Sistemi di Elaborazione dell'informazione II

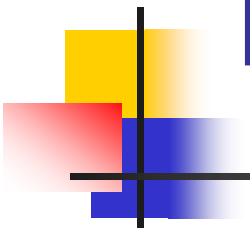
Corso di Laurea Specialistica in Ingegneria Telematica
II anno – 4 CFU
Università Kore – Enna – A.A. 2009-2010

Alessandro Longheu
http://www.diit.unict.it/users/alongheu
alessandro.longheu@diit.unict.it

Multimedia Information Retrieval

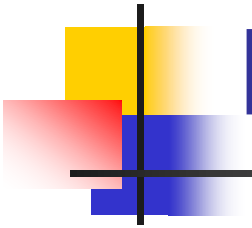
Multimedia IR

- *"Information Retrieval on the Web"* - KOBAYASHI and TAKEDA ACM Computing Surveys, Vol. 32, No. 2
- IR from multimedia databases (Multimedia IR) is a multidisciplinary research area, which includes topics from a very diverse range, such as analysis of text, image and video, speech, and nonspeech audio; graphics; animation; artificial intelligence; human-computer interaction; and multimedia computing
- **Query and retrieval of images** is one of the more established fields of research. Topics in this area include:
  - **search and retrieval from large image archives**
  - pictorial queries by **image similarity**
  - acquisition, storage, indexing, and retrieval of **map images**
  - **real-time fingerprint matching** from a very large database
  - querying and retrieval using **partially decoded JPEG data**
  - **retrieval of faces** from a database



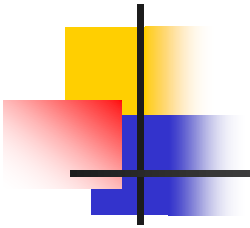
Multimedia IR

- **Finding documents that have images of interest is a much more sophisticated problem.** The number of retrievals for Photofinder website for instance were huge, but there was a considerable amount of noise after the first page of retrievals and there were many redundancies.
- **Query and retrieval of images in a video frame** or frames is a research area closely related to retrieval of still images from a very large image database. Several proposals are present, e.g. VisualSEEk, a tool for searching, browsing, and retrieving images, which allows users to query for images using visual properties of regions and their spatial layout.
- Other interesting research topics and applications in multimedia IR are **speech-based IR** for digital libraries and **retrieval of songs** from a database when a user hums the first few bars of a tune; the melody retrieval technology has been incorporated as an interface in a karaoke machine.



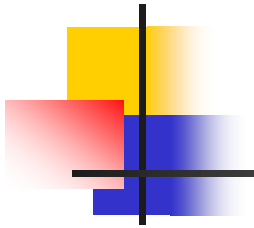
Multimedia IR

- Nel Web e nei DB "locali" esiste molta informazione non esclusivamente testuale:
  - audio (speech, musica…)
  - immagini,
  - video,
  - …
- Come recuperare efficacemente materiale multimediale?



Esempi di applicazione

- Web indexing:
  - Recupero di materiale multimediale dal Web,
  - Sistemi capaci di bloccare immagini pubblicitarie oppure illegali/indesiderate
- Trademark e copyright
- Accesso ai musei
- DB commerciali

