

# Che cosa è, come funziona: Uno standard pervasivo (MPEG-2 video)

Parte I

ing. Marzio Barbero e  
ing. Natasha Shpuza

## 1. Premessa

Dal momento in cui è stato normalizzato, il sistema MPEG-2 ha avuto una rapida diffusione ed è attualmente alla base della maggior parte dei sistemi, a livello mondiale, di diffusione digitale dell'informazione video.

Questa prima parte descrive gli algoritmi alla base della codifica video. La seconda parte, pubblicata nel prossimo numero, approfondirà aspetti legati alle applicazioni e alla qualità dei sistemi video compressi. Prossime schede avranno come oggetto il sistema MPEG-2 nel suo insieme e la codifica dell'audio.

## 2. Cenni storici

MPEG<sup>Nota 1</sup> nasce nel gennaio 1988 come Experts Group dell'ISO/IEC JTC 1/SC 2/WG 8 e diventa SC 29/WG 11 nel novembre 1989.

All'inizio vennero definite tre fasi di lavoro, caratterizzate dal bit-rate complessivo necessario per la codifica di immagini in movimento per la memorizzazione digitale (DSM, *Digital Storage Media*) a 1-1,5 Mbit/s, a 1,5-5 Mbit/s e 5-60 Mbit/s.

La prima fase intendeva definire uno standard utilizzabile per la memorizzazione di informazioni audiovisive sul disco ottico allora disponibile il CD (il Compact Disc è caratterizzato da una capacità di 680 MB ed in quegli anni

i lettori erano in grado di trasferire un flusso di dati appunto dell'ordine di 1,5 Mbit/s). Lo standard ISO/IEC 11172, noto come MPEG-1, ha quindi il titolo "Codifica di immagini in movimento e audio associato per media di memorizzazione digitale operanti fino a circa 1,5 Mbit/s", ma il riferimento al bit-rate è contenuto solamente nel titolo, infatti lo scopo era di definire uno "standard generico", in cui fosse specificato il minimo necessario, in pratica la sintassi e la struttura dei dati memorizzati o trasmessi, per consentire la decodifica dell'informazione audiovisiva, e lasciare ampia possibilità all'industria di ottimizzare i codificatori e i decodificatori.

La codifica video MPEG-1 è specificata nella parte 2 (11172-2:1993) ed basata sull'uso della DCT e della compensazione del movimento, analogamente allo standard ITU H.261 pubblicato nel 1990 per la videotelefonía e videoconferenza.

La seconda fase intendeva definire, malgrado l'obiettivo indicato nel titolo fosse limitato ai mezzi di memorizzazione digitale, una codifica adatta per le applicazioni diffuse (radiodiffusione e distribuzione via cavo) e dette origine allo standard ISO/IEC 13818, noto come MPEG-2. La prima sessione di MPEG-2 ebbe luogo nel luglio 1990, quando erano ancora in corso i lavori di definizione di MPEG-1. Nel 1992 fu eliminato il limite in termini di bit-rate, che intanto era passato da 5 Mbit/s a 10 Mbit/s, e la terza fase, inizialmente prevista per consentire la codifica di immagini in alta definizione (HDTV), conflui

Nota 1 - MPEG è l'acronimo di Motion Picture Expert Group, ovvero il WG 11 (Working Group 11, "Coding of Moving Pictures and Audio") dello SC 29 (Subcommittee 29, "Coding of audio, picture, multimedia and hypermedia information") del comitato congiunto JTC 1 (Joint Technical Committee 1, "Information Technology") dell'ISO e IEC (International Organisation for Standardisation e International Electrotechnical Commission). Il *Convenor* del gruppo è stato, fin dalla sua nascita Leonardo Chiariglione, dello CSELT (oggi TelecomItaliaLab), di Torino.

nella seconda fase (ciò spiega la mancanza di uno standard MPEG-3). Lo standard MPEG-2 video è del 1995 e coincide con lo standard ITU-T H.262.

Alla fine del 1996 MPEG-2 video fu scelto come base per il sistema televisivo digitale terrestre da introdurre in USA (ATV, *Advanced TV*). Fu scelto successivamente per la diffusione digitale da satellite, sempre in USA, per il progetto Direct TV. L'iniziativa europea per la diffusione digitale (DVB, *Digital Video Broadcasting*) diventò il maggior utilizzatore mondiale di MPEG. Seguì il DVD Forum e infine anche in Giappone MPEG-2 fu introdotto per la diffusione digitale sia di immagini in definizione standard (SDTV) che in alta definizione (HDTV).

Nel gennaio 1996 è stato definito il sistema di codifica per applicazioni professionali, in grado di operare sul video conforme alla Rac. ITU-R BT.601 senza apprezzabili degradamenti (*422 Profile*).

### 3. Toolkit, Profile, Level

Nella scheda precedente si è introdotto il termine *toolkit*, ovvero l'insieme degli utensili che vengono adottati per ridurre la ridondanza presente nell'informazione video e per eliminare eventuale informazione senza introdurre difetti rilevanti sull'immagine.

MPEG-2 adotta tutti i *tool* descritti: la trasformata coseno discreto (DCT) per sfruttare la correlazione spaziale, la codifica delle differenze fra quadri adiacenti e la compensazione del movimento per sfruttare la correlazione temporale, i codici a lunghezza variabile (VLC) per ridurre la ridondanza statistica, la matrice di visibilità e la quantizzazione dei coefficienti DCT per eliminare l'informazione meno rilevante.

