

Sistemi Multimediali I formati video

Ombretta Gaggi
Università di Padova

Video: fondamenti

Il video analogico è codificato da un segnale continuo variabile nel tempo

- può essere digitalizzato, ma non ulteriormente elaborato a causa della natura bidimensionale delle immagini

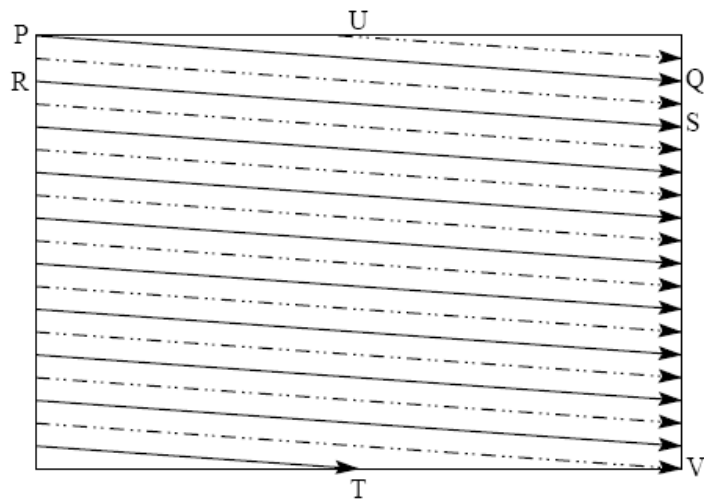
Il video digitale è rappresentato da una sequenza di immagini digitali

- accesso diretto a qualunque fotogramma
- editing video non lineare
- non sono necessari segnali supplementari (*blanking*, sincronizzazione, ...)

Sistemi Ipermediali - 2



Interlacciamento



Sistemi Ipermediali - 3



Interlacciamento: esempio



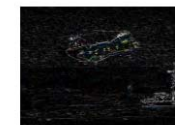
(a)



(b)



(c)



(d)

Sistemi Ipermediali - 4



Video: tipi di segnale

Video a componenti separate

- ogni segnale primario (RGB, YUV) è trasmesso come segnale separato
- consente la migliore riproduzione dei colori per l'assenza di fenomeni di interferenza tra i segnali
- richiede una banda passante molto elevata e una sincronizzazione molto precisa tra i tre segnali

Video composito

- i segnali di luminanza e di cromaticanza sono miscelati su una sola portante
- si introducono alcune interferenze tra i segnali

S-Video

- i segnali di cromaticanza sono miscelati in una sola portante mentre il segnale di luminanza viene spedito separatamente

Il video analogico utilizza normalmente un segnale composito (sempre per la trasmissione)

Il video digitale utilizza un segnale a componenti separate



Video: cavi per i diversi tipi di segnale

Video a componenti separate



Video composito



S-video



Video: proprietà

Profondità di colore

- la registrazione codifica immagini *true color*
- la riproduzione dipende dal dispositivo di visualizzazione

Risoluzione

- dipende dagli standard
- l'informazione di cromaticanza è sottocampionata

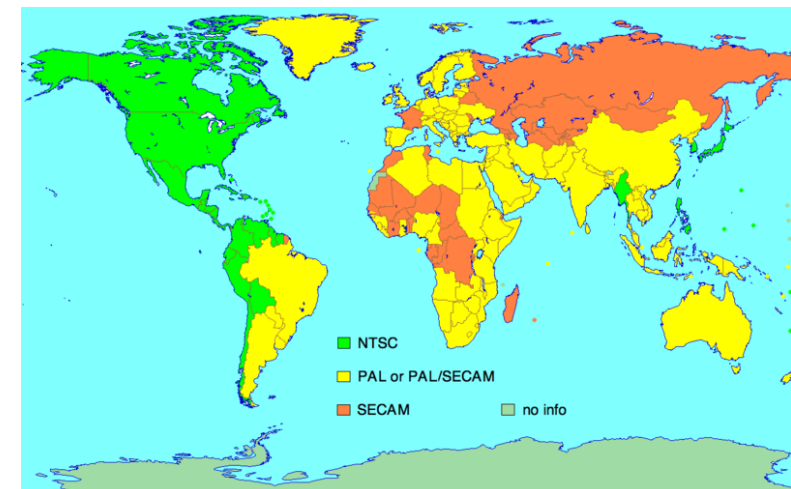
Frequenza dei fotogrammi

- PAL = 25 fotogrammi/sec
- NTSC = 29.97 (~ 30) fotogrammi/sec
- minimo ~ 15 fotogrammi/sec per evitare la percezione a scatti dei movimenti

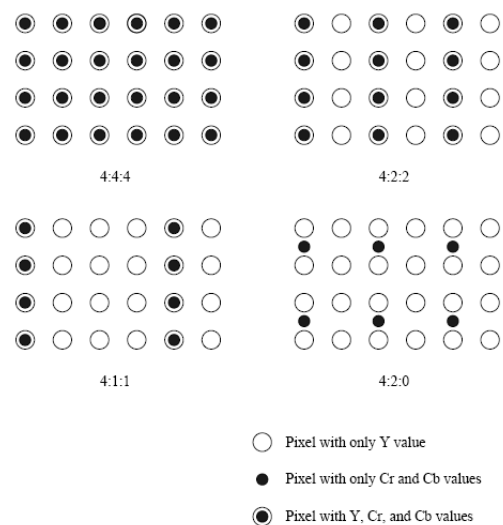
CCIR 601	720 x 480
NTSC	(525)
CCIR 601	720 x 576
PAL	(625)
CIF	352 x 288
QCIF	176 x 144



