'informazione, occorre campionare il segnale, ovvero prelevare, ad intervalli regolari, il valore del segnale audio: il risultato sarà quello di approssimare la funzione analogica, che rappresenta le curve del segnale, con una funzione fatta a rettangoli (Fig. 2.4).



**Figura 2.4:** – L'approssimazione del segnale analogico in digitale.

Ovviamente, più aumentano i campioni nell'unità di tempo, più alta sarà la fedeltà nella traduzione da analogico a digitale.

Lo standard PCM (Pulse-Code Modulation), usato per maneggiare un segnale audio in forma digitale non compressa, prevede l'impiego di 44100 campioni al secondo, ovvero 44100 rettangoli al secondo. Il WAV è il formato audio digitale per computer che organizza i dati campionati in modo simile al PCM.

L' aumento progressivo di potenza dei computer ha però favorito negli ultimi decenni lo sviluppo e la diffusione dei sistemi di compressione dei dati audio che permettono una forte riduzione delle dimensioni dei files.

Le codifiche di compressione dell'audio sono numerose ed utilizzano tecniche anche molto differenti l'una dall'altra.

Esistono però tre categorie principali:

- le **codifiche nel dominio del tempo**: si tratta di algoritmi che, elaborano il segnale campionato direttamente senza estrarre le informazioni spettrali(frequenze). L'obiettivo è quello di trovare correlazioni tra i campioni e/o proprietà della sorgente e della destinazione che permettano di ridurre il numero di bit usati per descrivere il valore di un campione audio.
- le **codifiche per modelli**: le codifiche per modelli sono tecniche legate ad una particolare sorgente sonora (come la voce) che si tenta di emulare tramite un modello più o meno semplificato. Le corde vocali e la gola hanno delle ben precise caratteristiche fisiche, il loro comportamento sarà quindi prevedibile sulla base di un modello.
- le **codifiche nel dominio delle frequenze**: Questi algoritmi sono accomunati dal fatto di esaminare e processare il segnale non nel dominio del tempo, ma nel dominio della frequenza. Ogni strumento musicale, ogni suono, e anche la voce, hanno una propria impronta spettrale caratteristica; è, cioè, costituita da una combinazione di frequenze contenute in uno spettro più o

meno ampio. Lavorando su tale spettro si può comprimere il segnale in misura molto maggiore di quanto non si riesca a fare nel dominio del tempo.



# Capitolo 3

## Fasi operative di un processo di post-produzione di un'emittenza televisiva

### 3.1 Utilizzo del formato MXF

MXF (abbreviazione di Material Exchange Format) è un formato sviluppato da Society of Motion Picture and Television Engineers (SMPTE).

Il formato MXF è un contenitore, e ha il compito di “impacchettare” il flusso audio e video processato dai codec e renderlo così fruibile per i software di editing o media player. Tra le altre cose, sono proprio i contenitori a determinare il formato file video “finale”.

È stato creato da giganti del settore radiotelevisivo con l'obiettivo di ottenere un formato di file universale che potesse consentire l'interscambio di contenuti multimediali su diverse piattaforme-server, workstation e simili apparecchiature di editing digitale utilizzate nella produzione televisiva.

File MXF rappresentano contenitori per memorizzare dati video e audio, nonché altri flussi di bit. Essi possono anche includere metadati relativi per la descrizione di contenuti multimediali all'interno di file “.mxf”.

Sono spesso utilizzati per registrare video professionali in alta definizione, con lo scopo di implementarli in sistemi di computer e altri dispositivi.

#### Universalità del formato MXF

Obiettivo principale della standardizzazione del formato MXF è la facilitazione della condivisione di dati video tra varie piattaforme, come fotocamere, lettori video, sistemi informatici, ecc.

MXF è ottimizzato per lo scambio e l'archiviazione di contenuti multimediali. Nella sua struttura basilare un file MXF può contenere soltanto un tipo di media, ma in realtà lo

standard proposto prevede una serie di possibilità cioè dei modelli operativi che ci permettono una maggiore flessibilità del lavoro. Ad esempio è possibile inserire all'interno dello stesso file due o più tracce audio da riprodurre in sincronia, oppure una sequenza di brani audio o parti di video.

Per quanto riguarda la metadattazione del formato MXF vi sono dei dati primari che devono essere presenti in ogni caso all'interno del file ed una serie di dati aggiuntivi opzionali. In base alle sue caratteristiche sarebbe possibile utilizzare il formato MXF in qualsiasi fase di produzione, dall'acquisizione del materiale fino all'eventuale messa in onda di questo, parlando anche ad esempio di notiziari.