Un sottoinsieme degli utensili, o tutti, concorrono alla definizione di un *profile*, un profilo specifico per una tipologia di applicazioni.

Nell'ambito di un singolo *profile*, possono

essere definiti uno o più *level*, livelli di definizione dell'immagine, in termini di campioni e righe attive.

I concetti di *tool*, *profile* e *level* vengono adottati in MPEG-2 per consentire di comprendere in un unico standard generico i requisiti individuati durante il processo di normalizzazione.

In tabella 1 sono riassunte le caratteristiche principali dei profili e livelli previsti da MPEG-2. Quello più diffuso è il MP@ML, ovvero il profilo principale (*Main Profile*) al (@) livello principale (*Main Level*), è alla base dei sistemi DVB e DVD. Il 422P@ML è usato in produzione e postproduzione.

#### 3.1 Main Profile

Il MP prevede più livelli, corrispondenti a dimensioni massime delle *picture*, si va dall'alta definizione, formato studio, al formato SIF, quello adottato in MPEG-1.

È comunque il livello principale, cioè MP@ML quello più diffuso, in grado di codificare immagini basate sulla Rac. ITU-R BT.601, caratterizzate da un sottocampionamento della crominanza in verticale (convenzionalmente indicata come 4:2:0). Il decoder deve essere in grado di decodificare flussi con bit-rate fino ad un massimo di 15 Mbit/s.

#### 3.2 4:2:2 Profile

È stato definito per applicazioni professionali: ad esempio per la videoregistrazione in studio o per la trasmissione su rete di contribuzione o di distribuzione primaria. Opera quindi su un segnale conforme alla Rac. ITU BT.601 ed è caratterizzato da un bit-rate massimo (50 Mbit/s) che consente una qualità molto elevata (quasi la trasparenza) del segnale anche nel caso di più co-decodifiche in cascata.

#### 3.1 Simple Profile

Il SP è praticamente identico al MP@ML tranne per il fatto di non prevedere l'uso di *B-picture*, la cui funzione è spiegata nel seguito, e di conseguenza limita l'uso di memorie di quadro video e riduce i costi realizzativi.

<b>profili e livelli</b>	<b>SIMPLE</b> 4:2:0	<b>MAIN</b> 4:2:0	<b>422P</b> 4:2:2 e 4:2:0	<b>SNR</b> 2 layer 4:2:0	<b>SPATIAL</b> 3 layer 4:2:0	<b>HIGH</b> 3 layer 4:2:2 e 4:2:0
<b>HIGH</b> 60 frame/s		1920x1088 80 Mbit/s I,P,B	1920x1088 300 Mbit/s I,P,B			1920x1152 100 Mbit/s I,P,B
<b>HIGH-1440</b> 60 frame/s		1440x1152 80 Mbit/s I,P,B			1440x1152 60 Mbit/s I,P,B	1440x1152 80 Mbit/s I,P,B
<b>MAIN</b> 30 frame/s	720x576 15 Mbit/s I,P	720x576 15 Mbit/s I,P,B	720x608 50 Mbit/s I,P,B	720x576 15 Mbit/s I,P,B		720x576 20 Mbit/s I,P,B
<b>LOW</b> 30 frame/s		352x288 4 Mbit/s I,P,B		352x288 4 Mbit/s I,P,B		

Tab. 1 - Tabella riassuntiva dei principali parametri che caratterizzano profili e livelli previsti dallo standard MPEG-2.

### 3.3 SNR Scalability Profile

I sistemi analogici sono caratterizzati da quello che in inglese è indicata come *graceful degradation*, ovvero la qualità dell'immagine e del servizio degradano progressivamente al peggioramento delle condizioni del canale. Nei sistemi digitali di trasmissione del video, invece, la qualità è praticamente costante, indipendente dalle condizioni del canale, fino a quando le tecniche di protezione e mitigazione degli errori non riescono più a far fronte alla crescita della probabilità di errore e, in modo brusco, si passa in condizioni di mancanza del servizio.

Nel caso del profilo SNR si vuole ottenere un comportamento simile alla *graceful degradation* e i dati sono assegnati a due strati (*layer*) del flusso informativo. Il *layer* base contiene l'informazione indispensabile ad una decodifica completa dell'immagine, anche se quantizzata in modo più grossolano e quindi caratterizzata da un rapporto segnale/

rumore inferiore; in genere è trasportato da un canale più protetto, caratterizzato da una QoS (*Quality of Service*) più elevata, garantendo quindi il livello minimo di qualità. L'informazione aggiuntiva del secondo *layer* consente, in condizioni ottimali del canale, di decodificare l'immagine alla massima qualità.

### 3.4 Spatial e Temporal Scalability

In un sistema scalabile spazialmente il flusso di dati è diviso in un *layer* base sufficiente a decodificare immagini a più bassa risoluzione (ad esempio SDTV, a definizione convenzionale) e in un ulteriore strato, che contiene l'informazione aggiuntiva che consente di ricostruire l'immagine originaria, a definizione più elevata (quella associata al livello HDTV-1440, cioè caratterizzata da una risoluzione doppia sia in orizzontale che in verticale).

Nel caso di scalabilità temporale un *layer* è relativo ad una ripetizione di immagine pari a 25 o 30 Hz, mentre un ulteriore *layer* contie-

ne le informazioni necessarie a riprodurre le immagini a 50 o 60 Hz.