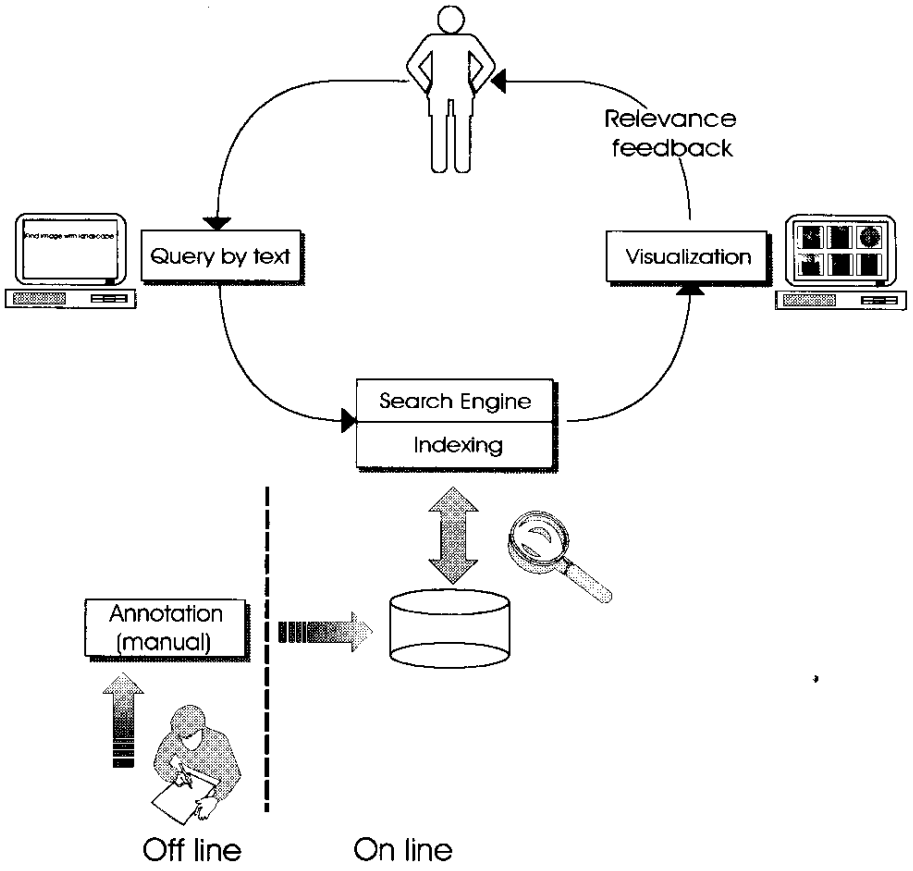


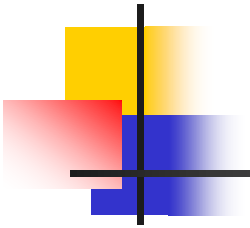
Esempi di applicazione

- Immagini satellitari (applicazioni militari, government, …)
- Immagini mediche
- Entertainment
- Criminal investigation
- …

Prima generazione di multimedia information retrieval systems

- Off-line: il materiale multimediale viene associato con una descrizione testuale (annotazione manuale) "content descriptive metadata"
- On-line: utilizzo di tecniche di IR testuali basate sul "keyword match"





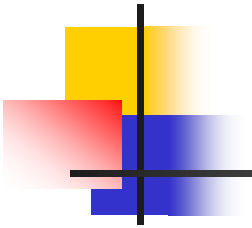
Limiti dell'approccio esclusivamente testuale

- L'annotazione manuale di grossi DB multimediali è impensabile…
- Non è facile descrivere testualmente il contenuto percettivo di un'immagine, di un video o di un brano musicale…
- Ad esempio:
  - Cercare una canzone conoscendone solo il ritornello
  - Cercare una determinata azione in un video sportivo
  - Cercare dipinti che hanno un determinato dosaggio di colori

Limiti dell'approccio esclusivamente testuale

- Google Image può restituire fino ad un 85% di documenti NON rilevanti anche per query semplici (e.g., oggetti specifici) [1]

- [1] Fergus, Fei-Fei, Perona, Zisserman, Learning Object Categories from Google's Image Search, ICCV 05