Analizziamo le caratteristiche di un file che poi verrà trasmesso:

#### Generale

Nome completo	: C:
\Users\Public\Desktop\FILE ANALIZZATO.mxf	
Formato	: MXF
Nome commerciale	: XDCAM HD422
Versione formato	: 1.2
Dimensione	: 416MiB
Durata	: 58s 160 ms
Bitrate totale	: 60,0 Mb/s
Data codifica	: 2020-05-29 11:07:05.812

#### Video

Formato	: MPEG Video
Nome commerciale	: XDCAM HD422
Impostazioni formato, GOP	: M=3, N=12
Durata	: 58s 160 ms
Modalità bitrate	: Costante
Bitrate	: 50,0 Mb/s
Larghezza	: 1.920 pixel
Altezza	: 1.080 pixel
Rapporto aspetto visualizzazione	: 16:9
Frame rate	: 25,000 FPS
Standard	: PAL
Spazio colore	: YUV
Croma subsampling	: 4:2:2
Profondità bit	: 8 bit
Modo compressione	: Con perdita
Timecode del primo frame	: 00:00:00:00

**Figura 3.1:** esempio di un file analizzato

Utilizzando un software per analizzare le caratteristiche di un file otteniamo una tabella (vedi fig.3.1) dove possiamo individuare tutte le informazioni di cui abbiamo bisogno.

Nella prima parte della fig.3.1 abbiamo le informazioni generali del file.

- **Nome completo: C:\user\public\desktop\FILE ANALIZZATO.mxf**

Ci fornisce le indicazioni della cartella di destinazione e il nome del file.

- **Formato: MXF**

Questo è il contenitore, ci dice il formato del file, da questo possiamo già capire se si tratta di un video, un audio, file di testo, ecc

- **Nome commerciale: XDCAM HD422**

XDCAM è un sistema di videoregistrazione professionale utilizzato dalla Sony, utilizza il profilo 4:2:2 del codec MPEG-2

- **Versione formato: 1.2**

Ogni nuova versione introduce miglioramenti

- **Dimensione: 416MiB**

Spazio che occupa in memoria

- **Durata: 58s 160 ms**

- **Bit-rate totale: 60,0 Mb/s**

è la quantità di informazioni digitali (bit) che è trasferita nell'unità di tempo

- **Data codifica: 2020-05-29 11:07:05.812**

Nel dettaglio il video ha le seguenti caratteristiche:

## **Video**

- **Formato: MPEG Video**

Tipo di codec utilizzato per comprimere il video

- **Nome commerciale: XDCAM HD422**

(visto nella parte precedente)

- **Impostazioni formato, GOP: M=3, N=12**

Vedere fig.2.1

- **Durata: 58s 160 ms**

Durata del filmato

- **Modalità bit-rate: Costante**

Bit-rate costante vuol dire che la codifica di ogni segmento utilizza un numero costante di bit.

- **Bitrate: 50,0 Mb/s**

Come abbiamo già visto il bit-rate è la quantità di informazioni digitali (bit) che è trasferita nell'unità di tempo

- **Larghezza: 1.920 pixel**

indica che l'immagine sul monitor è formata da 1920 colonne

- **Altezza: 1.080 pixel**

indica che l'immagine sul monitor è formata da 1080 righe

- **Rapporto aspetto visualizzazione: 16:9**

Il rapporto tra la larghezza e l'altezza di un'immagine

- **Frame rate: 25,000 FPS**

Rappresenta il numero di fotogrammi trasmessi in un secondo

- **Standard: PAL**

Il **PAL** (Phase Alternating Line) è un metodo di codifica del colore utilizzato nella televisione

- **Spazio colore: YUV**

Lo YUV è uno spazio colore utilizzato per la codifica del video

- **Croma subsampling: 4:2:2**

Sottocampionatura cromaticità. Per ottimizzare la compressione si dedica più banda alla luminanza (Y) e meno alla differenza delle componenti cromatiche (Cb e Cr) perché l'occhio umano è più sensibile alla luminanza. Ogni 4 campioni di Y ce ne sono 2 di Cb e 2 di Cr

- **Profondità bit: 8 bit**

Indica sia il numero di bit usati per indicare il colore di un singolo pixel, sia il numero di bit usati per ogni componente di colore di un singolo pixel.

- **Modo compressione: Con perdita**

utilizzo tecnica di compressione lossy cioè con perdita di informazione

- **Timecode del primo frame: 00:00:00:00**

Si tratta di un segnale di ampio utilizzo in diversi ambiti per la sincronizzazione di segnali e per la scalettatura del materiale registrato su supporti audio/video.

