

# Che cosa è, come funziona: Advanced Video Coding (AVC - H.264): Il prossimo futuro

Parte I

ing. Marzio Barbero e  
ing. Natasha Shpuza

## 1. Premessa

La definizione della codifica video MPEG-2 risale alla prima metà degli anni novanta, negli ultimi anni gli esperti hanno proseguito il loro lavoro per ottenere codifiche sempre più efficienti, soprattutto per le applicazioni a basso bit-rate. L'evoluzione degli algoritmi, ma soprattutto delle capacità di elaborazione integrabili nei *microchip* ha consentito la definizione del sistema AVC (*Advanced Video Coding*) che ha buone possibilità di trovare ampia applicazione nei prossimi anni.

Lo standard specifica la codifica del video (VCL, *Video Coding Layer*), che è oggetto di questa scheda, e il formato con cui organizzare i dati video per il trasporto e la memorizzazione (NAL, *Network Abstraction Layer*).

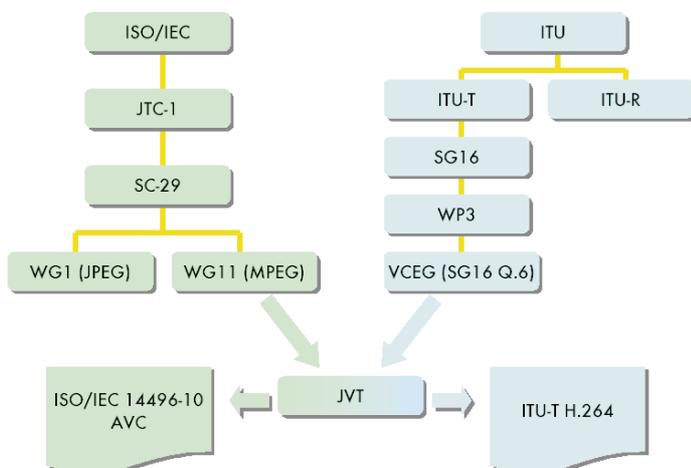
## 2. Cenni storici

La codifica delle informazioni video è oggetto di studio dei gruppi di normalizzazione ISO/IEC (MPEG, *Motion Picture Expert Group*) e ITU (VCEG, *Video Coding Experts Group*), il cui lavoro portò alla definizione della parte 2 di MPEG-2 e allo standard ITU-T H.262 nel 1995. L'ITU sviluppò indipendentemente l'H.263 e due estensioni (pubblicate sotto forma di annessi) e denominate H.263+ e H.263++, mentre in MPEG si procedeva allo sviluppo della parte 2, relativa alla codifica video, dello standard MPEG-4 partendo, come base da H.263.

Nel 2001 fu deciso, per evitare divergenze nello sviluppo ed i problemi di sincronizzazione fra i due organismi di standardizzazione, di stabilire un gruppo congiunto, il JVT (*Joint Video Team*) per portare a termine il lavoro di definizione di un unico sistema di codifica video (figura 1).

Nella riunione MPEG-4 del marzo 2003 a Pattaya venne approvato il nuovo sistema di codifica, AVC (*Advanced Video Coding*), come parte 10 dello standard MPEG-4 ISO/IEC 14496-10. In ambito ITU lo standard, inizialmente indicato provvisoriamente come H.26L, sarà pubblicato come ITU-T H.264.

Fig. 1 - Il Joint Video Team comprende esperti di MPEG e dell'ITU-T e ha completato le specifiche che sono pubblicate separatamente dai due organismi di standardizzazione. Gli standard sono tecnicamente equivalenti, anche se non necessariamente identici.



### 3. Algoritmi e profili

Lo standard AVC, così come avviene nel caso di MPEG-1 e MPEG-2, non definisce un CODEC (cioè codificatore e decodificatore, *CODer* ed *DECoder*), bensì la sintassi del flusso dati (*stream syntax*) e il metodo di decodificarlo. I *tool*, cioè gli algoritmi, adottati, non sono sostanzialmente diversi da quelli illustrati per MPEG-2 video: la maggiore efficienza di codifica è dovuta alla cura dei dettagli di ciascun elemento funzionale (figura 2).

