

Principali algoritmi di compressione per voce, immagini e video

1

Sommario

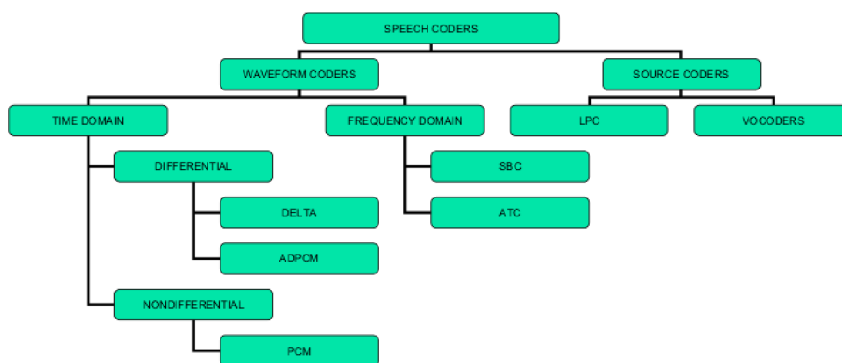
- Compressione della voce
- Compressione di immagini
 - JPEG
- Compressione di video
 - algoritmi
 - principali standard

2

Compressione della voce

3

Principali metodi compressione voce



Courtesy of R. Z. Zaputowycz

4

Pulse Code Modulation (PCM)

- Il segnale vocale viene
 - filtrato a 3.3 kHz
 - campionato a 8 kHz
 - quantizzato su 8 bit
- Quantizzatore logaritmico
 - preciso a volume basso
 - approssimato a volume alto
 - adatto alla percezione umana
- ITU-T G.711
 - 64 kb/s
 - due tipi di curve logaritmiche: uLaw, aLaw

5

Differential PCM (DPCM)

- Il segnale vocale viene
 - filtrato a 3.3 kHz
 - campionato a 8 kHz
 - quantizzato su 16 bit
- Codifica della differenza tra due campioni successivi
- Eliminazione della correlazione statistica

6

Adaptive Differential PCM (ADPCM)

- Il segnale vocale viene
 - filtrato a 3.3 kHz
 - campionato a 8 kHz
 - quantizzato su 16 bit
- Codifica della differenza tra due campioni successivi
- Quantizzazione della differenza variabile nel tempo (**adattativa**)
 - bassa precisione quando il volume e' alto
 - alta precisione quando il volume e' basso
- ITU-T G.726
 - vari bitrate: 40, 32, 24, 16 kb/s

7

Compressione di immagini

8

Immagini in bianco e nero

- Campionamento
 - Griglia di pixel
 - Risoluzione data da colonne x righe
- Quantizzazione
 - 8 bit/pixel --> Matrice di valori interi su 8 bit
 - La quantizzazione può essere uniforme o non-uniforme (p.es., logaritmica), a seconda della distribuzione di probabilità delle intensità e il comportamento percettivo umano.

9

Immagini a colori

- Ogni colore è dato dalla sovrapposizione di 3 componenti
- Tricromia additiva
 - R (rosso)
 - G (verde)
 - B (blu)
- Tre matrici per rappresentare la griglia di pixel

10

Rappresentazione YUV

- In principio la TV era solo in bianco e nero
- Poi avvio delle trasmissioni a colori
 - necessità di poter vedere le trasmissioni a colori anche su televisori in bianco e nero
- Passaggio da rappresentazione RGB a YUV:
 - Y, luminanza (toni di grigio)
 - U e V cromaticanza
- I televisori in bianco e nero usano solo la componente Y
- I televisori a colori decodificano anche U e V

11

Conversione RGB --> YUV

- La componente Y è una media pesata delle componenti RGB
 - $Y = rR + gG + bB$
- $U = B - Y$
- $V = R - Y$
- La conversione RGB --> YUV è una trasformazione lineare
 - reversibile
 - rappresentazione matriciale

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} r & g & b \\ -r & -g & 1-b \\ 1-r & -g & -b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

12

Percezione delle comp. YUV

- L'occhio umano è molto sensibile ai contorni
- I contorni sono veicolati dalla componente Y















- Le componenti U e V possono avere una risoluzione spaziale minore della componente Y senza evidenti perdite di qualità



- Opportunità di compressione (lossy)













13

Formato CIF

Base				3 matrici 352 x 288
4:4:4				3 matrici 352 x 288
4:2:2				1 matrice 352 x 288 e 2 matrici 176 x 288
4:1:1				1 matrice 352 x 288 e 2 matrici 176 x 144

14

Formato Quarter-CIF (QCIF)

Base				3 matrici 176 x 144
4:4:4				3 matrici 176 x 144
4:2:2				1 matrice 176 x 144 e 2 matrici 88 x 144
4:1:1				1 matrice 176 x 144 e 2 matrici 88 x 72