Poi oltre al video vengono analizzate tutte le 8 tracce audio comprese nel file

Tutto ciò è di fondamentale importanza per un'emittente televisiva, appunto perché deve gestire e trasmettere immagini e a questo si deve aggiungere il fatto che oggi si punta molto sulla qualità delle immagini e quindi è molto importante che dalla loro origine fino alla visione finale i file mantengano la qualità desiderata. Quindi nella catena produttiva televisiva da un lato bisogna gestire i file in maniera da non perdere la qualità e dall'altro poterli utilizzare e condividere in maniera rapida. Un telegiornale, occupandosi di notizie, ha tempi di produzione molto ristretti e questo implica che tutto il sistema sia estremamente affidabile per non avere perdite di tempo prezioso.

Naturalmente con l'avvento del digitale tutta la gestione e l'archiviazione è affidata ai computer che permettono le varie fasi del processo produttivo.

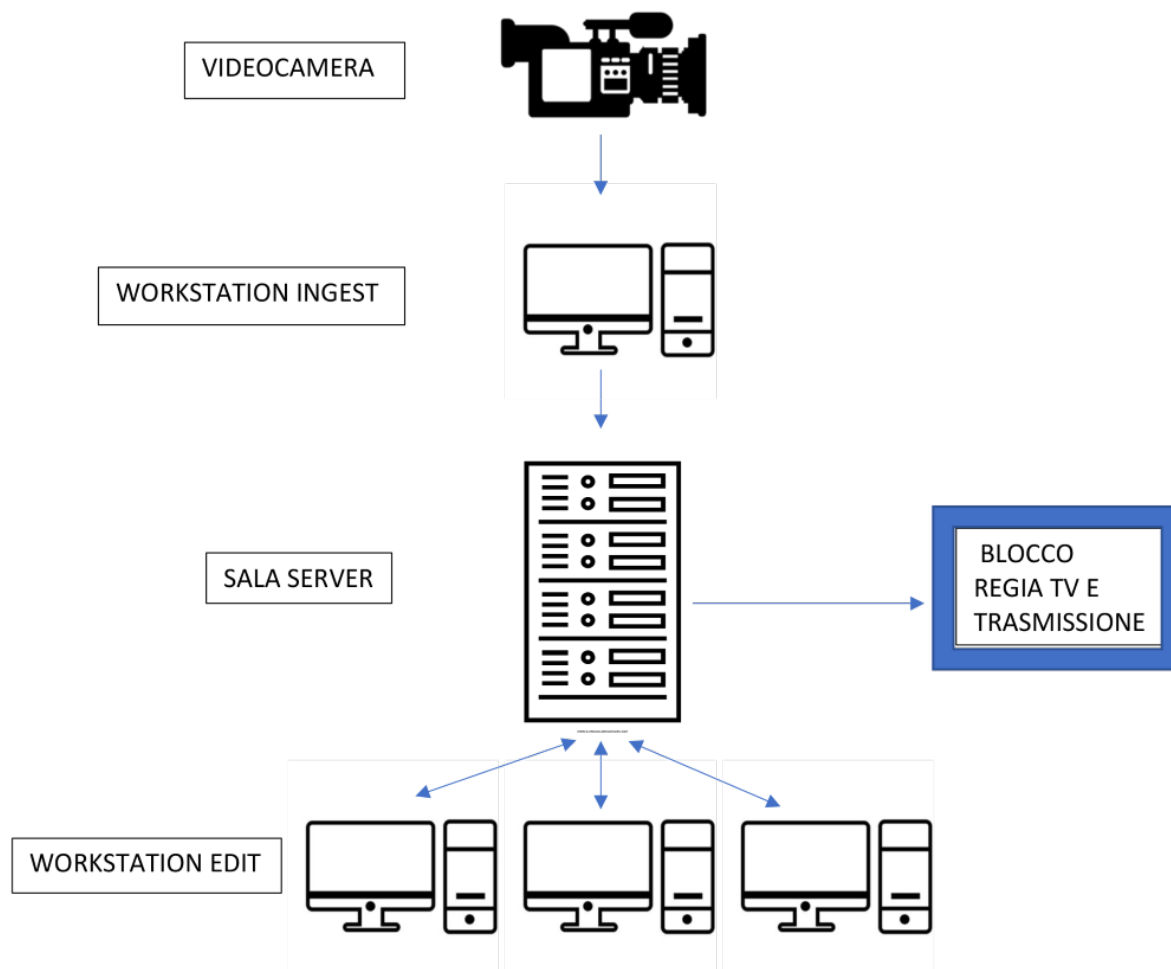
Da un punto di vista tecnico il montaggio (fase di editing) è il momento in cui le varie riprese, dette “girato”, vengono messe in sequenza ovvero secondo la consecuzione temporale prevista così da dare luogo al racconto delle notizie. La fase di editing inizia con la visione analitica del girato, la scelta delle inquadrature e la loro disposizione in sequenza.