Mappa della distribuzione geografica dei sistemi TV



Campionamento



Video: occupazione di memoria

Il video non compresso è di dimensioni molto elevate

- La televisione ad alta definizione (HDTV) richiede un bit-rate che può superare 1 Gbps

I dati devono essere compressi

- un'ora di video MPEG-1 a qualità VHS (352 x 288, 25 fotogrammi/sec) occupa un CD-ROM (~600 Mbyte)

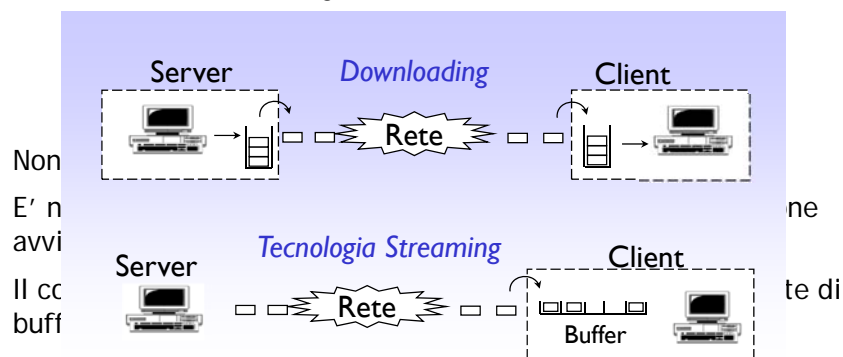
Si utilizzano tecniche di compressione con perdita

- eliminazione della ridondanza spaziale e temporale
- codifica *intra-frame* e *inter-frame*



Video: tempo di trasferimento

Il caricamento da rete del video ha gli stessi problemi del caricamento delle immagini, e inoltre...



Motion JPEG

È il primo tentativo di video digitale

Codifica un segnale video come una sequenza di fotogrammi: ogni frame è codificato come un'immagine JPEG

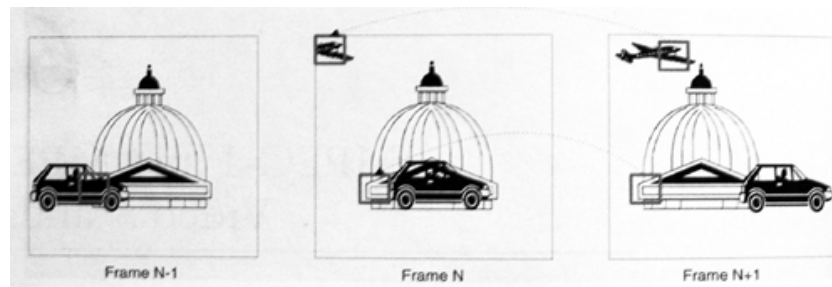
Non sfrutta in alcun modo l'evidente correlazione tra un frame e il successivo



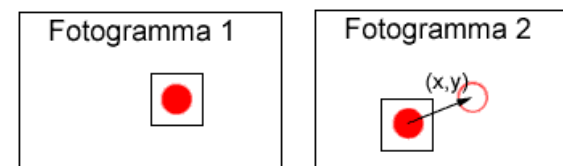
Ridondanza Temporale

Nella codifica dei fotogrammi di un video è possibile omettere molte informazioni perché, a meno di cambi di scena, ci sono poche differenze tra due immagini che distano di un breve lasso di tempo

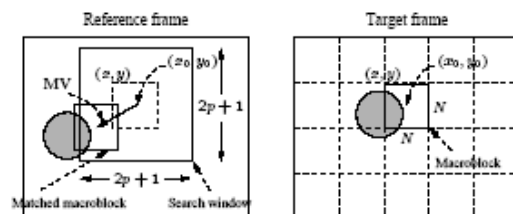
Le differenze tra un fotogramma ed il successivo sono per lo più dovute solo a traslazioni di parti di esso



Esempio



Algoritmi di ricerca



$$MAD(i, j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)|$$

N – size of the macroblock,

k and l – indices for pixels in the macroblock,

i and j – horizontal and vertical displacements,

$C(x+k, y+l)$ – pixels in macroblock in Target frame,

$R(x+i+k, y+j+l)$ – pixels in macroblock in Reference frame.

$$(u, v) = [(i, j) \mid MAD(i, j) \text{ is minimum, } i \in [-p, p], j \in [-p, p]]$$



Sequential Search (Full Search)

L'algoritmo *Sequential Search* ricerca tutto lo spazio $(2p+1) \times (2p+1)$ un macroblocco simile (minimo MAD) al macroblocco preso in considerazione