Soluzioni in fase di studio

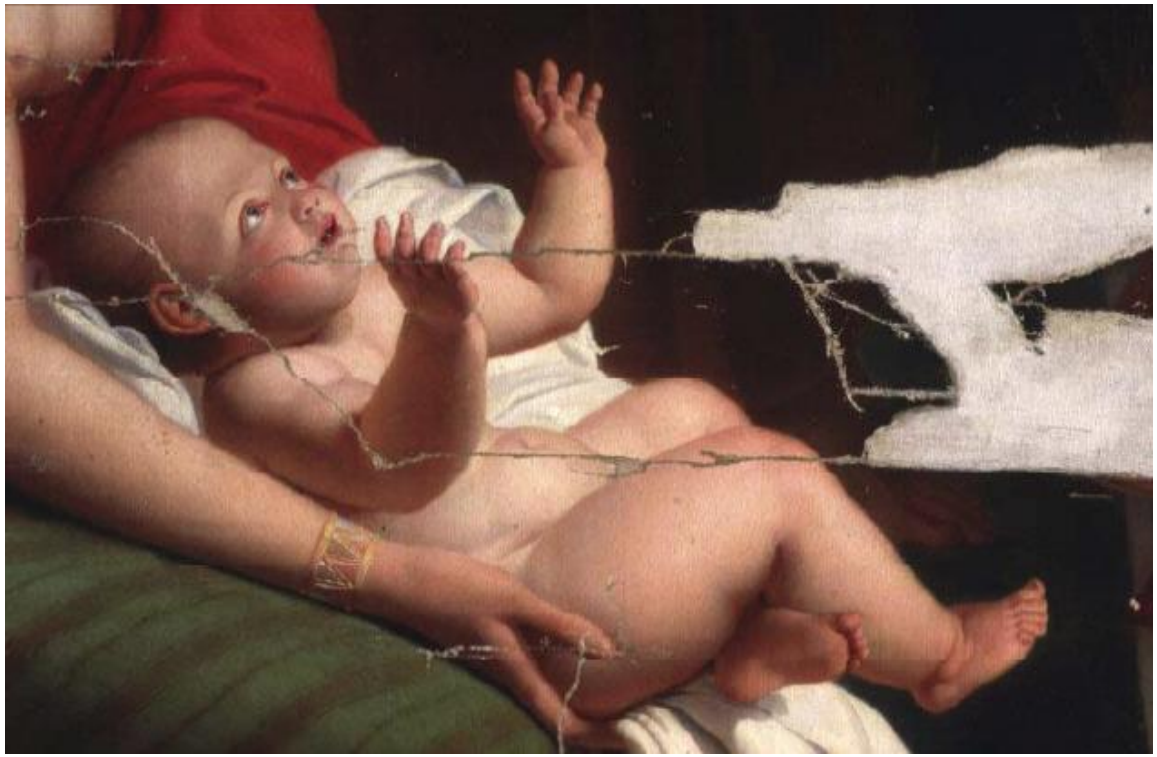
- **Sistemi "Content Based", CBIR**:
  - Modellano direttamente sia la query che gli oggetti del DB in uno spazio percettivo visivo/auditivo
- Sistemi di annotazione automatica:
  - Fase di pre-processing ("information extraction"): viene estratta informazione da elementi non testuali e memorizzata in maniera testuale o simbolica; il retrieval (on-line) avviene con tecniche "tradizionali"

Elementi tipici di un sistema di **CBIR** (Content Based IR)

- Dal punto di vista dell'utente:
  - La query è un oggetto multimediale (e.g., un'immagine, un disegno, un elemento audio, …)
  - L'output è una lista di elementi ordinati in base alla somiglianza percettiva con la query
  - Esistono strumenti opzionali di interazione per visualizzare collezioni di immagini o fornire feedback al sistema

Esempio: query by image example

La query è un particolare

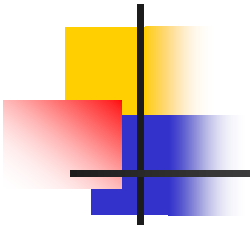


Esempio: query by image example

Query e particolare possono **non** essere identici.
Ad es. la query può essere scelta da un' immagine prima di un restauro

Esempio: query by image example

immagine presa da: A. Del Bimbo, Visual Information Retrieval

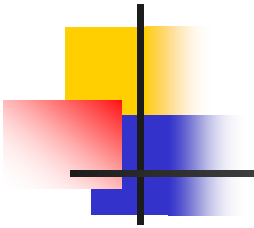


Elementi tipici di un sistema di CBIR

- Dal punto di vista del sistema:
  - **Rappresentazione** dell'oggetto multimediale (e.g., spazio delle feature)
  - Modellazione del concetto di **similitudine** percettiva (e.g., attraverso appositi algoritmi di matching)
  - Utilizzo di strutture dati particolari che permettano un **indexing** efficiente dello spazio delle feature
  - Gestione dell'eventuale **relevance feedback**