Lo standard supporta la codifica del video nel formato 4:2:0, interlacciato o progressivo. Una *sequence* è organizzata in *frame*, costituiti da due *field* (*top* e *bottom field*). Una sequenza progressiva è codificata come quadri (*frame picture*), consistenti in due semiquadri (*field*), acquisiti nello stesso istante di tempo.

Sono previsti differenti profili, indirizzati ad applicazioni differenti:

- *Baseline Profile*, destinato ad applicazioni in cui si richieda un ridotto ritardo dovuto alla codifica, ad esempio videotelefonica o applicazioni mobili, ...
- *eXtended Profile*, per applicazioni mobili e *streaming*, ...
- *Main Profile*, per applicazioni diffuse, formato video interlacciato, ...

### 4. La codifica

Per ciascun elemento funzionale nel seguito si descrivono brevemente i miglioramenti apportati in AVC rispetto ad MPEG-2, che possono essere sintetizzati in:

- applicazione della trasformata su blocchi più piccoli
- miglioramenti relativi alla valutazione e alla compensazione del movimento
- filtro di ricostruzione nel *loop* di decodifica per ridurre l'effetto di blocchettizzazione
- miglioramento della codifica entropica.

**Predizione intra e inter con stima del movimento strutturata ad albero e precisione a 1/4 pel.**

**Filtro di ricostruzione nel loop di decodifica.**

**Trasformata applicata a blocchi 4x4**

**Matrice di pesatura e Quantizzazione**

**Riordino dei coefficienti**

**Codifica con parole a lunghezza variabile basata sul contesto (CAVLC o CABAC)**

Fig. 2 - Il processo di compressione è analogo a quello applicato negli standard precedenti, in particolare MPEG-2, ma, a spese di una maggiore complessità, consente un migliore sfruttamento delle ridondanze statistiche e riduce la percezione soggettiva delle distorsioni.

#### 4.1 Macroblock e slice

I macroblocchi sono anche in AVC costituiti da 16x16 elementi di immagine: 16x16 campioni (*sample*) di luminanza e 8x8 campioni per ciascuna componente di cromaticità  $C_B$  e  $C_R$ .

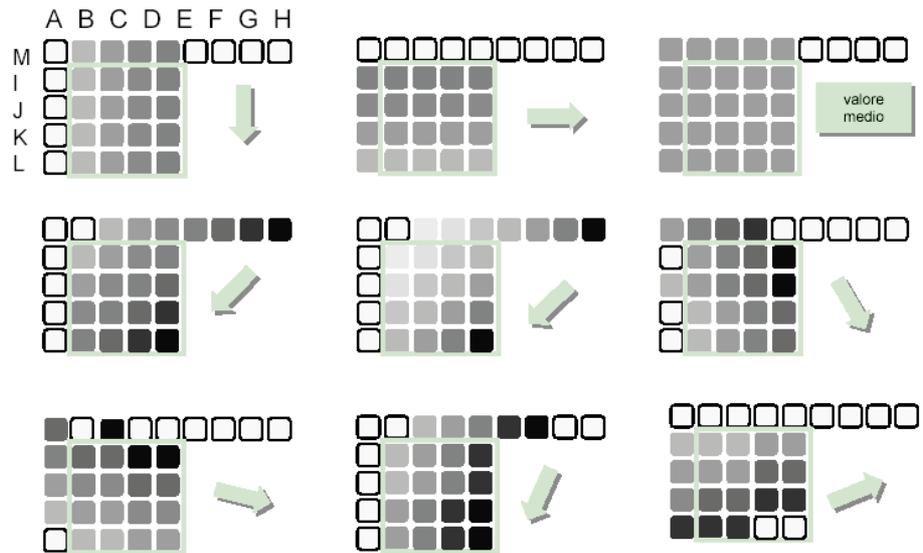
I blocchi (*block*) sono costituiti da 4x4 campioni (un quarto della dimensione adottata in MPEG-2).