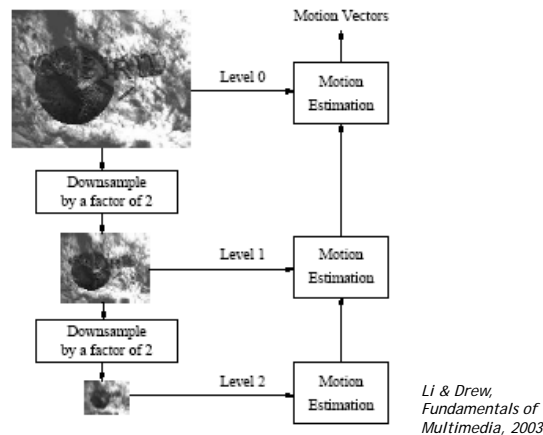
- il macroblocco target viene confrontato, bit per bit, con un macroblocco centrato in ogni posizione dello spazio di ricerca, e viene calcolato il MAD
- nel vettore di movimento viene salvata la differenza di posizione del macroblocco target con quello con MAD minimo

È un algoritmo computazionalmente molto costoso: $O(p^2 N^2)$



Ricerca Gerarchica

L'algoritmo di ricerca gerarchica lavora per livelli di approssimazione in cui la prima stima viene fatta su un'immagine a risoluzione minore



Standard video H.261

H. 261, sviluppato da CCITT nel 1988-1990

- progettato per applicazioni di videoconferenza e videotelefonia su linee telefoniche ISDN
- immagini in formato CIF (352 x 288) e QCIF (176 x 144), 4:2:0
- il bit-rate è $p \times 64 \text{ Kb/sec}$, $1 \leq p \leq 30$

Codifica:

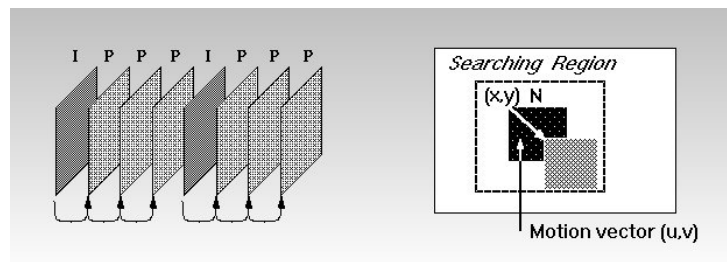
- la codifica e la decodifica devono avvenire real-time con un ritardo massimo di 150 ms
- la frequenza di refresh in input deve essere di 29.97 fps (video non interlacciato) mentre quella in output varia tra 10 e 15 fps
- utilizza lo spazio di colori *YCbCr* con le componenti di cromaticità sottocampionate
- prevede 2 tipi di frame: intra-frames (*I-frames*) e inter-frames (*P-frames*)
- *Intra-frames*: sono i frame sui quali è costruito il video
- *Inter-frames*: sono codificati da informazioni di altre immagini



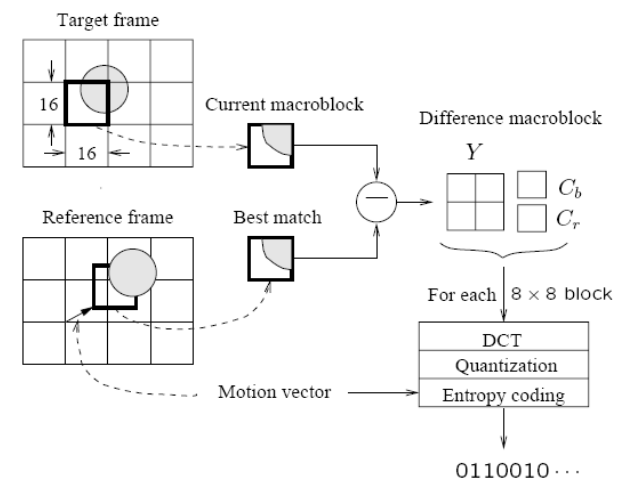
Standard video

H.261 e H.263 basano la codifica di un fotogramma sull'analisi delle differenze rispetto al fotogramma precedente

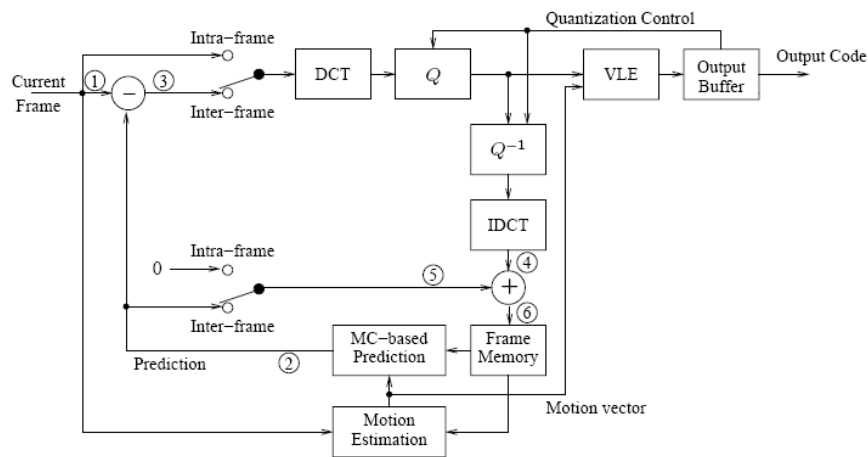
- sono codificate solo le differenze e il contenuto ricostruito per confronto
- la compensazione di moto stima il movimento di porzioni di immagine tra fotogrammi successivi e codifica la differenza rispetto alla stima (H.263)