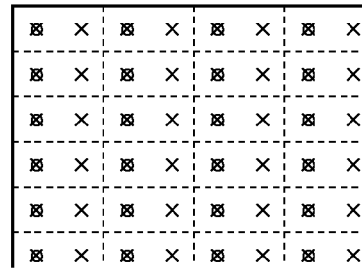
15

Sottocampionamento di U e V

- Prima di sottocampionare le componenti U e V bisogna passarle in un filtro passa-basso al fine di ridurre la larghezza e soddisfare il teorema di Nyquist
- ESEMPIO DIMOSTRATIVO: supponiamo di dover sottocampionare del 50% un'immagine con righe bianche e nere alternate:
- filtrando ottengo un'immagine grigia (OK)
- senza il sottocampionamento avrei un'immagine tutta nera o tutta bianca (**aliasing**)

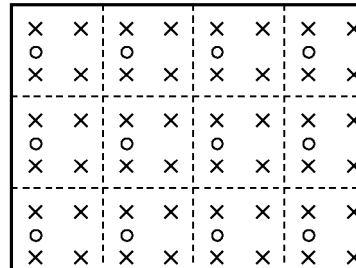
16

Sottocampionamento: campioni e pixel



X Represent luminance samples
O Represent chrominance samples

4:2:2



X Represent luminance samples
O Represent chrominance samples

4:1:1

17

Sottocampionamento: guadagno

- Formato QCIF
- 8 bit per campione (non per pixel !)
- **4:4:4**
 - $176 \times 144 \times 3 = 76464$ byte
- **4:2:2**
 - $176 \times 144 \times 2 = 50976$ byte
- **4:1:1**
 - $176 \times 144 \times 1,5 = 38232$ byte

18

Sottocampionamento: applicazioni

- **4:4:4**
 - applicazioni studio, mediche e per archiviazione
 - utile quando occorre elaborare più volte l'immagine
- **4:2:2 e 4:1:1**
 - sufficiente per applicazioni domestiche
- **ATTENZIONE**

La compressione 4:2:2 e 4:1:1 si può effettuare solo su matrici YUV

19

JPEG: passi

- Conversione RGB --> YUV
- Sottocampionamento 4:4:4 --> 4:1:1 (**lossy**)
- Eliminazione della correlazione statistica mediante Discrete Cosine Transform (DCT) su blocchi di 8x8 campioni
 - 16x16 pixel --> 4 blocchi Y, 1 U e 1 V
- Quantizzazione dei coefficienti DCT (**lossy**)
- Codifica run-length
- Codifica di Huffman delle coppie run-length

20

Discrete Cosine Transform

- Permette il passaggio dal dominio spaziale a quello delle frequenze spaziali.
 - Sottocaso dello Sviluppo in Serie di Fourier
- In forma matematica, da matrice y a matrice z via moltiplicazione per la matrice che definisce la trasformata
 - $z = A y$ trasformazione diretta
 - $y = A^{-1} z$ trasformazione inversa
 - matrice z : matrice 8x8 dei coefficienti delle componenti cosinusoidali a varie frequenze
- Operazione **lineare** e **reversibile**
 - in teoria non introduce distorsione (**lossless**)
 - in pratica arrotondamento: y e z sono interi, A è float

21

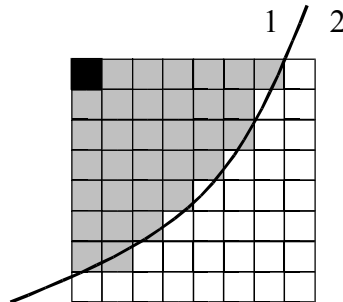
DCT su immagini

- Primo coefficiente --> frequenza nulla --> valor medio di luminosità (oppure U o V)
- Segnale poco ripido --> coefficienti nulli a frequenze elevate
- Segnale ripido --> alti coefficienti a frequenze elevate
- Le immagini naturali contengono molte zone di colore piatte o sfumate
 - molto spesso i coefficienti a frequenze elevate sono nulli
- I contorni degli oggetti introducono alti coefficienti a frequenze elevate

22

Matrice trasformata

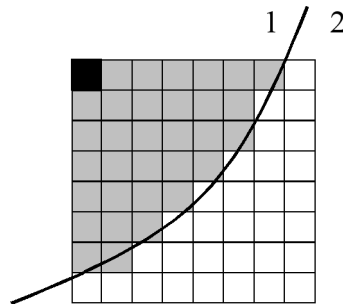
- La cella nera contiene la componente continua
- Le sfumature contribuiscono ai coefficienti nell'area 1
- I bordi invece ai coefficienti nell'area 2
- Spesso, per immagini naturali, l'area 2 contiene molti coefficienti piccoli o nulli.