### 3.5 MultiView Profile

Il MVP è stato approvato nel luglio 1996 per la codifica in modo efficiente di sequenze video ottenute da due telecamere che riprendono la stessa scena. In questo caso il *layer* base contiene una delle immagini ed è utilizzato per fornire la predizione ad un secondo *layer*, in modo da consentire la riproduzione di un video stereoscopico, su apposito ricevitore dotato di display stereoscopico.

## 4. Organizzazione dell'informazione video

### 4.1 Formati d'immagine, a scansione progressiva o interlacciata

Lo standard H.261, sviluppato per la videoconferenza su linee ISDN, era in grado di codificare immagini in formato CIF (*Common Interface Format*, 352 pixel x 288 righe a 30 quadri al secondo), un compromesso derivato dai due formati interlacciati previsti dalla Rac. ITU-R BT.601. Il compromesso, e la conseguente complicazione dei terminali, rese difficoltosa

la diffusione del sistema.

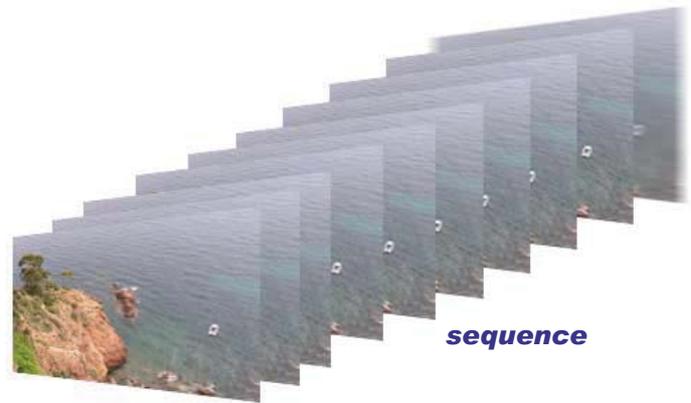
MPEG-1 introdusse il quindi il formato SIF (*Standard Image Format*) che è ricavato sottocampionando sia orizzontalmente che verticalmente i due formati previsti dalla ITU-R BT.601 e ottenendo in un caso immagini progressive da 352 pixel per 288 righe a 25 quadri al secondo e nell'altro caso immagini da 352 pixel x 240 righe e 30 quadri al secondo. Il sottocampionamento verticale consiste semplicemente nella non codifica di uno dei semiquadri. In questo modo, se si opera nell'ambito di uno dei formati, telecamere e display sono di facile reperimento e il codec deve effettuare sottocampionamenti e interpolazioni non particolarmente complessi.

MPEG-2 MP@ML è in grado di codificare immagini a piena risoluzione (720 pixel x 576 righe per 25 quadri al secondo, oppure 720x480 per 30 quadri al secondo), sia nel caso di scansione progressiva che nel caso di scansione interlacciata. Nel caso di scansione interlacciata i due semiquadri (*field*) possono essere codificati contemporaneamente, come una singola immagine o *picture* (modalità *frame picture*) oppure possono essere codificati come due *picture* distinte (modalità *field picture*). Entrambe le modalità possono essere utilizzate all'interno della stessa sequenza.

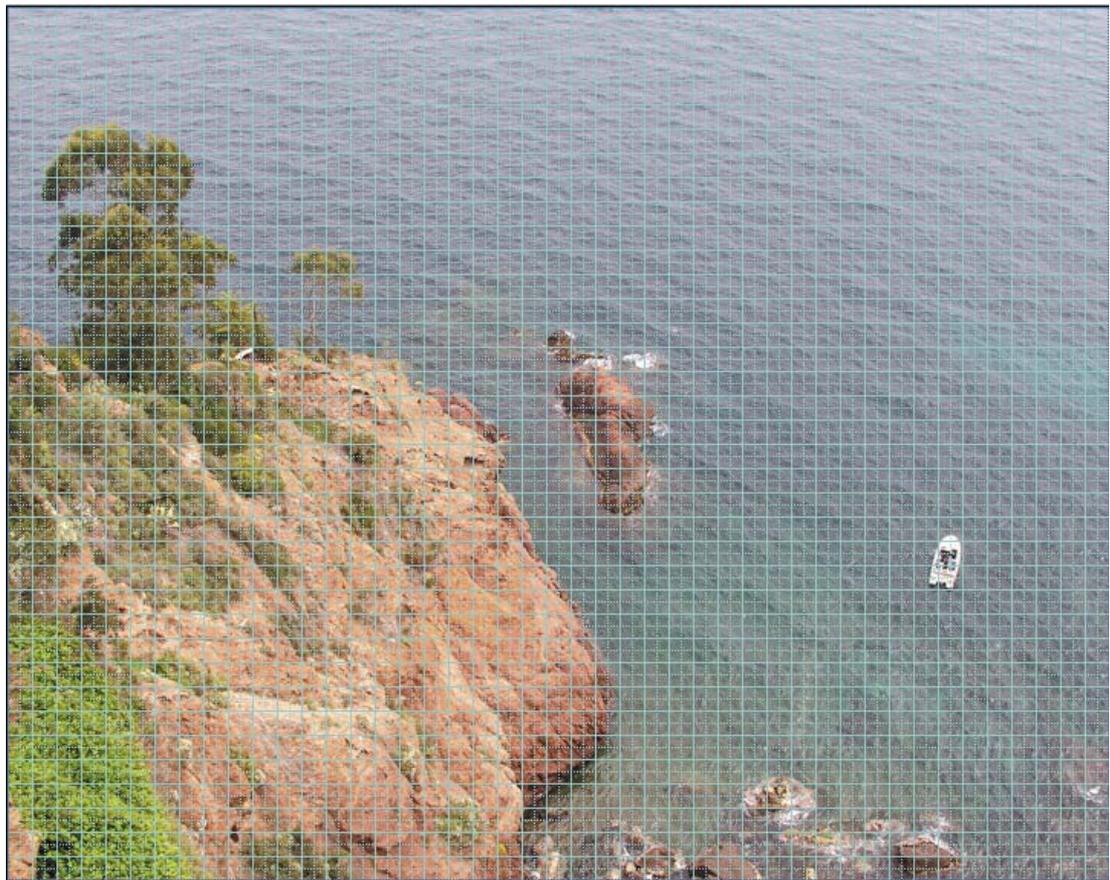
Fig. 1 - Nella struttura di campionamento denominata 4:2:0 i campioni di cromaticanza  $C_R$  e  $C_B$  sono ottenuti mediante filtraggio e sottocampionamento a partire dal 4:2:2 e, a causa della struttura interlacciata, non sono in posizione equidistante rispetto ai campioni d'origine. Nella terminologia MPEG i due semiquadri video sono denominati *top field* e *bottom field*.