P-frame



Encoder H.261



Standard video H.263

H. 263 (1996), codifica migliorata per bit-rate bassi

- il formato delle immagini è variabile da 128 x 96 fino a 1480 x 1152
- l'algoritmo di compressione è migliore e in grado di codificare flussi video con bit-rate inferiori a 64 Kbps
- include dei meccanismi di correzione degli errori

La codifica intraframe prevede:

- **PB-frame** per aumentare il frame-rate senza aumentare il bit-rate
- vettori di movimento senza restrizioni
- predizione avanzata: la precisione nella compensazione del moto è portata a mezzo pixel



Vettore di moto

Una volta calcolate le differenze tra due fotogrammi, si trasmettono solo il verso e l'entità dello spostamento (**motion vector**)

H.263 permette al vettore di moto di riferirsi a pixel al di fuori dei confini dell'immagine (**unrestricted motion vector mode**) associando il pixel più vicino giacente sul bordo dell'immagine, a quello indicato dal MV, esterno all'immagine.

Integer Pixel Motion Estimation

- l'immagine viene divisa in macroblocchi (MB) 16x16 o 8x8
- per ogni macroblocco è calcolato il vettore di moto cercando, nel frame precedente, un MB che sia il più simile possibile a quello corrente.
- si cerca in un intorno della posizione originaria muovendosi orizzontalmente e verticalmente di ± 15 pixel a passi di un pixel per volta.



MPEG, Motion Picture Expert Group (1)

La prima versione di MPEG risale al 1991 e permette la compressione di una sequenza di immagini e la memorizzazione in CD

Permette l'accesso casuale al filmato e le ricerche veloci

L'algoritmo di compressione è molto complesso ma fortemente asimmetrico: assicura una decompressione in real-time

Come lo standard H.261, il video MPEG utilizza lo spazio di colori **YCbCr** (8 bit) con le componenti di cromaticità sottocampionate

La risoluzione della luminanza non può superare 768x576 pixel

Non supporta il video interlacciato



MPEG, Motion Picture Expert Group (2)

Sono permesse diverse risoluzioni e diverse frequenze di refresh (da 23.98 fps a 60 fps)

L'informazione video contiene:

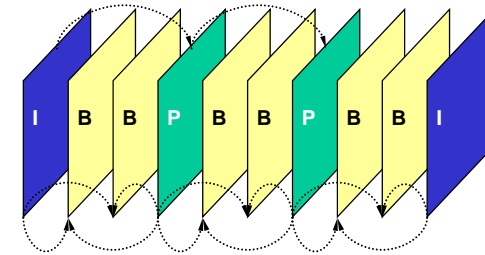
- *ridondanza spaziale* → codifica delle singole immagini
 - ✓ codifica JPEG
- *ridondanza temporale* → relazione tra fotogrammi successivi
 - ✓ codifica diversificata dei singoli fotogrammi



Algoritmo di compressione MPEG

MPEG estende gli algoritmi di compressione H.261 and H.263 con uno schema di previsione del movimento più sofisticato

- i fotogrammi di tipo I (*Intra coded frame*) sono codificati interamente con un algoritmo JPEG in modo indipendente ma con qualità più bassa
- i fotogrammi di tipo P (*Predictive coded frame*) sono codificati in base ad una stima riferita al fotogramma I o P precedente
- i fotogrammi di tipo B (*Bidirectionally predictive coded frame*) sono codificati in base a due stime di movimento relative ai fotogrammi precedenti e seguenti (stima bidirezionale)



Fotogrammi MPEG (1)

Gli *Intracoded frame*

- richiedono maggiore spazio di memorizzazione
- bloccano la propagazione degli errori dovuti alla trasmissione
- rendono possibile l'accesso casuale

I *Predictive coded frame*

- il calcolo delle differenze viene fatto sul valore assoluto delle componenti di luminanza
- sono più "piccoli" ma propagano gli errori di trasmissione

I *Bidirectional predictive coded frame*

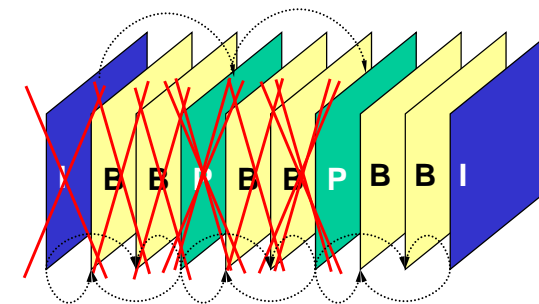
- sono i più compressi

Più I-frame permettono l'accesso casuale in più punti, ma aumentano il bit-rate

- IBBPBBPBBIBBPBBPBB...
- ci deve essere almeno un I-frame ogni 15 frame



Fotogrammi MPEG (2)



Motion Compensation Prediction (1)

3 fasi

- stima del movimento degli oggetti e creazione del vettore movimento
- stima dei frame utilizzando le informazioni ricavate dalla fase precedente
- confronto tra il frame stimato e l'originale per il calcolo dell'errore