23

Quantizzazione dei coefficienti DCT

- L'occhio è poco sensibile alla precisione dei coefficienti alle alte frequenze spaziali
- Tutti i coefficienti vengono divisi per delle costanti note al decodificatore
- Al risultato della divisione viene troncata la parte decimale
- Quantizzazione lineare

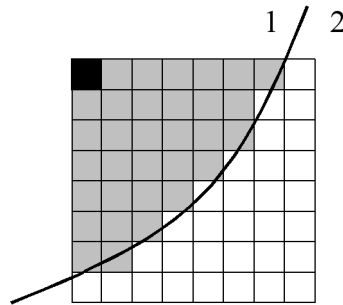


8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

24

Quantizzazione dei coefficienti DCT

- I già bassi valori dei coefficienti nell'area 2 diventano spesso nulli
- Maggiore è il divisore e maggiore è l'errore di quantizzazione (irreversibile)
- Introduzione di distorsione (effetto blocchi delle immagini JPEG molto compresse)



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

25

Effetto della quantizzazione dei coefficienti DCT



alto bitrate

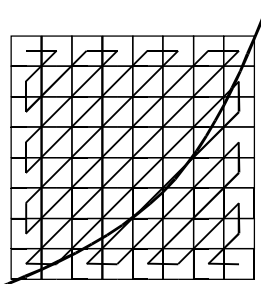


basso bitrate

26

Codifica Run-length dei coefficienti

- Percorso seguito per la scansione dei coefficienti
- Massimizza la probabilità di trovare sequenze di valori uguali
- Dopo la metà del percorso è molto probabile che ci sia un'unica sequenza di zeri che diventa un unico simbolo run-length (simbolo ad alta probabilità)



27

Codifica di Huffman

- Le coppie run-length (valore del coefficiente, ripetizioni) vengono codificate con codici binari a lunghezza variabile.
- Le coppie più frequenti sono rappresentate da codici più corti.

28

Compressione del video

29

Ridondanza spaziale e temporale

- Il Video si può assimilare ad una sequenza di immagini dette **frame**
 - in genere 20-30 frame/s per qualità film.
- La forma di codifica video più semplice consiste nel codificare i frame come immagini indipendenti
 - MotionJPEG --> ogni frame è un JPEG
- Nei video oltre alla correlazione SPAZIALE (DCT – JPEG) esiste anche la correlazione TEMPORALE (cioè tra frame diversi)
- Per eliminare la correlazione temporale si usa la **codifica differenziale con compensazione del moto**

30

Codifica differenziale con compensazione del moto

- Il primo frame viene codificato come in JPEG
 - frame indipendente e autoreferenziante
 - frame INTRA (**I-frame**).
- Per ciascun blocco 16 x 16 pixel (**macroblocco**)
 - Ricerca nel frame precedente dell'area d'immagine più simile al macroblocco (**motion estimation**)
- Codifica INTRA (come in JPEG) della differenza tra le 2 matrici (**residuo**) e codifica della posizione relativa tra le due aree (**vettore di moto**).

31

Motion estimation

- E' una fase computazionalmente intensa
 - consumo di tempo o necessita' di CPU potente
 - consumo di energia
- Si usano soluzioni SUB-OTTIME:
 - si riduce il range di ricerca
 - si compie la ricerca su versioni sottocampionate dei 2 frame (e quindi di dimensioni ridotte)

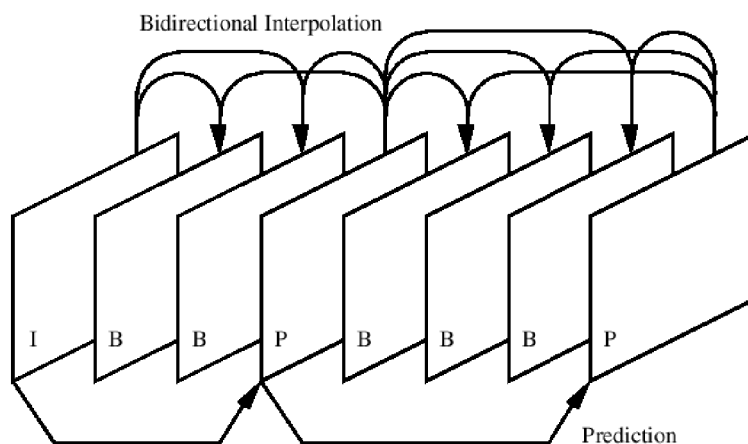
32

Tipi di frame

- Frame di tipo INTRA (I)
 - codifica JPEG indipendente dai frame vicini
- Frame di tipo PREDITTIVO (P).
 - codifica differenziale con compensazione del moto rispetto ad un frame precedente
 - dipendente da tale frame
- Frame di tipo BI-DIRECTIONAL (B).
 - codifica differenziale con compensazione del moto rispetto ad un frame precedente e ad uno successivo
 - aumenta la probabilità di avere residui bassi
 - dipendente da entrambi i frame
 - alta intensita' computazionale

33

Tipi di frame (2)



34

Display vs. Coding Order

- Sequenza di frame come viene catturata o riprodotta (**display order**):

I B P B P B P B P I

Frame # 0 1 2 3

- Ordine di codifica e trasmissione dei frame (**coding order**):

Frame # 0 2 1.....