Fig. 2- Una *sequence* è una successione di *picture* video. Nel caso del profilo principale a livello principale MP@ML, ogni *picture* è costituita da un massimo di 720 campioni (*sample*) di luminanza per ciascuno delle 576 righe video (valore massimo). La *picture* è organizzata in fette (*slice*) costituiti da *macroblock* contigui. Il macroblocco corrisponde alla porzione di immagine a cui può venire applicata la compensazione del movimento ed è composto da 16x16 elementi di immagine. E' a sua volta formato da 4 blocchi (*block*) costituiti da 8x8 campioni (*sample*) di luminanza, mentre per quanto riguarda le due componenti di cromaticanza, queste sono sottocampionate sia in orizzontale che in verticale (struttura 4:2:0): a 16x16 campioni di luminanza corrispondono quindi due blocchi 8x8 di cromaticanza (uno per ciascuna delle due componenti  $C_R$  e  $C_B$ ).



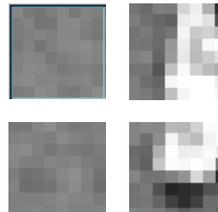
**picture**



**slice**



**macroblock**



**block**

**sample**



## 4.2 Campionamento 4:2:0

La Rac. ITU-R BT.601 prevede le due strutture di campionamento 4:4:4 e 4:2:2, e il 422 *Profile* dello standard MPEG-2 consente di operare direttamente sui campioni della struttura 4:2:2.

Per gli altri profili si prevede invece un ulteriore sottocampionamento delle due componenti di cromaticità, riducendone la risoluzione verticale: questa struttura di campionamento viene convenzionalmente indicata come 4:2:0 (figura 1). In questo modo si ha una riduzione del 25% del numero di campioni su cui effettuare la compressione, ma si introduce un degradamento che è particolarmente critico nel caso di co-decodifiche successive, nel caso dei formati interlacciati. Tali profili sono quindi adatti ad applicazioni relative alla distribuzione all'utente finale.

## 4.3 Dalla Sequence al Sample

Il video in ingresso ad un codificatore MPEG è considerato una struttura organizzata secondo quanto indicato in figura 2.

La sequenza è una successione di *picture* video. Ciascuna *picture* è suddivisa in *slice*, fette costituite da un numero arbitrario di *macroblock* consecutivi, i macroblocchi sono raggruppamenti di 16 elementi di immagine

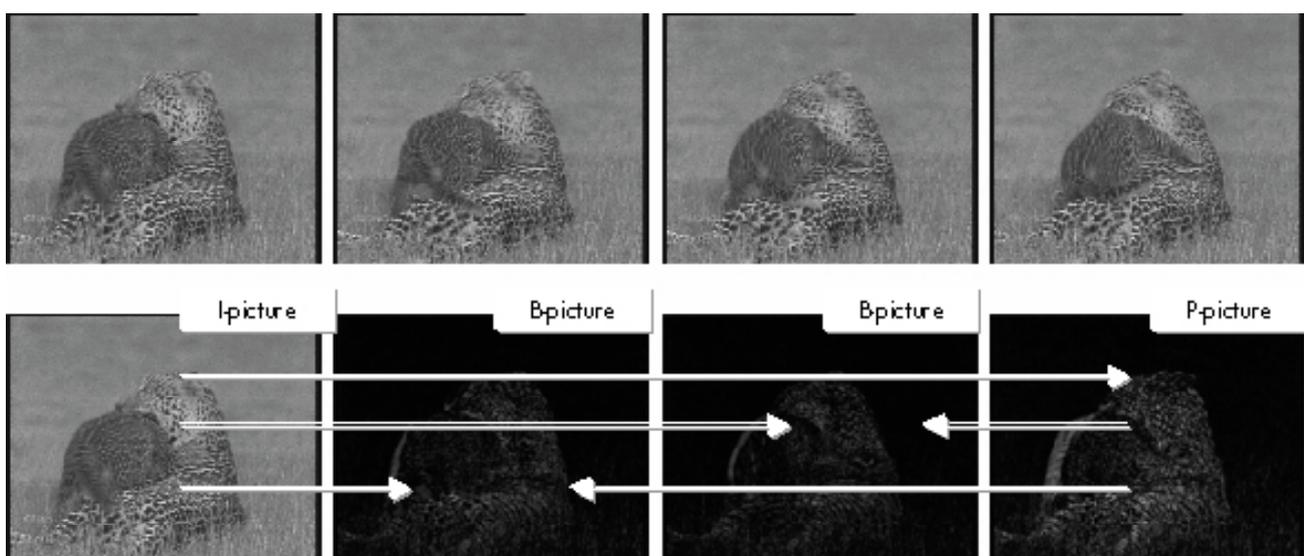
in orizzontale per 16 elementi in verticale. Ogni macroblocco è a sua volta costituito da 4 *block*, blocchi di 8 per 8 *sample*, campioni di luminanza, più 2 blocchi 8x8 per ciascuna delle due componenti di cromaticità (nei formati 4:2:0).