I macroblocchi sono organizzati in *slice*, un sottoinsieme di immagine decodificabile indipendentemente dalle altre. L'ordine di trasmissione dei macroblocchi non è necessariamente quello originario nell'immagine, ma è indicato dal codificatore in una apposita mappa (*Macroblock Allocation Map*).

Sono definiti 5 differenti tipi di *slice*.

I primi tre, analogamente a quanto visto per MPEG-2, sono I (*intra*), P (*predictive*) e B (*bi-predictive*) e le predizioni sono ottenute a parti-

Fig. 3 - Nella codifica intra le predizioni per i blocco 4x4 di luminanza sono determinate in base ai valori dei campioni qui indicati con le lettere A-M, precedentemente decodificati. Le frecce indicano la direzione di maggior correlazione e sulla cui base si determinano i valori di predizione. Ad esempio, per il modo 0 (verticale) in alto a sinistra, i valori predetti sono ottenuti a partire da quelli precedentemente decodificati ABCD, mentre nel modo 2 (DC) in alto a destra, tutti i 16 valori sono ottenuti come media di quelli adiacenti al blocco, cioè A-D e I-L.



re dalle *picture* precedentemente codificate. In AVC più *picture* possono essere utilizzate per le predizioni e pertanto codificatore e decodificatore memorizzano le *picture* utilizzate per le predizioni in una apposita memoria (*multi-picture buffer*) e il controllo per la gestione del *buffer* è specificato nel flusso dati.

Nelle applicazioni di *streaming* via internet spesso lo stesso video è codificato a differenti bit-rate ed il decoder tenta di accedere al flusso a più elevato bit-rate, che fornisce una più

elevata qualità, ma se le condizioni del canale non lo permettono, commuta al flusso a bit-rate più basso. Quando si utilizza MPEG-2 queste operazioni di commutazione possono essere effettuate a livello di GOP, in corrispondenza di una *I-picture* e ciò implica l'uso di GOP relativamente corti e lo sfruttamento non ottimale della ridondanza temporale dell'informazione video. In AVC sono stati pertanto definiti ulteriori due tipi di *slice*, denominati SI (*Switching I*) e SP (*Switching P*) che consentono un'efficiente commutazione fra flussi di dati a bit-

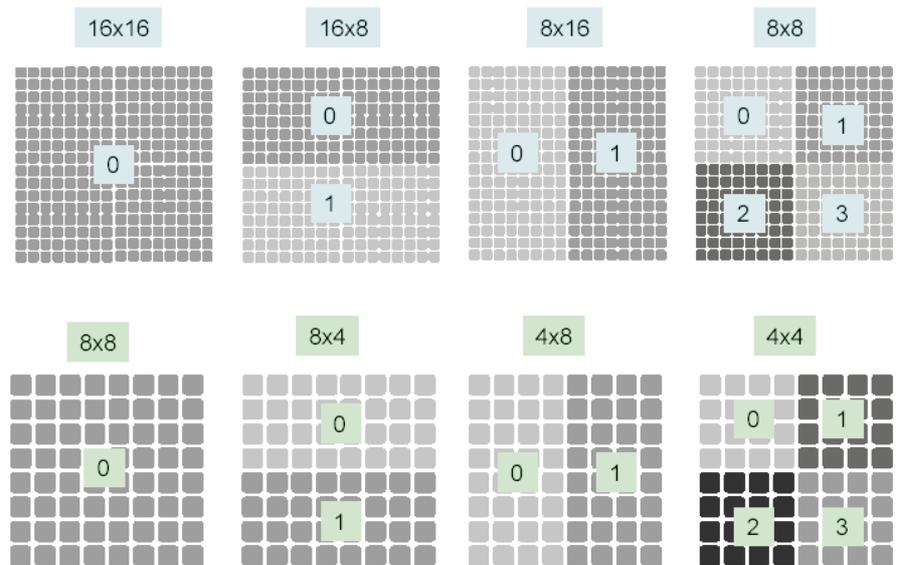


Fig. 4 - La segmentazione dei macroblocchi (luminanza) al fine della compensazione del movimento prevede i 4 modi indicati nella parte alta della figura.