- Si Codifica il frame 0, poi il frame2, dopodichè avendo informazioni su 0 e 2 si puo' codificare il frame 1 di tipo B. --> **RITARDO ALGORITMICO**

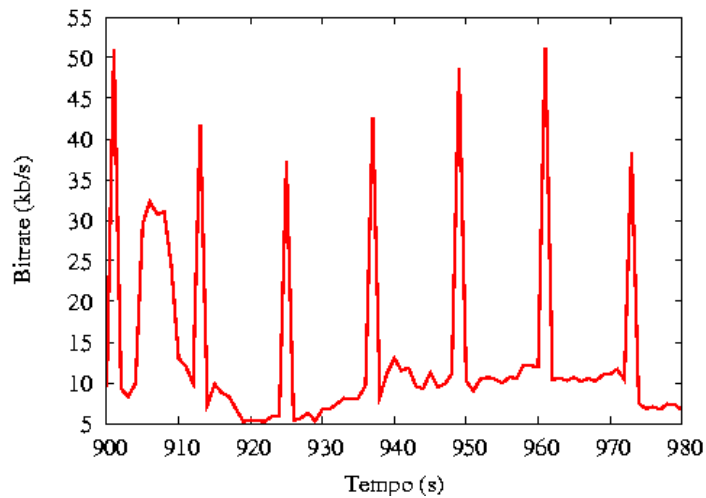
35

Andamento del bitrate

- I frame di tipo I hanno bit-rate più elevato rispetto agli altri perché non implementano la compensazione del moto.
- I frame di tipo B invece hanno il bit-rate più basso perché implementano una doppia compensazione del moto.
- I frame di tipo P hanno una dimensione intermedia.
- La trasmissione di video compresso genera traffico VBR.

36

Trasmissione di video compresso



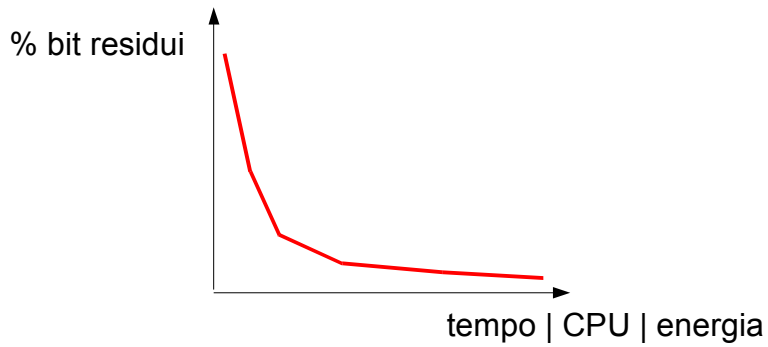
37

Intensita' computazionale

- E' il reciproco dell'andamento del bit-rate !
- I frame di tipo I hanno ridotta complessità perché non c'è compensazione del moto.
- I frame di tipo B invece hanno elevata complessità perché effettuano una doppia compensazione del moto.
- I frame di tipo P sono una via di mezzo (effettuano una sola compensazione del moto).
- Occorre tenere conto di tale variabilità nel progetto dell'architettura del codificatore (hardware, software).

38

Intensita' computazionale (2)



39

Ritardo

- Il ritardo è MINIMO per tutti i frame di tipo I
- MEDIO per i frame I + P
- ALTO per i frame I + P + B dovuto al ritardo algoritmico (quindi indipendente dalla potenza della CPU !)
- Per applicazioni interattive (videoconferenza) si usano solo frame I e P
- Frame di tipo B si usano per lo streaming

40

Standard di compressione video

- ITU-T H.261
 - videoconferenza
- ISO MPEG-1
 - VideoCD (~1.4 Mb/s)
- ISO MPEG-2 / ITU-T H.262
 - DVD (~3 Mb/s)
 - DVB: digitale terrestre e satellite (~2 Mb/s)
- ITU-T H.263
 - videoconferenza
- ISO MPEG-4
 - DivX, XviD (700 kb/s – 5 Mb/s quasi lossless)
- ITU-T H.264 **stato dell'arte !!!**