## 4.2 Picture di tipo I, P e B: il GOP

La DCT viene applicata ai singoli blocchi 8x8. Se il macroblocco è codificato in modo I (*Intra*), cioè sfruttando la sola ridondanza spaziale, la DCT viene applicata direttamente ai campioni, di luminanza o di cromaticità, corrispondenti ai 6 blocchi che costituiscono il macroblocco. Se l'intera *picture* è codificata sfruttando esclusivamente la correlazione spaziale, viene denominata *I-picture* e tutti i macroblocchi sono di tipo I.

Se si sfrutta la ridondanza temporale, i coefficienti della DCT sono calcolati sulle differenze fra i campioni e i valori di predizione ottenuti a partire dai campioni di una *picture* precedente, eventualmente utilizzando anche la compensazione del movimento. Le immagini che contengono macroblocchi codificati sfruttando la predizione temporale sono denominate *P-picture* (*P*=*Predicted*). Una *P-picture* contiene normalmente anche dei blocchi di tipo intra, se non è stata individuata una buona predizione

Fig. 3 - In questo esempio quattro *picture* sono codificate, la prima come *I-picture*, la quarta come *P-picture* e le due intermedie come *B-picture*. Le differenze fra i valori dei campioni e le predizioni (le frecce indicano da quale *picture* sono ottenute le predizioni) sono inferiori nel caso delle *B-picture*, con un conseguente risparmio in termini di bit-rate.



temporale per tali porzioni d'immagine.

MPEG prevede un'ulteriore tipologia di immagini, le *B-picture*, per i macroblocchi delle quali i valori di predizione sono ottenuti a partire da *I-picture* o *P-picture* che precedono e seguono (*B=Bidirectionally predicted*) quella attuale.

Poiché non risulta facile predire il futuro, per poter codificare i macroblocchi delle *B-picture* occorre avere delle memorie di quadro in modo che, con un opportuno ritardo, il codificatore abbia contemporaneamente a disposizione sia la *picture* di tipo I o P che precede che la *picture* di tipo I o P che segue (figura 3). Periodicamente è forzata la codifica di una *I-picture* in modo che il decodificatore al momento dell'accensione o a seguito di errori sul canale di trasmissione, possa recuperare la corretta decodifica della sequenza.

Una *I-Picture* è la prima di un gruppo di immagini denominato GOP (*Group of Pictures*). La lunghezza del GOP non è specificata da MPEG. Nel caso in cui non si utilizzi la correlazione temporale, le *picture* sono tutte di tipo I e la lunghezza del GOP è pari a 1. Si possono costruire strutture di GOP composte solo dalla *I-picture* seguita da una o più *P-picture*. Se vi sono *B-picture*, la struttura tipica è quella

denominata IBBP, cioè vi è una coppia di *B-picture* preceduta e seguita da *I*- o *P-picture*.

In genere una *I-Picture* richiede, per la codifica, una quantità di dati superiore rispetto a quelle predette, per cui si ha un risparmio in termini di bit-rate per GOP lunghi, comunque tipicamente la lunghezza del GOP non supera 12 o 15, ma questo valore non è necessariamente fisso e non è indicato da nessun parametro all'interno del flusso binario: la scelta è operata dal codificatore e il decodificatore si adatta di conseguenza.