Visto che il montaggio video permette di modificare delle sequenze di ripresa video e di aggiungere effetti e tracce sonore gli attuali software permettono di visualizzare già subito le immagini mentre vengono trasferite su un computer o server. Questo perché oltre al file originale viene creato un altro file detto proxy (file di bassa qualità utilizzato solo nella fase di editing) che rimane associato al file originale, e consente di visualizzare un video utilizzando meno risorse hardware a discapito della risoluzione.

Esistono molti software di montaggio video sia a livello professionale che amatoriale. A livello generale permettono tutti di creare un video ma ognuno ha le proprie caratteristiche che lo rendono differente dagli altri. La prima differenza è proprio l’interfaccia grafica, infatti ogni software ha una propria organizzazione dello spazio di lavoro e delle varie cartelle. Un'altra caratteristica è la quantità di funzioni aggiuntive che rendono un software basilare o completo. Il montaggio video professionale deve coprire un numero sempre maggiore di funzioni per soddisfare le diverse esigenze e quindi richiede maggiori risorse anche a livello hardware.

Parte fondamentale è la compatibilità tra il software di montaggio e i file (girati) prodotti dalla videocamera.

I passaggi che il file compie, dalla sua creazione fino alla trasmissione finale, in un'emittenza televisiva sono i seguenti:



**Figura 3.2:** Schema dei passaggi che compie un file (dall' origine alla trasmissione)

**VIDEOCAMERA:** La videocamera è costituita da tre elementi fondamentali: la lente, il sensore e il processore d'immagine. A questi si aggiunge il registratore da cui viene registrato il segnale. L'immagine individuata passa dalla lente (ottica) e viene catturata dal sensore, che in base alle proprie dimensioni la trasmette più o meno ricca di dettagli al processore. Questi la elabora, definisce, adatta e allinea alle dimensioni e caratteristiche volute o imposte definendo la risoluzione del video. Anche le telecamere negli ultimi anni hanno avuto una notevole evoluzione e permettono di essere impostate in base alle proprie esigenze (scegliendo anche tra più formati di registrazione).

WORKSTATION di INGEST: è una postazione periferica naturalmente collegata alla sala server ed è dedicata ad ingestare (acquisire) i contenuti nel sistema grazie alla presenza di lettori di scheda e apparati per l'acquisizione dei vari supporti di registrazione.

SALA SERVER: naturalmente è il cuore di tutto il sistema perché gestisce tutte le operazioni e le funzioni.

E' composta da:

- Server per l'ingest e il playout dei contenuti
- Server per la gestione degli accessi, ricerche e visualizzazioni (sono molti gli utenti che utilizzano contemporaneamente il sistema)
- Server che realizzano i render
- Server per la transcodifica dei file provenienti da fonti esterne
- Server per l'indicizzazione e gestione delle risorse
- Sistema di storage (dischi di archiviazione)

Tutti i file sia quelli registrati dalla videocamera che quelli lavorati in post-produzione rimangono nel sistema principale che li mette a disposizione delle varie postazioni (workstation e regia TV) così da poter essere utilizzati o trasmessi.

WORKSTATION di EDITING: è una postazione periferica sempre collegata alla sala server ed è dedicata al montaggio, ricerca e metadattazione. In base alle necessità di produzione si possono avere diverse workstation che lavorano contemporaneamente.

Quindi, a livello tecnico, un emittente televisiva è organizzata in questo modo.

Un operatore dotato di una telecamera si occuperà di fare le riprese e quindi di registrare le immagini. In questo caso vengono usate telecamere digitali che salvano i file in formato “.mxf” come visto in fig.3.1. Una volta terminata questa fase l'operatore consegnerà il supporto di registrazione, di solito una scheda SXS (memoria flash creata da Sony), al tecnico che inizierà la fase di INGEST (tramite una workstation ci sarà il trasferimento dei file nel sistema centrale). Oltre all'acquisizione delle immagini tramite scheda ci sono anche altre fonti di ingresso come le piattaforme web e tutti i vecchi supporti di registrazione analogica dove comunque in ogni caso il sistema converte e salva il file in formato MXF.