Si memorizzano solo il vettore movimento e l'errore nella stima

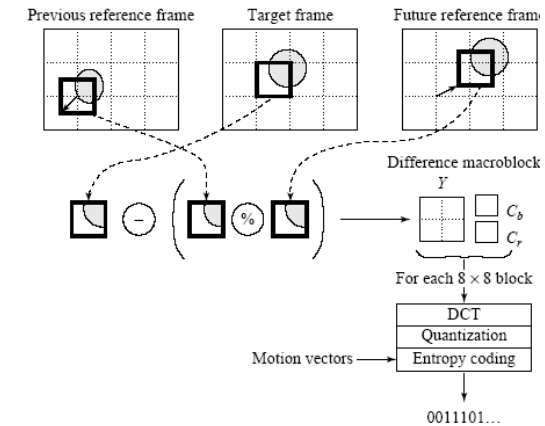
MPEG utilizza una precisione a *mezzo bit*:

- ogni blocco 16x16 viene espanso per interpolazione ad un blocco virtuale 32x32
- si cerca nel macro blocco la posizione del blocco originale
- il risultato è ottenuto dall'interpolazione del blocco virtuale 32x32 contenente il blocco originario spostato
- lo spazio di ricerca è di ± 512 pixel per la precisione a mezzo pixel e di ± 1024 pixel per la precisione a pixel interi

Nell'algoritmo di ricerca risiede la complessità dell'operazione



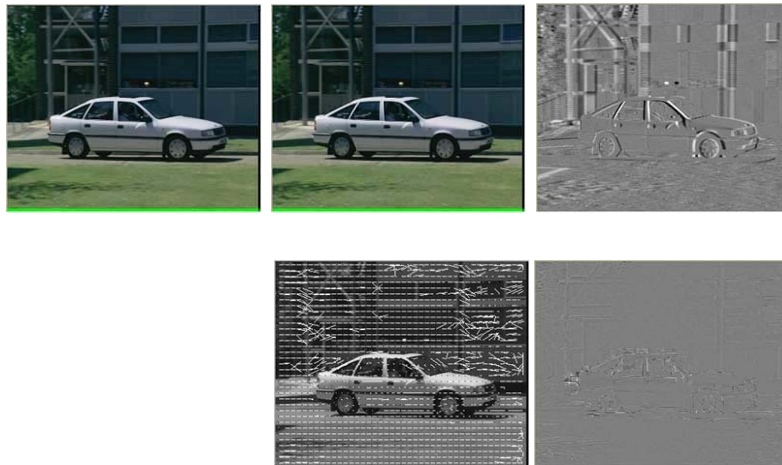
Motion Vector per i B-frame



Li & Drew, Fundamentals of Multimedia, 2003



Motion Compensation Prediction (2)



Dimensioni dei Macroblocchi

Uno dei problemi principali riguarda la dimensione dei macroblocchi a cui applicare l'algoritmo di *motion compensation prediction*

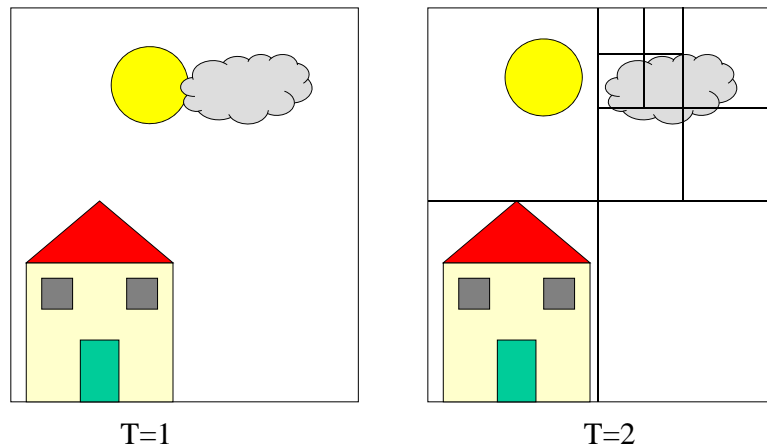
- blocchi di grandi dimensione → predizione poco precisa
- blocchi di piccole dimensioni → aumento della complessità della codifica

Blocchi a dimensioni variabili:

- Quad-tree methods
- Binary-tree methods
- H.26L



Metodi Quad-Tree



Blocchi a dimensioni variabili

Pro

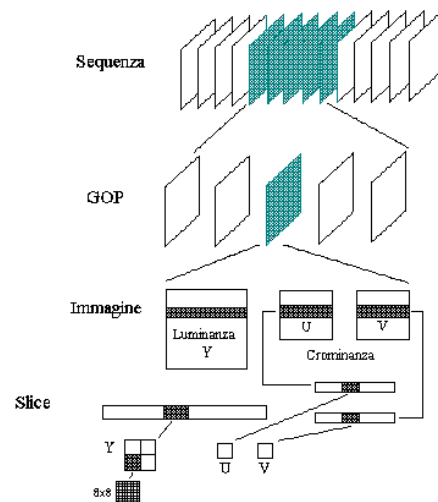
- la predizione è molto più accurata
- quanto più la predizione è accurata, tante meno sono le differenze da codificare