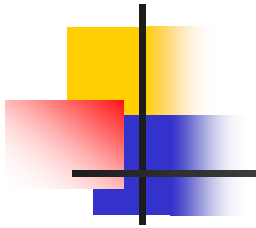
Rappresentazione tramite Feature di un'immagine

- Una feature è una rappresentazione, tramite valori numerici, di tutta o parte dell'immagine. In genere una feature è una caratteristica facilmente misurabile dell'immagine (ad esempio istogramma dell'intensità dei pixel in un colore dato)
- L'immagine in esame viene quindi descritta usando i valori di un insieme di feature pre-scelte
- Se I' = I, la feature è globale, mentre se I' $\subset$ I la feature è locale
- Per le feature locali, è importante capire come selezionare le sottoparti da rappresentare, ma sono più robuste ad occlusioni, parti mancanti, separazione dal background
- Lo spazio delle feature è un concetto simile allo spazio vettoriale usato per rappresentare la frequenza dei termini nei documenti testuali
- E' la rappresentazione dei dati più usata nei vari sistemi content based (immagini, video, audio) ma non è l'unica



Similitudine percettiva

- Nel text retrieval, la similitudine tra due documenti può essere stimata contando i termini in comune
- Ad esempio, rappresentando la frequenza dei termini di un documento in un apposito spazio vettoriale, la distanza tra punti in tale spazio descrive la dissimilarità tra documenti
- Lo spazio delle feature si comporta in maniera analoga…
- Nello spazio delle feature la differenza percettiva tra I_1 e I_2 è proporzionale ad una determinata misura di distanza (non necessariamente Euclidea): $dist(x(I_1),x(I_2))$
- Data la query Q, l'output del sistema è una lista di immagini I_1, I_2, … ordinata massimizzando $dist(x(Q),x(I_j))$