## 5. La codifica

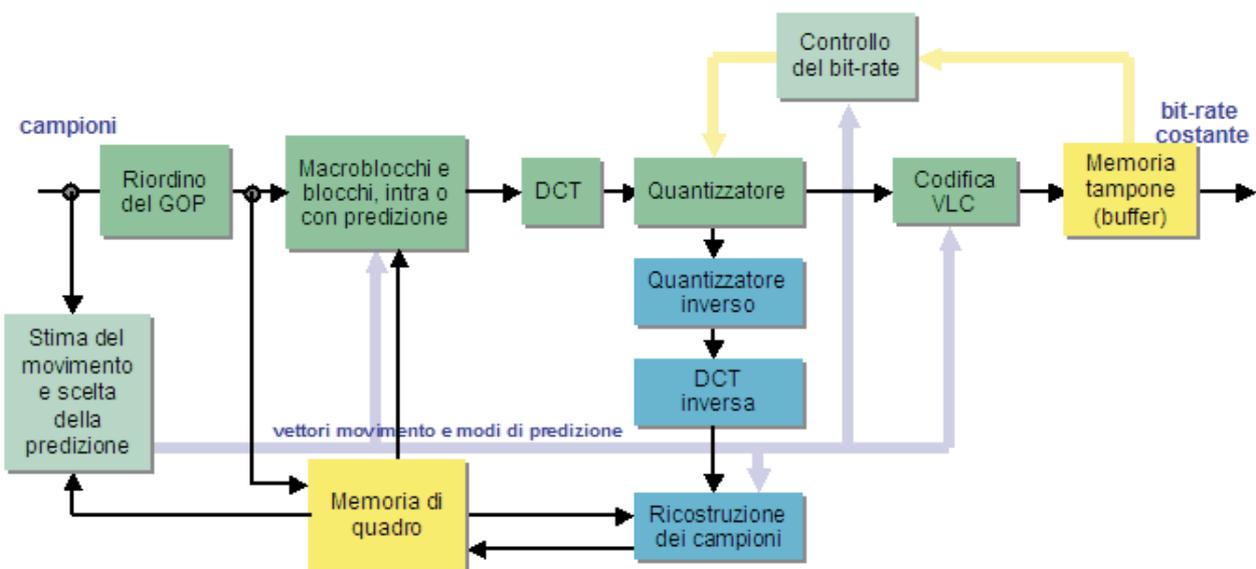
### 5.1 Schema del codificatore

Lo schema di un codificatore MPEG è riportato in figura 4. Il processo di codifica segue lo schema indicato in figura 5.

Il video viene organizzato in funzione della struttura a GOP prescelta.

Nel caso in cui la *picture* è di tipo P o B, viene determinato il macroblocco di predizione ottimale, sfruttando la compensazione del

Fig. 4 - Schema a blocchi di un codificatore MPEG.



movimento.

A tutti i blocchi 8x8 (di luminanza e cromaticità) viene applicata la DCT. Nel caso dei macroblocchi di tipo I la precisione dei coefficienti è fino a 10 bit nel caso del MP@ML, mentre nel caso del 422P arriva a 11 bit.

Nel caso in cui i macroblocchi siano di tipo P o B, la DCT è applicata sulle differenze fra i campioni video attuali e quelli predetti a partire dai campioni contenuti nelle memorie di quadro. Nelle memorie di quadro sono memorizzati i valori delle *picture* ricostruite localmente, ma identiche (a parte gli eventuali errori di trasmissione) a quelle presenti nei decodificatori.

Ai coefficienti appartenenti ai blocchi sono applicate le matrici di pesatura (figura 6).

La quantizzazione dei coefficienti DCT è la principale causa di perdita di informazione (e di qualità dell'immagine) nella codifica MPEG. Per ogni macroblocco viene trasmesso il parametro *mquant* che indica il fattore di scala con cui vengono codificati i coefficienti, il codificatore opera la scelta per ottenere la massima qualità (il minimo errore di quantizzazione) con il bit-rate disponibile.

Successivamente si effettua la scansione di tipo zig-zag (figura 7) in modo da operare sulla sequenza dei 64 coefficienti generati per ciascun blocco.

La parole a lunghezza variabile vengono assegnate in base a due informazioni: il numero di zeri che precede un coefficiente non nullo (*run*) e il valore di tale coefficiente (*level*). In base alla coppia *run+level* viene generata una parola VLC la cui lunghezza può variare da 1 a 16 bit + segno. Molte combinazioni *run+level* non sono associate ad una parola VLC, ma sono codificate come *run* (6 bit) + *level* (12 bit). Una parola specifica (EOB, *end-of-block*) indica che tutti i coefficienti del blocco successivi a quello precedentemente codificato sono nulli.

## 5.2 Predizioni

Il predittore normalmente basa la stima del

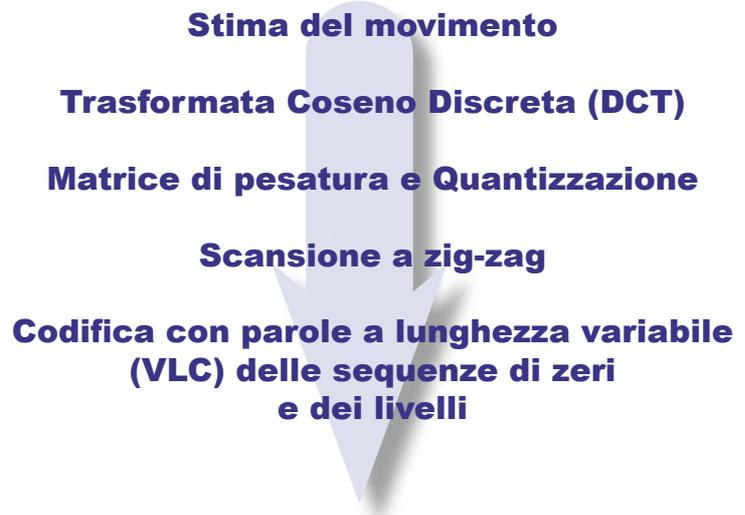


Fig. 5 - Il processo di compressione è basato, nell'ordine, in: stima del movimento (applicata solo nel caso delle *picture* B e P), trasformata DCT applicata ai blocchi 8x8, quantizzazione, scansione dei coefficienti e codifica VLC.