Contro

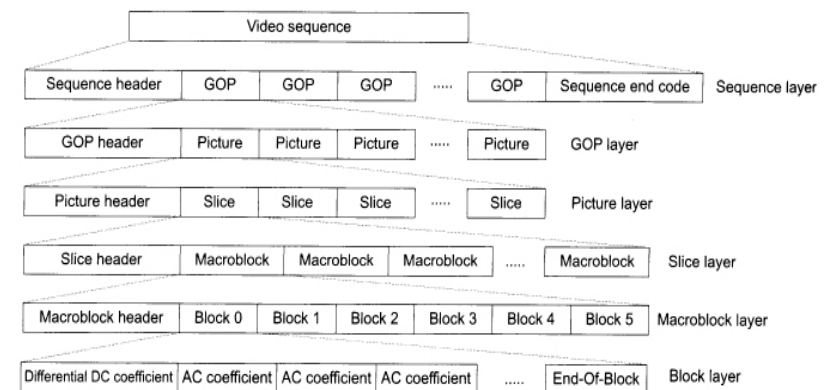
- computazione molto costosa
- la descrizione della delimitazione dei macroblocchi (chiamati *regioni*) è molto complessa



La struttura MPEG (1)



La struttura MPEG (2)



Prestazioni & Applicazioni

Considerando immagini in formato CIF (352 x 288), la conversione in MPEG comporta una qualità comparabile e un fattore di compressione di circa 30:1

Si possono ottenere fattori di compressioni più alti a discapito della qualità

Applicazioni:

- video su cd (cd dimostrativi, museali,...)
- videogiochi
- educazione a distanza (ma non real-time)
- ...



La famiglia MPEG (1)

MPEG-1

- CD-ROM video di media qualità
- la qualità è comparabile alla qualità delle registrazioni su nastro VHS
- la decodifica non richiede hardware specializzato su PC standard di mercato

MPEG-2

- DVD-ROM video di alta qualità (bitrate superiori a 4Mbps)
- la qualità è comparabile o superiore rispetto alle trasmissioni televisive commerciali
- è il formato standard per le applicazioni *consumer* di fascia alta
- richiede hardware specifico per la decompressione a meno di non impegnare completamente il PC
- supporta il video interlacciato



Profili e livelli MPEG-2

Table 11.5: Profiles and Levels in MPEG-2

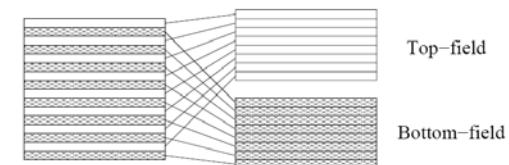
Level	Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Table 11.6: Four Levels in the Main Profile of MPEG-2

Level	Max Resolution	Max fps	Max Pixels/sec	Max coded Data Rate (Mbps)	Application
High	1,920 × 1,152	60	62.7 × 10 ⁶	80	film production
High 1440	1,440 × 1,152	60	47.0 × 10 ⁶	60	consumer HDTV
Main	720 × 576	30	10.4 × 10 ⁶	15	studio TV
Low	352 × 288	30	3.0 × 10 ⁶	4	consumer tape equiv.



Motion Prediction in MPEG-2 (1)



(a) Frame-picture vs. Field-pictures

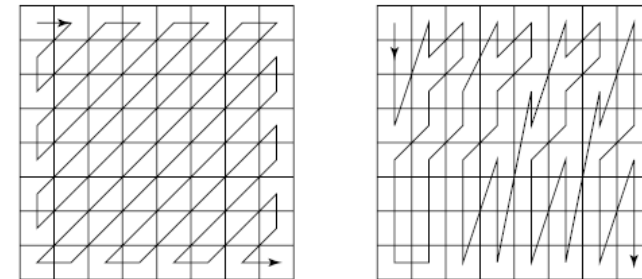


Motion Prediction in MPEG-2 (1)

- MPEG-2 supporta 5 differenti modalita' di motion prediction:
 - Frame prediction for frame-picture
 - Field prediction for field-picture
 - Field prediction for frame-picture
 - 16x8 MC for field-pictures
 - Dual-prime for P-pictures



Zig-zag scan vs Alternate scan



Differenze con MPEG-1

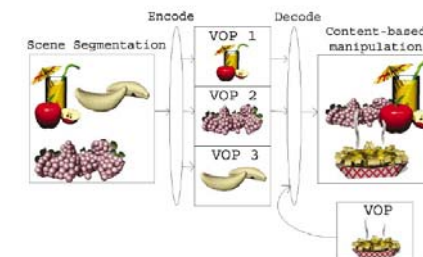
- Migliore resistenza agli errori
- Supporta il sottocampionamento cromatico 4:2:2 e 4:4:4
- Quantizzazione non lineare
- Maggiore flessibilità del formato video



La famiglia MPEG (2)

MPEG-4 (1999)

- permette di integrare *stream* video e oggetti creati in modo indipendente
- e' ottimizzato per 3 diversi bitrate: < 64Kbps, 64-384 Kbps, 384-4Mbps
- consente l'indicizzazione dei singoli componenti di una scena
- si rivolge alle applicazioni che coinvolgono sistemi multimediali complessi e interattivi
- "...one single technology for playing every where...": supporto per diversi dispositivi e bande di rete a disposizione



MPEG-4



Esempi di applicazione

- Video streaming su Internet
- Video su cellulari
- Content-based storage and retrieval
- DVD interattivi
- Produzione televisiva
- Monitoraggio e sorveglianza remota
- Infotainment
- Virtual meeting