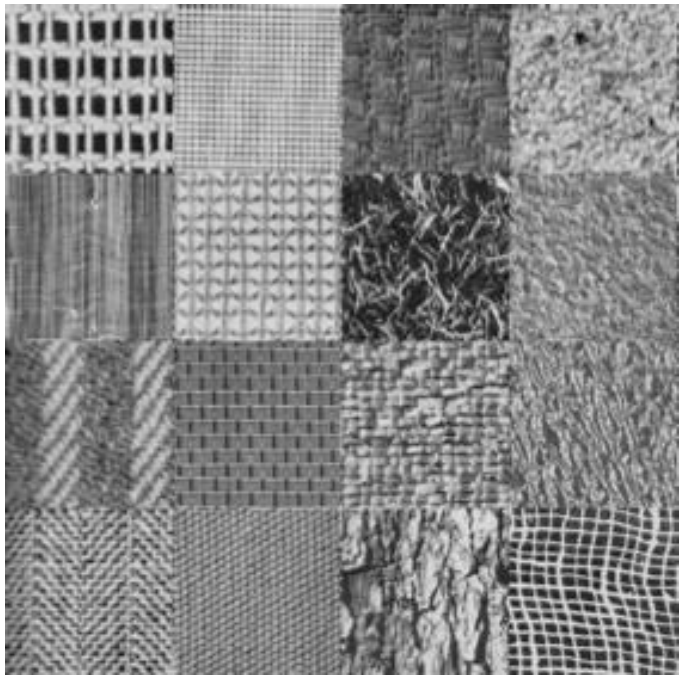


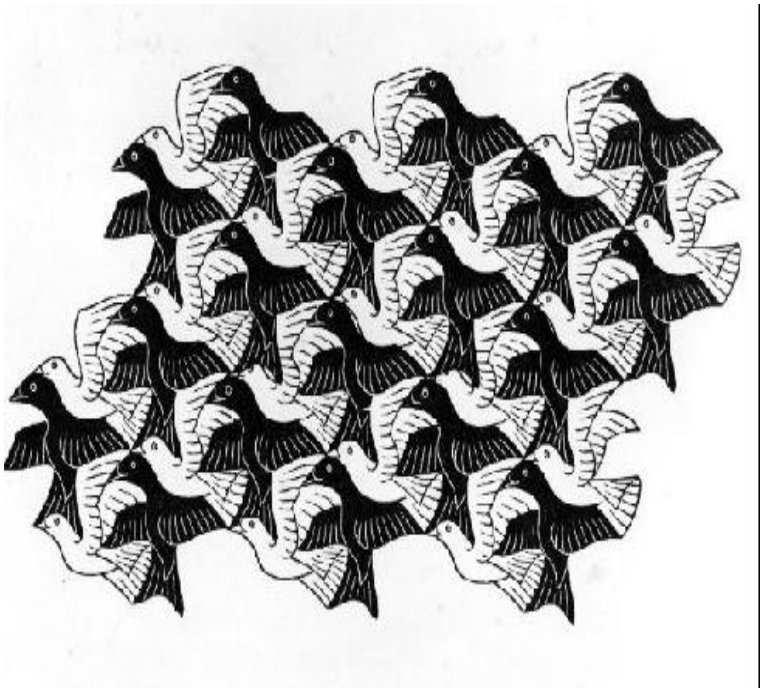
Indexing

- Problema: come indicizzare efficientemente dei dati in uno spazio multidimensionale?
- La maggior parte delle più comuni strutture dati per la ricerca efficiente si basano su un ordinamento totale dei valori rappresentati:
  - $x_i <= x_j \ V \ x_j <= x_i \ (0 <= i,j <= N)$
- Ad esempio, nel caso delle keyword l'ordine alfabetico stabilisce un ordinamento totale
- In R^k ciò non è più vero, si adotta allora in genere il k-d tree, una generalizzazione dell'albero di ricerca binario in *k* dimensioni nel quale ad ogni livello dell'albero viene presa in considerazione ciclicamente una delle *k* feature

Image, video e audio retrieval

…in questo caso si adottano tecniche specifiche, ad esempio retrieval by texture:



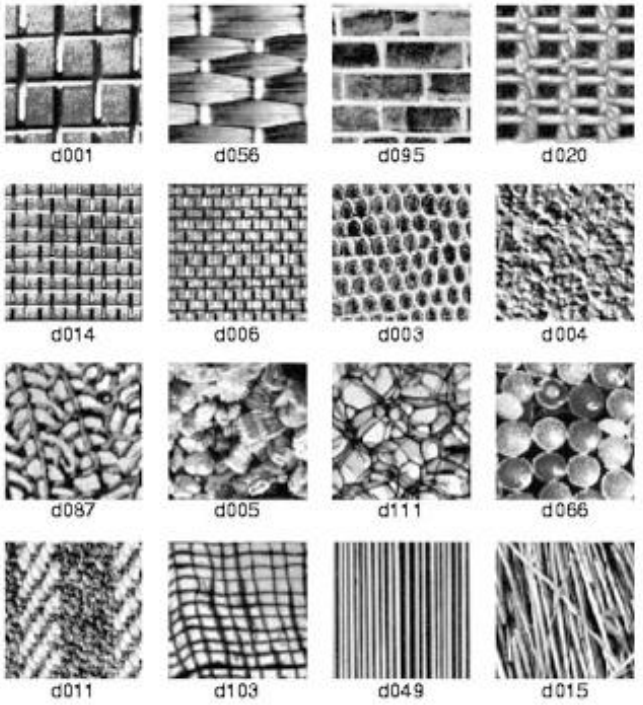


Image, video e audio retrieval

- Si segue in genera un approccio statistico che rileva la similitudine con texture campione:

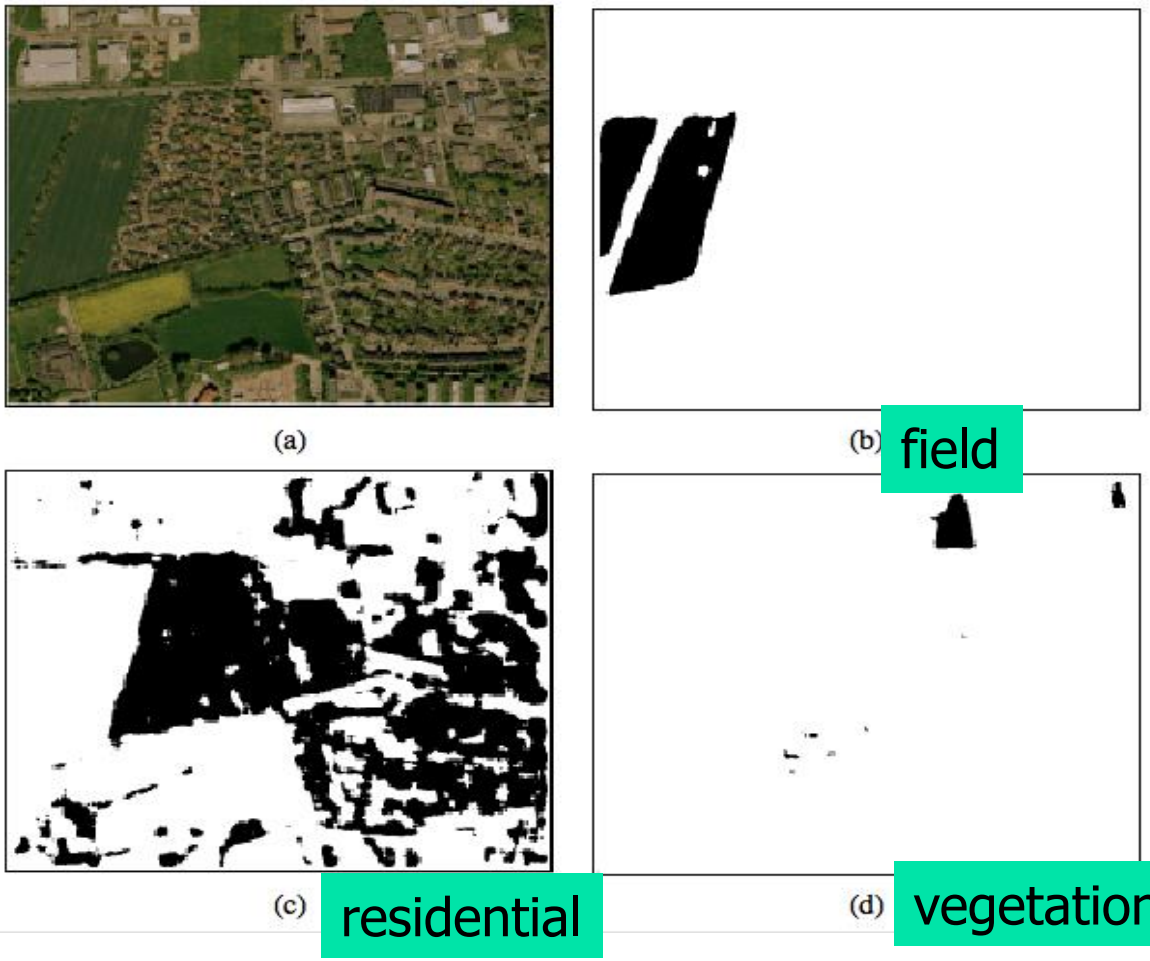
Le features più utilizzate sono le Tamura features:

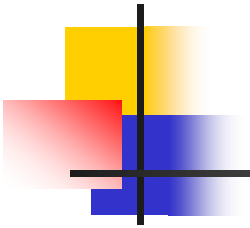
- Contrasto
  - ➔ **Distribuzione dell'intensità dei pixel**
- Coarseness
  - ➔ **Granularità della tessitura**
- Direzionalità
  - ➔ **Direzione dominante della tessitura**