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Fig. 6 - La matrice di pesatura di sinistra è utilizzata per dividere i 64 coefficienti nel caso di blocchi di tipo I, mentre la matrice di destra è utilizzata nel caso dei blocchi di tipo P e B. Sono applicate sia ai blocchi di luminanza che a quelli di cromaticità.

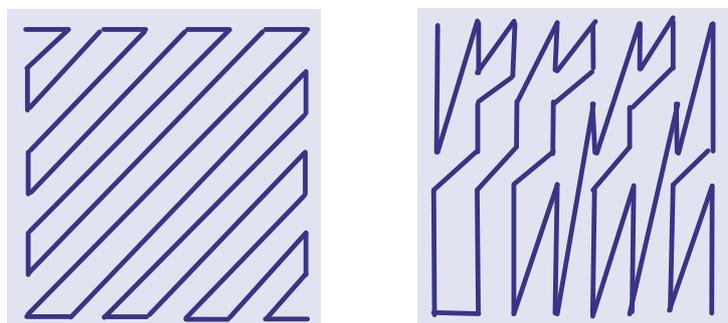
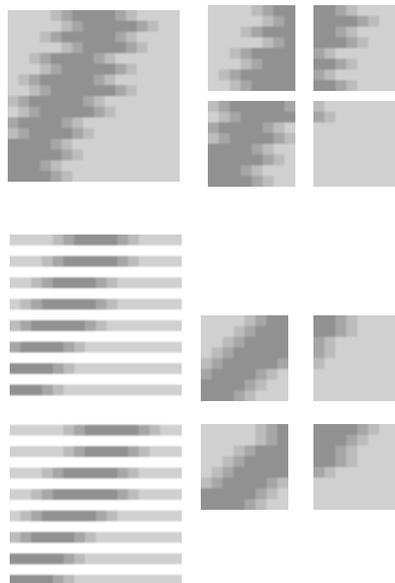


Fig. 7 - I 64 coefficienti DCT sono trasmessi secondo un percorso a zig-zag, dapprima quello relativo alla componente continua e via via quelli relativi alle frequenze spaziali più alte. Il percorso di sinistra è quello utilizzato nel caso di codifica di video in formato progressivo, quello di destra è invece utilizzato nel caso di video in formato interlacciato.

Fig. 8 - Nel caso di video interlacciato, è possibile scegliere se codificare separatamente la luminanza relativa ai due semiquadri. Nel caso di movimenti veloci è possibile che i blocchi ottenuti codificando separatamente le otto righe video appartenenti al *top field* e le otto righe video appartenenti al *bottom field* presentino una correlazione maggiore rispetto alla codifica nei quattro blocchi ottenuti combinando i due *field* in un *frame*. Nell'esempio sono messe a confronto le due possibilità: il *frame* origina quattro blocchi (in alto) che presentano una minore correlazione rispetto ai quattro blocchi originati dai due *field* considerati separatamente (in basso).



movimento sulle immagini memorizzate ottenute grazie alla decodifica locale e la stima è fatta con una precisione pari a mezzo pixel: la predizione ottimale sfrutta al meglio la ridondanza temporale e riduce l'effetto del rumore presente sull'immagine, grazie alle operazioni di media (e quindi filtraggio) su più pixel.

Sono previsti differenti tipi di predizione.

Nella *field prediction*, le predizioni sono ottenute indipendentemente per ciascun semiquadro usando i dati memorizzati relativi ad uno o più semiquadri decodificati. Nelle *field picture* tutte le predizioni sono di tipo *field*.

Nella *frame prediction*, le predizioni sono ottenute da uno o più quadri precedentemente decodificati e la selezione *field prediction* o *frame prediction* è effettuata a livello del singolo macroblocco.

Nel caso di movimenti veloci, ci possono essere vantaggi nel codificare separatamente la luminanza relativa ai due semiquadri (figura 8): per ciascun semiquadro sono originati due blocchi di luminanza ed un vettore movimento (per questi macroblocchi vengono dunque trasmessi due vettori movimento, ciascuno relativo a 16x8 campioni).

I modi di predizione prescelti vengono codificati a lunghezza variabile e inviati al decoder.

Analogamente vengono codificati con VLC i vettori movimento: nel caso di MP@ML i valori codificabili sono da -1024 a +1023,5 in orizzontale e da -128 a +127,5 in verticale.

### 5.3 Qualità costante o bit-rate costante (VBR, CBR)

Si è visto che la perdita in qualità è legata soprattutto all'uso di un fattore di quantizzazione *mquant* più elevato. Al suo crescere diminuisce la precisione con cui vengono trasmessi i coefficienti DCT e conseguentemente aumenta il rumore di quantizzazione, ma si riduce il bit-rate: infatti in ogni blocco vi sono più coefficienti nulli ed il livello di quelli non nulli si riduce, consentendo l'utilizzo di parole VLC più brevi.

Pertanto se si mantiene costante il valore di *mquant*, le immagini sono codificate a qualità costante, ma il bit-rate varia in funzione della criticità dell'immagine rispetto alla codifica.

Per certe applicazioni è conveniente utilizzare una codifica a qualità costante, e conseguente a bit-rate variabile: la codifica VBR (*variable bit-rate*) è spesso adottata per il video memorizzato, ad esempio su supporto ottico (DVD).