Formati non proprietari: DivX e XviD

È un codec di compressione video basato sullo standard MPEG-4. Un CODEC (COder - DECoder) svolge sia le funzioni di codifica che di decodifica

DivX nasce come codec **illegale**: DivX;-) è una versione *crakkata* del codec MPEG-4 introdotto da Microsoft nel formato proprietario **WMV** per sopperire ad alcune sue limitazioni:

- il codec Microsoft permetteva la creazione di video nel solo formato proprietario wmv
- le opzioni configurabili sono pochissime
- molti parametri sono fissati per lo stream di dati in tempo reale via Internet

DivX (v. 4.0 in poi) è una riscrittura completamente **open source** e **free** del codec originario

DivX (v. 5.0) è diventato di nuovo un formato proprietario

- **Xvid**



DivX: compressione

DivX raggiunge livelli di compressione molto alti: un video di due ore in formato DVD può essere memorizzato in due CD con perdita di qualità quasi nulla

Utilizza la codifica MP3 per l'audio e MPEG-4 per la parte video:

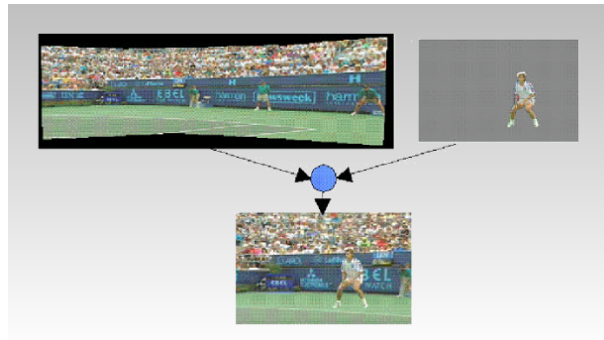
- compressione efficiente di immagini e video
- la gamma di bitrate offerta varia tra i 5 kbps e i 10 Mbps
- robustezza agli errori
- rindondanza spaziale → *codifica intra-frame*
- rindondanza temporale → *codifica inter-frame*



MPEG-4 Video

Una scena animata può essere scomposta in due parti

- il movimento dello sfondo è limitato ai movimenti della telecamera, quindi può essere codificato come immagine fissa + movimenti codificati (sprite panorama)

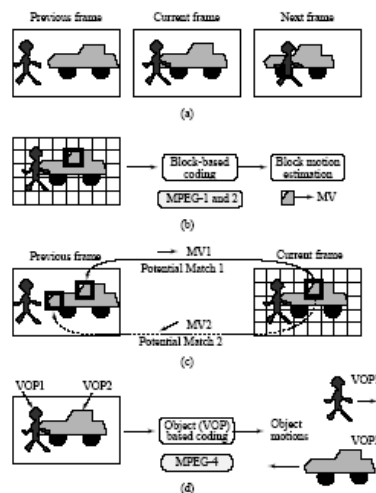


Descrizione gerarchica di una scena in MPEG-4

1. Video-object Sequence (VS): la scena completa; può contenere sia oggetti naturali che sintetici
2. Video Object (VO): un oggetto particolare della scena. Ha una forma arbitraria, corrispondente ad un oggetto o al background della scena stessa
3. Video Object Layer (VOL): supporta la codifica scalabile per livelli; ogni VO può avere più VOL (codifica scalabile) oppure uno solo (codifica non scalabile)
4. Group of Video Object Plane (GOV): è un livello opzionale che consente di considerare sequenze di VOP
5. Video Object Plane (VOP): uno snapshot di un VO in un particolare momento



Codifica a frame vs codifica object-oriented

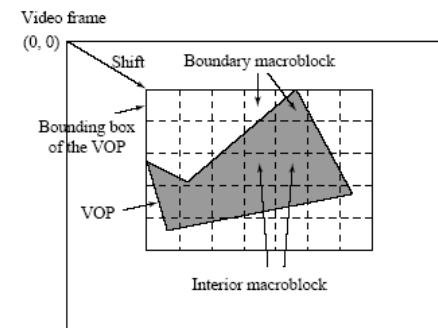


Motion compensation in MPEG4

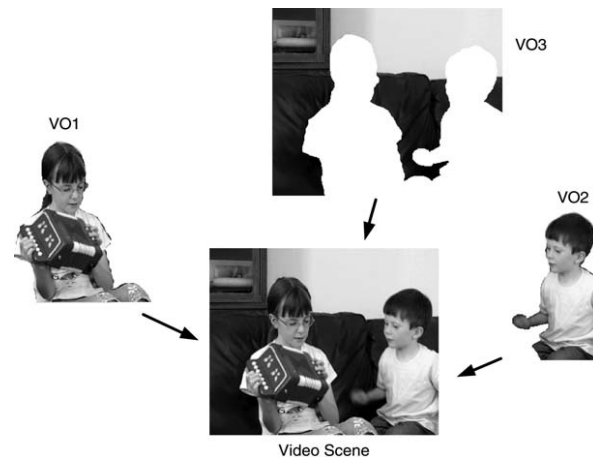
La forma di ogni VOP è arbitraria e deve essere codificata insieme alla *texture* (toni di grigio utilizzati)

Ogni VOP viene diviso in blocchi 16x16 e si cerca il movimento dell'oggetto globale

