



# Lezione 10 - Indicizzazione e recupero del video

<https://www.youtube.com/watch?v=gEXthIUCD0I>

## Metodologie di indicizzazione:

- Per metadati → Informazioni strutturate in RDBMS tradizionali (oggettivi e incompleti)
- Per testo → Sottotitoli o trascrizioni dei dialoghi trattati con tecniche di IR standard
- Audio → Segmentazione in parlato e non parlato, con tecniche di speech-recognition
- Contenuto → 2 metodi:
  - Singolo fotogramma → Video considerato come una serie di immagini indipendenti
  - Gruppi di fotogrammi → Frame suddivisi in “shots”. L’indicizzazione è basata sui frame rappresentativi di ogni shot

I migliori risultati derivano con la combinazione di due o più metodi

## Indicizzazione e ricerca basata sugli SHOT

Uno shot è un gruppo di frame contigui con delle caratteristiche comuni:

- Fanno parte della stessa scena
- Non interrotti da uno stacco di telecamera
- Contengono lo stesso evento/azione

Sono necessarie anche tecniche di segmentazione dei video

## Segmentazione automatica

Per verificare passaggi tra shot occorre misurare la differenza fra 2 frame consecutivi

- Metodo 1 → Si calcola la somma, pixel per pixel, delle differenze tra i 2 frame consecutivi
  - Non garantisce buoni risultati poiché la presenza di oggetti in movimento causa grandi differenze anche all'interno della stessa scena
- Metodo 2 → Si calcola la differenza tra gli istogrammi di colore di 2 frame consecutivi:
  - Basato sul fatto che il movimento di oggetti all'interno della stessa scena causa piccole variazioni nell'istogramma dell'immagine

Definiamo la distanza tra frame:

$$SD_i = \sum_j |H_i(j) - H_{i+1}(j)|$$

in cui:

- $H_i(j)$  → Valore del j-esimo bin dell'istogramma dell'i-esimo frame

Se  $SD_i$  supera una soglia prefissata allora assumiamo che si sia verificato un cambio di shot

## Shot e Transizioni

Quando tra una scena e l'altra di un video sono utilizzate transizioni la differenza di frame per frame è sempre molto bassa

Sarebbe necessario usare una soglia molto più bassa ma con il rischio di trovare molti cambi di shot non accurati

## Segmentazione a due soglie

Introduciamo due soglie diverse: una più alta  $T_b$ , per determinare i cambi di camera, e una più bassa  $T_s$ , per determinare i frame nei quali avviene una transizione graduale

Confrontando il un frame con il successivo, se la differenza è maggiore di  $T_b$  si introduce un cambio di shot, se è minore di  $T_b$  ma superiore a  $T_s$  il frame è dichiarato come un potenziale frame di transizione

Fino a quando si hanno potenziali frame di transizione consecutivi, la loro distanza viene sommata e quando raggiunge la soglia  $T_b$  si genera un cambio di shot

## Problema di zoom e panning

Le operazioni di panning e zoom generano dei cambiamenti gradualmente che possono portare all'individuazione di falsi shot

Per limitare questo fenomeno si possono introdurre tecniche di analisi del movimento di blocchi di pixel tra frame successivi

- Il computer capisce quando c'è uno zoom nel momento in cui tutti i punti intorno all'oggetto ingrandito tendono a spostarsi verso l'esterno
- Il panning viene rilevato nel momento in cui tutti i punti dell'immagine si spostano nella stessa direzione

## Cambiamenti di illuminazione

Per evitare falsi cambi di shot dovuti al cambio di illuminazione, si utilizzano diversi step di elaborazione, e si applicano operazioni di normalizzazione ai colori

- Normalizzazione:

$$R'_i = \frac{R_i}{\sqrt{\sum_{i=1}^N R_i^2}}$$

$$G'_i = \frac{G_i}{\sqrt{\sum_{i=1}^N G_i^2}}$$

$$B'_i = \frac{B_i}{\sqrt{\sum_{i=1}^N B_i^2}}$$

- Conversione cromatica:

$$r_i = \frac{R'_i}{R'_i + G'_i + B'_i}$$

$$g_i = \frac{G'_i}{R'_i + G'_i + B'_i}$$

## Utilizzo degli Shot

Dopo aver individuati gli shot occorre rappresentarli ed indicizzarli in modo da essere gestiti dalla fase di ricerca

Per ogni shot si determinano uno o più frame rappresentativi che verranno utilizzati sia per l'indicizzazione che per la ricerca

- Prendere un solo r-frame
  - Non sempre è significativo
- Prendere N r-frame al secondo
  - Gestisco la lunghezza ma non il contenuto, e potrebbero essere sovrabbondanti nel caso di uno shot uniforme
- Suddivisione in sotto-shot
  - Usando le stesse tecniche per determinare gli shot, si introduce un nuovo livello e si genera un r-frame per ogni sotto-shot

## Individuazione degli r-frame

- Metodo 1 → Si considera come r-frame il primo frame dello shot (o sotto-shot)
- Metodo 2 → Si calcola un frame “medio”, facendo la media dei colori di tutti i frame dello shot pixel per pixel
- Metodo 3 → Si calcola la media di tutti gli istogrammi dei frame dello shot e scelgo come r-frame il frame il cui istogramma è più vicino all'istogramma medio

## Tecniche basate sulle informazioni di movimento

Tecniche che memorizzano le informazioni derivanti dai “motion vectors”:

- Direzione prevalente di movimento
- Uniformità di movimento
- Panning
- Zoom

Ogni shot viene indicizzato attraverso i motion vectors che lo caratterizzano

## Rappresentazione del video

## Presentazione video tramite MICON

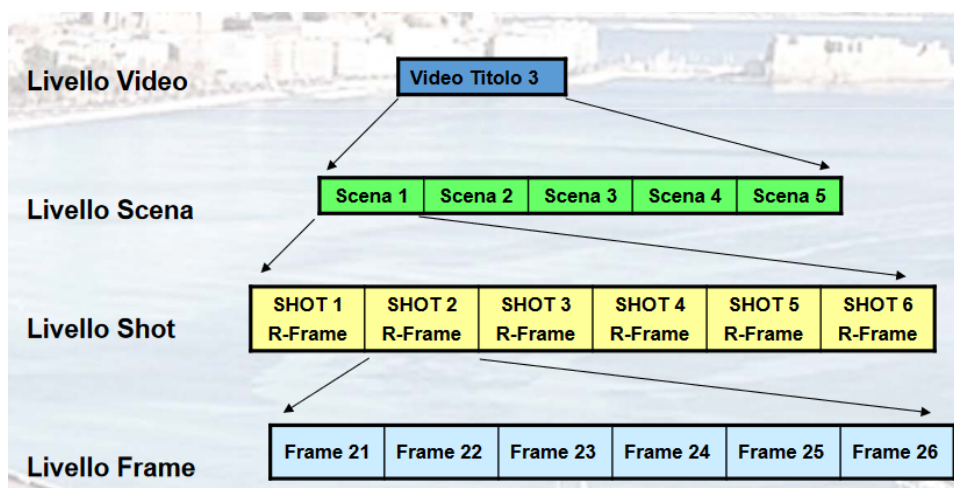
- MICON → Motion Icon → Rappresentazione compatta degli shot

Un MICON è caratterizzato da tre dimensioni:

- Un r-frame
- Profondità → Durata temporale
- Pixel presenti sui bordi orizzontali e verticali → Danno un'idea del movimento

## Browser video gerarchico

Strumento che consente di ispezionare in maniera gerarchica i vari elementi di un video

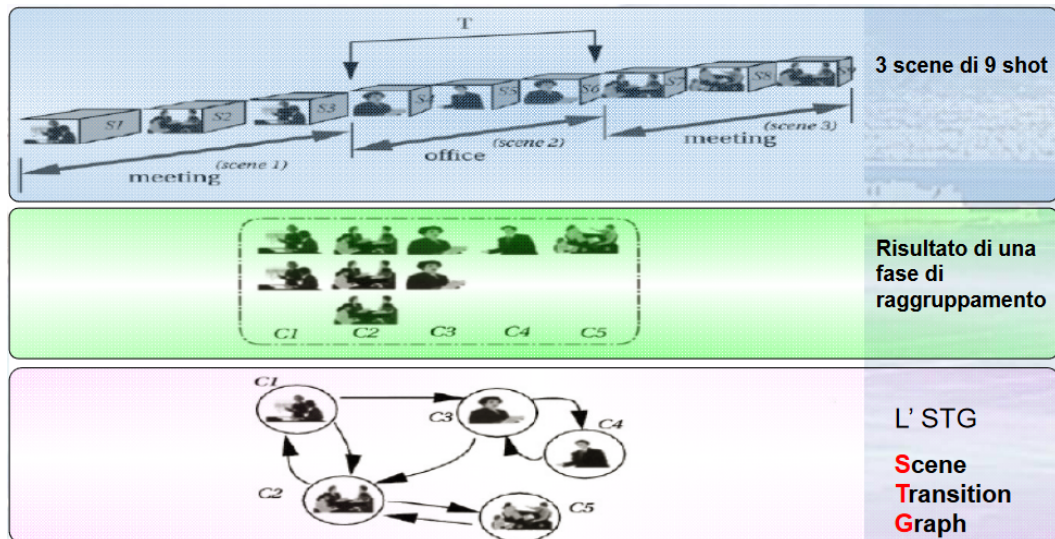


## Rappresentazione compatta

- Storyboard → Collezione di frame rappresentativi
- Mosaicatura → Unione di frame diversi che nel loro insieme descrivono un elemento complesso
- Scene Transition Graph → Struttura a grafo che cattura sia il contenuto che il flusso temporale di un video

## Scene Transition Graph (STG)

Consiste in una rappresentazione di un video mediante un grafo orientato  $G = (V, E)$ , in cui  $V$  sono i vertici (ogni vertice rappresenta un cluster di shot simili),  $E$  sono gli archi (Due vertici sono connessi da un arco se esiste uno shot in un cluster  $i$  che precede uno shot in un cluster  $j$ )



## Tipi di query e vettori

- Point Query → Query rappresentata da un vettore, che ricerca un oggetto nel DB che sia simile o identico a quello descritto dalla query
- Range Query → Rappresentata da un vettore e un range di distanza. Ritorna tutti gli oggetti la cui distanza dalla query è inferiore o uguale a quella specificata
- K-nearest Neighbour → Query rappresentata da un intero k: i k oggetti con le distanze più piccole dal vettore di query sono ritrovati

## Riduzione dello spazio di ricerca tramite filtraggio

Si introducono dei criteri per focalizzare la ricerca solo in alcune “zone” dello spazio multidimensionale

Un filtraggio efficiente consente di escludere dal confronto una gran parte dei vettori multidimensionali

### Classificazione

Utilizza informazioni associate ai dati multimediali per ridurre lo spazio di ricerca (es: L'utente può essere interessato ad un intervallo di date di produzione o a prodotti di un autore/regista)

### Disuguaglianza triangolare

L'obiettivo fondamentale del filtraggio è la riduzione del numero dei confronti (tra la query e gli N vettori) → Partendo dagli N vettori presenti nel database, solo un sottoinsieme di m vettori è scelto come base di confronto

Viene applicata la disuguaglianza triangolare inversa, secondo la quale la differenza tra le lunghezze di due lati è minore della lunghezza del lato rimanente