Mentre le immagini vengono salvate nel sistema di archiviazione viene fatta una prima metadattazione descrivendo sommariamente l'argomento. Finita la fase di ingest si passa alla fase dedicata all'editing, cioè al montaggio del filmato, qui vengono gestiti contributi sonori,



video e altro per realizzare il contributo che verrà utilizzato durante la trasmissione. Ultimata anche questa fase il lavoro è completato e si passa al salvataggio del video finale. Visto che si è lavorato in bassa qualità nel momento conclusivo il sistema ricollegherà ai file proxy quelli in alta qualità creando il file finale pronto per essere trasmesso. Poi attraverso apparecchiature dedicate, situate in regia TV, potremo trasmettere il flusso video in uscita.

## **3.2 Trasmissione DVB-T e DVB-T2**

Il passaggio dalla trasmissione analogica a quella digitale ha portato numerosi vantaggi, come l'ampliamento dei contenuti, più qualità e una maggiore copertura del segnale sul territorio.

Il DVB (Digital Video Broadcasting) è uno standard tecnico per la trasmissione televisiva digitale terrestre. È stato proprio il passaggio al digitale terrestre "DVB-T" che ha permesso la trasmissione dei primi programmi in HD (High Definition) perché sfrutta le compressioni audio/video del codec H264 AVC per i canali HD, insieme al sistema di modulazione delle frequenze OFDM.

Nei prossimi anni il successore, il DVB-T2 (Digital Video Broadcasting - Second Generation Terrestrial) ha lo scopo da un lato di migliorare la ricezione con apparati fissi e portatili, dall'altro di far crescere il bit-rate rispetto all'attuale DVB-T. Il DVB-T2 in sinergia con il codec HEVC permetterà di godere di una maggiore qualità video, rendere più efficiente l'uso delle frequenze a disposizione e ridurre il rischio interferenze. Il miglioramento prestazionale passa da 24 Mbit/sec attuali a 33-35 Mbit/sec del DVB-T2, che insieme a una compressione video migliore consentirà ai media televisivi di trasmettere più canali con maggiore qualità usando le stesse risorse. Infatti il nuovo compressore HEVC erede dell'H.264 AVC, migliora la qualità video, raddoppia il rapporto della compressione dei dati e supporta l'ultra definizione 8K con risoluzioni fino a 8192 x 4320 pixel.

Con il passaggio al codec HEVC (H.265) si potranno quindi ospitare fino a 11 canali HD (con risoluzione a 720p) o fino a 5 Full HD (con risoluzione a 1080p) sullo stesso multiplex (MUX).

Attualmente invece con lo standard DVB-T con codec MPEG-2 avendo a disposizione una capacità netta di banda di circa 24 Mb/s per multiplex, possiamo trasmettere 5-6 canali in SDTV o 2 canali HDTV 720 o 1080p.

# Conclusione

Negli ultimi decenni abbiamo assistito ad un'evoluzione tecnologica senza precedenti in svariati campi, sia per entità del cambiamento, sia per rapidità. Quest'ultimo fattore è stato il più incisivo, differenziando il panorama tecnologico in modo netto rispetto al passato. In poco tempo siamo passati da un mondo multimediale prettamente analogico ad un altro in cui tutti i contenuti sono disponibili in formato digitale. Grazie alla digitalizzazione ed alla nascita di Internet abbiamo la possibilità di accedere in ogni momento ad una quantità infinita di informazioni, sia testuali che audiovisive. A differenza di venti anni fa, oggi è comune la condivisione di immagini e video. Hanno certamente contribuito a questo la nascita dei social network, dei siti di video sharing e la diffusione di dispositivi sempre più versatili e compatti come smartphone e tablet, i quali hanno cambiato le modalità di fruizione dei contenuti. Negli ultimi anni è aumentata la richiesta di maggior qualità visiva, il tutto combinato con l'esigenza di diminuire la quantità di bit da memorizzare o trasmettere. Ed è proprio per venire incontro a queste crescenti necessità che è stato di fondamentale importanza lo sviluppo di nuove e sempre più efficienti tecniche di rappresentazione dei contenuti multimediali e il futuro sarà ancora basato su questo.



# Bibliografia

[1] <http://www.crit.rai.it/eletel/> n. 01/2014

[2] Vivienne Sze, Madhukar Budagavi, Gary J. Sullivan, "High Efficiency Video Coding (HEVC): Algorithms and Architectures", Springer International Publishing, 2014

[3] Keith Jack, "Video Demystified: A Handbook for the Digital Engineer", 4 ed., Elsevier, 2005