Per applicare la DCT (che richiede matrici quadrate) alla MC si usa il padding



Composizione di Video Object in MPEG4



Maschere



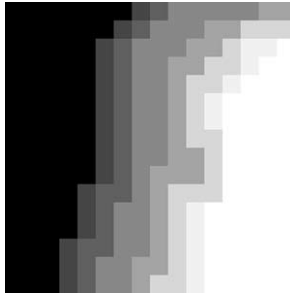
Maschere - Dettaglio



Composizione con uno sfondo diverso



Maschera a scala di grigi



Codifica DivX

DivX utilizza una tecnica di codifica a bit-rate variabile:

- alloca maggior bit-rate per scene di azione e
- una minore quantità di banda trasmissiva per scene più "calme"

Il codec non conosce *a priori* la quantità di banda necessaria dalle scene future quindi non è in grado di sfruttare a pieno il bit-rate

Codifica a 2 passaggi:

1. verifica delle scene d'azione presenti
2. codifica vera e propria



Formati proprietari

AVI (Audio Video Interleave)

- formato proprietario Microsoft, standard *de facto* nei sistemi Windows
- è un caso particolare di file RIFF (Resource Interchange File Format) utilizzato per codificare numerosi formati diversi secondo principi simili
- supporta numerosi *codec* diversi anche proprietari

QuickTime

- formato proprietario Apple, standard nei sistemi Macintosh
- esistono player su tutte le piattaforme
- supporta numerosi *codec* diversi anche proprietari
- supporta QuickTime Virtual Reality e MIDI
- consente una forma limitata di *streaming*

Real Media

- progettato espressamente per applicazioni *streaming* di audio e video
- il livello di compressione e la qualità di riproduzione sono variabili
- standard *de facto* multiplatforma per le trasmissioni *streaming*



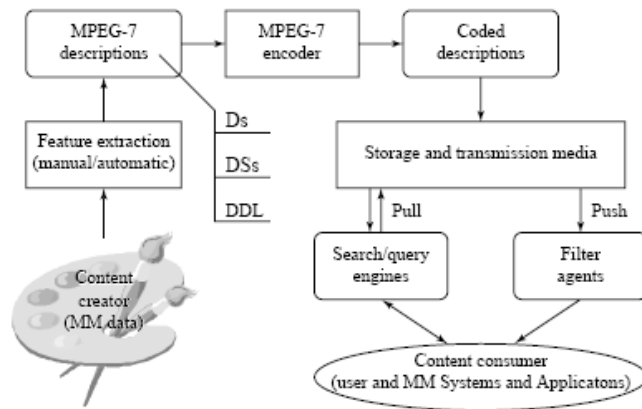
La famiglia MPEG (3)

MPEG-7

- definisce come rappresentare in modo standard un descrizione di contenuto
- associa agli oggetti che compongono un'applicazione multimediale una serie di descrittori per consentirne la classificazione e la ricerca per contenuto
- definisce contenitori generici per oggetti di media differenti secondo standard diversi
- combina descrizioni estratte automaticamente dai media, con descrizioni offerte da un utente umano
- si rivolge ad applicazioni di *information retrieval* basate sul contenuto
- è uno standard definito recentemente (settembre 2001)
- *NON* definisce come estrarre e codificare le descrizioni di contenuto e come utilizzare tali descrizioni



Possibili applicazioni di MPEG-7

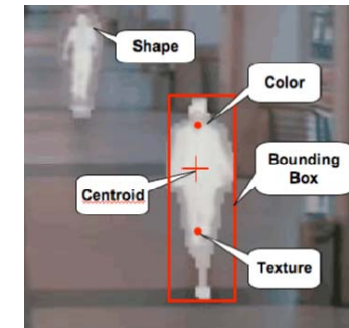


Li & Drew, Fundamentals of Multimedia, 2003




Interazione uomo - sistema IR

- Un utente si esprime in termini di colore, forma e posizione, magari fornendo un'altra immagine (schizzo) dalla quale partire per la ricerca
- Un sistema di IR si basa su proprietà specifiche del media cercato (ex. Immagini: istogramma dei colori, texture, forme,...)



Caratteristiche e descrittori

Color GoF/GoP Color Scalable Color Color Layout Color Structure Dominant Color	Texture Homogeneous Text. Texture Browsing Edge histogram	Shape  Region Shape Contour Shape 3D Shape 2D-3D Multiple View
Motion Camera Motion Motion Trajectory Parametric Motion Motion Activity	Localization Bounding Box Region Locator Spatio-Temporal Locator	Other Face Recognition



Livelli di descrizione del contenuto

- Feature: una caratteristica dei dati
- Descriptor (D): una definizione (sintattica e semantica) di una caratteristica
- Description Scheme (DS): specifica della struttura e delle relazioni tra D e DS
- Description: un insieme di D e DS che descrivono in contenuto multimediale, la sua struttura, la memorizzazione, gli utilizzi possibili, etc.
- Description Definition Language (DDL): un linguaggio (le regole sintattiche) per definire D e DS



La famiglia MPEG (4)

- MPEG - 21 (~ 2003)
 - E' pensato per la protezione dei contenuti digitali
 - Descrizione dei contenuti e dei diritti di chi li ha creati
 - Deve fornire un'interfaccia che faciliti l'utilizzo dei media (ricerca, tecniche di caching, etc)