Se è scelto il modo 8x8, ciascuna delle 4 partizioni possono essere ulteriormente suddivise in sottopartizioni, come indicato in basso nella figura.

rate differente, senza rinunciare al massimo sfruttamento della ridondanza temporale.

#### 4.2 Predizione e codifica Intra

Nella codifica intra è sfruttata la sola correlazione spaziale: per aumentare l'efficienza vengono codificate le differenze fra i campioni del macroblocco e i campioni precedentemente codificati, tipicamente quelli posizionati sopra e a sinistra (figura 3) e sono definiti 9 modi distinti di predizione. Nel caso di aree piatte, con scarso dettaglio si può adottare la codifica intra sull'intera area 16x16 ed in tal caso sono definiti altri 4 modi di predizione per l'intero macroblocco.

#### 4.3 Predizione e codifica inter

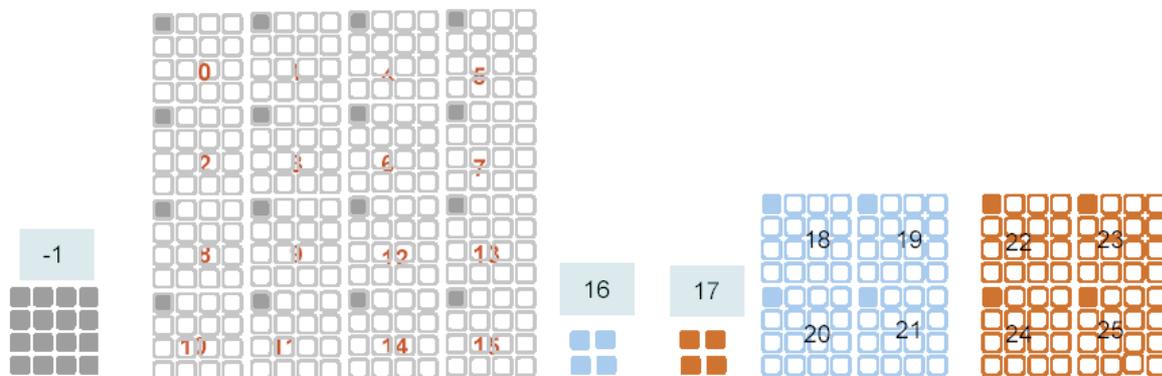
Nella codifica di tipo inter si parte da una predizione ottenuta, sfruttando la correlazione temporale, da uno o due quadri precedentemente codificati. La predizione può essere ottenuta mediante una stima ed una compensazione del movimento (*motion compensated prediction*). A differenza dagli standard precedenti, la dimensione del blocco su cui si effettua la predizione può variare da 16x16 fino a 4x4 (figura 4). Questo metodo di partizionare i *macroblock* in *sub-block* è denominato *tree structured motion compensation* e in fase di

codifica sono possibili molteplici scelte che hanno implicazioni differenti sul numero di bit necessario a codificare i vettori movimento e le differenze residue: in genere dimensioni elevate del blocco sono convenienti in aree piatte, mentre in aree ricche di dettagli si può trarre vantaggio dall'uso di aree ridotte.

La precisione per i vettori movimento si incrementa da 1/2 elemento di immagine, utilizzato in MPEG-2, a 1/4 di elemento di immagine. Per ottenere questa precisione si utilizza un filtro digitale (6-tap FIR, *Finite Impulse Response*) che fornisce, a partire dalla somma pesata dei valori dei 6 campioni di luminanza adiacenti, i valori interpolati a 1/2 e una successiva interpolazione bilineare permette di ricavare i valori a 1/4. Nel caso della cromaticità e di formato 4:2:0 la precisione è portata a 1/8, che corrisponde al valore 1/4 per la luminanza.