In altri casi invece è vincolata la velocità, ad esempio perchè l'informazione deve essere trasferita su un canale a bit-rate costante (CBR, *constant bit-rate*).

In questo caso diventa determinante il ruolo della memoria tampone (*buffer*) indicato nello schema di figura 4. il bit-rate medio dei dati in ingresso deve coincidere con quello all'uscita del *buffer*, costante ed imposto dal canale. Per ottenere tale risultato *mquant* viene variato dinamicamente in funzione del livello di riempimento del *buffer*. Se il *buffer* tende a riempirsi, i coefficienti DCT vengono quantizzati più grossolanamente, il flusso di dati in ingresso diminuisce e il *buffer* tende a svuotarsi. Quando il livello di riempimento è basso, si migliora la quantizzazione e il livello torna a salire.

L'ottimizzazione dell'algoritmo di gestione del buffer deve tenere in considerazione il fatto che le I-picture generano una quantità di dati molto superiori a quelle predette. Le B-picture invece possono contribuire molto poco alla quantità di dati complessiva, anche perché spesso si accetta per esse una quantizzazione più grossolana, poiché il sistema psicovisivo umano è considerato poco sensibile alle fluttuazioni di qualità video se avviene con frequenza prossima alla periodicità IBBP.

## 6. Video elementary stream

Il risultato della codifica video è un flusso di dati binari denominato *Elementary Stream* (ES): esso contiene tutta l'informazione necessaria a decodificare un segnale video.

La figura 9 è una rappresentazione schematica dell'organizzazione dei dati presenti nello *stream* e può essere utilizzata per ricapitolare brevemente gli algoritmi e le funzioni precedentemente descritti.

A livello superiore troviamo la *sequence*, l'intestazione (*sequence header*) contiene le informazioni di base, necessarie per iniziare la decodifica, quali le dimensioni delle *picture*, il formato dell'immagine (*aspect ratio*), la frequenza di quadro, le tabelle di quantizzazione. L'intestazione, data la sua importanza, è ripetuta periodicamente (ad esempio due volte al secondo).

Raggruppati nella sequenza vi sono i *Group Of Pictures*. L'intestazione (*GOP header*) con-

tiene le informazioni necessarie per la riproduzione temporalmente corretta del video (*time code*) e alcuni flag utilizzati nell'editing.

Le *picture* costituiscono ciascun GOP. L'intestazione (*picture header*) contiene un riferimento temporale, l'indicazione del tipo di *picture* (I,P,B).

La *picture* è costituita da *slice*. L'intestazione (*slice header*) è identificata, come le altre intestazioni, da un codice (*start code*) che non può essere duplicato all'interno del flusso. E' l'entità minima, all'interno del flusso elementare video, grazie alla quale è possibile ottenere la sincronizzazione e quindi la corretta decodifica. In genere una *slice* corrisponde ad un insieme di macroblocchi pari a 16 righe video, ma in applicazioni in cui occorra una veloce e sicura sincronizzazione, è possibile avere più *slice*, al limite una *slice* in corrispondenza di ciascun macroblocco.

L'intestazione del macroblocco (*macroblock header*) contiene tutta l'informazione necessaria a decodificare correttamente la porzione di immagine (64 elementi di immagine) che lo costituiscono: l'indirizzo spaziale all'interno dell'immagine, i vettori movimento, i modi di predizione e di trasformazione (*field/frame*), il fattore di quantizzazione.

Seguono i coefficienti DCT codificati VLC (*run+level*) e le parole EOB (*End Of Block*) separano i quattro blocchi (*block*) di luminanza e i due di crominanza che (nel MP@ML) costituiscono un macroblocco.

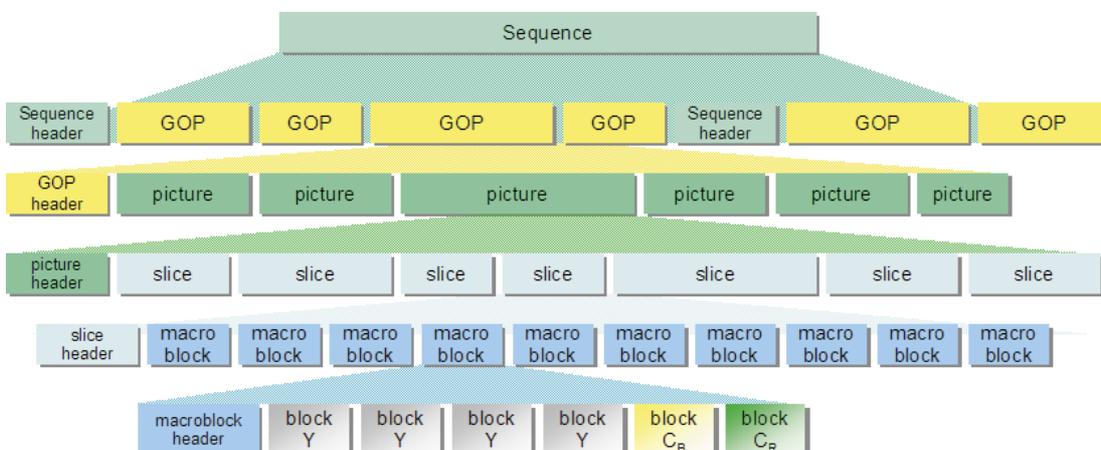


Fig. 9 - Schema dell'organizzazione del video elementary stream.