Esiste una correlazione fra i vettori movimento delle sottopartizioni adiacenti: essa è sfruttata calcolando un valore di predizione MVP dei vettori movimento relativi ad un macroblocco. Il valore di MVP è calcolato sia in codifica che in decodifica sulla base della struttura in termini di sottopartizioni che costituiscono il *macroblock*. In questo modo al decoder vengono inviati solo gli MVD, i valori delle

Fig. 5 - Ordine di trasmissione dei dati di un macroblocco. Se codificato nel modo 16x16 intra, tutti i coefficienti DC sono raggruppati nel blocco -1 trasmesso per primo. Vengono trasmessi i blocchi 0-15 (nel caso di modo intra i valori corrispondenti alla posizione DC sono posti a 0). Analogamente i blocchi 16 e 17 contengono i coefficienti DC delle componenti di cromaticità  $C_B$  e  $C_R$ , seguono i blocchi 18-25.



differenze fra i vettori movimento e il valore predetto MVP.

#### 4.4 Trasformata e quantizzazione

Si utilizzano tre trasformate che dipendono dal tipo di dati che devono essere elaborati (figura 5):

- una trasformata 4x4 dei 16 coefficienti DC nel caso di macroblocchi intra 16x16
- una trasformata 2x2 per i coefficienti DC delle crominanze di tutti i macroblocchi
- una trasformata 4x4 di tutti gli altri dati differenze.

Il tipo di trasformata adottato è basato sulla DCT (*Discrete Cosine Transform*), ma sono state apportate delle modifiche affinché le operazioni richiedano somme e scalamenti effettuabili con numeri interi a 16 bit in modo da non avere perdita di precisione effettuando la trasformazione diretta seguita da quella inversa.

Esistono 52 passi di quantizzazione, denominati  $Q_p$  e questa ampia gamma di valori permette al codificatore di raggiungere il miglior compromesso fra qualità e bit-rate.

#### 4.5 Filtro di ricostruzione

L'effetto di blocchettizzazione è uno dei degradamenti caratteristici delle tecniche di compressione che operano su macroblocchi di campioni video: è particolarmente visibile e fastidioso.

AVC introduce un filtro apposito che è applicato prima della trasformata inversa sia nel codificatore, prima della ricostruzione delle immagini utilizzate per le predizioni, sia nel

decodificatore. Si ottengono due principali vantaggi: una minore visibilità dei bordi dei blocchi e una migliore predizione inter con compensazione del movimento (nel caso di predizione intra i macroblocchi sono filtrati, ma la predizione è ottenuta dai macroblocchi ricostruiti non filtrati).

#### 4.6 Codifica VLC

I simboli che rappresentano i parametri relativi ai modi di codifica e predizione, i vettori movimento e i coefficienti della trasformata vengono codificati con codici a lunghezza variabile.

Lo standard specifica diversi tipi di codifica entropica: una codifica a lunghezza variabile (VLC, *Variable Length Coding*) basata su tabelle di assegnazione statiche oppure basate sul contesto CAVLC (*Context Adaptive Variable Length Coding*) e CABAC (*Context Adaptive Binary Arithmetic Coding*).

Il CAVLC utilizza diverse tabelle VLC specificamente ottimizzate per i vari elementi sintattici in base a quelli precedentemente trasmessi. A seguito della predizione, trasformazione e quantizzazione i valori relativi ai coefficienti sono molto spesso nulli o molto piccoli: la codifica a lunghezza variabile sfrutta le sequenze di zero (codifica *run-level*), l'elevata frequenza di valori +1 e -1, e la correlazione fra il numero di coefficienti non nulli di un blocco e quello nei blocchi adiacenti.

Il CABAC, che è utilizzato nel *Main Profile*, sfrutta in modo ancora più efficiente la correlazione fra simboli perché utilizza la statistica dei simboli precedentemente codificati per stimare la probabilità condizionata, usata per selezionare uno fra i diversi modelli.