Image, video e audio retrieval

- Le texture campione permettono ad esempio di selezionare solo porzioni di un'immagine





Image, video e audio retrieval

- Un approccio alternativo è quello sintattico, che fa uso di grammatiche (cosiddette) visive, ad esempio una regola di produzione per la rappresentazione di texture (Rosenfeld):

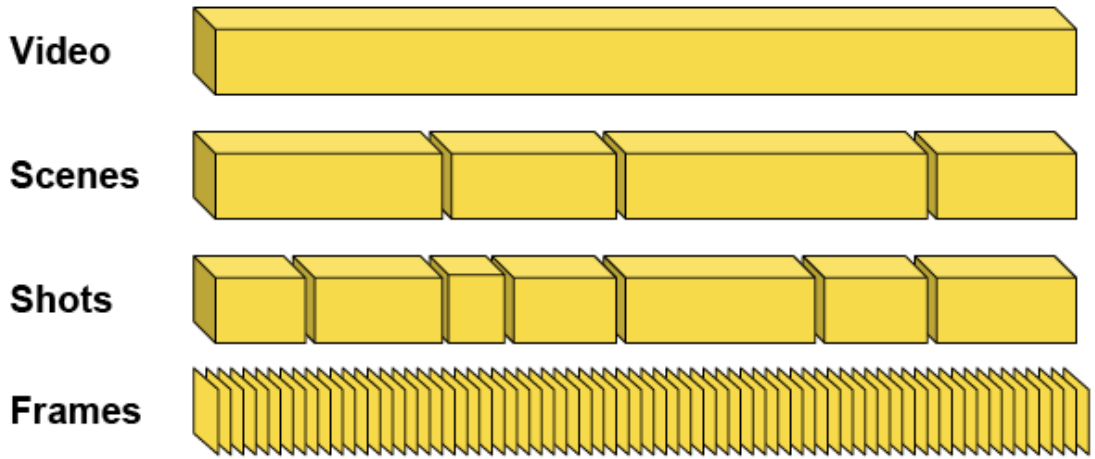
$$0_x b \rightarrow 0_x 1, \quad x \in \{U, D, L, R\}$$

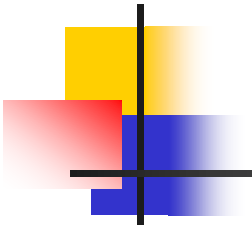
- Essa riassume le seguenti 4 regole:

$$\begin{matrix} 0 \\ b \end{matrix} \rightarrow \begin{matrix} 0 \\ 1 \end{matrix}, \quad \begin{matrix} b \\ 0 \end{matrix} \rightarrow \begin{matrix} 1 \\ 0 \end{matrix}, \quad 0b \rightarrow 01, \quad b0 \rightarrow 10$$

Image, video e audio retrieval

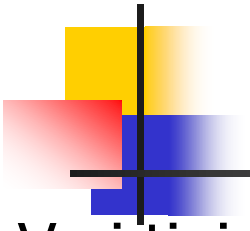
- Un video è una sequenza di immagini, ognuna detta **frame**. Lo **shot** è una sequenza di frame consecutivi ripresi da una singola telecamera.
- **Scena**: un insieme di shot consecutivi che rispettano le 3 unità aristoteliche di spazio, tempo e azione





Image, video e audio retrieval

- Il procedimento prevede la segmentazione di video: individuando gli "editing effects" (cuts, dissolvenze, ….) tra uno shot e l'altro è possibile suddividere (automaticamente) un video in shot
- Molto più difficile è individuare le scene (concetto semantico)
- I video possono essere rappresentati con dei "key frame" rappresentativi di ogni shot. Un key frame è trattabile come una "still image": Si applicano quindi le tecniche note per le immagini (spazio delle feature, ricerca per colore, texture, forma, ecc.)
- Alternativamente, è possibile cercare in un video informazione relativa al movimento (e.g., una particolare traiettoria in un video sportivo, …)



Image, video e audio retrieval

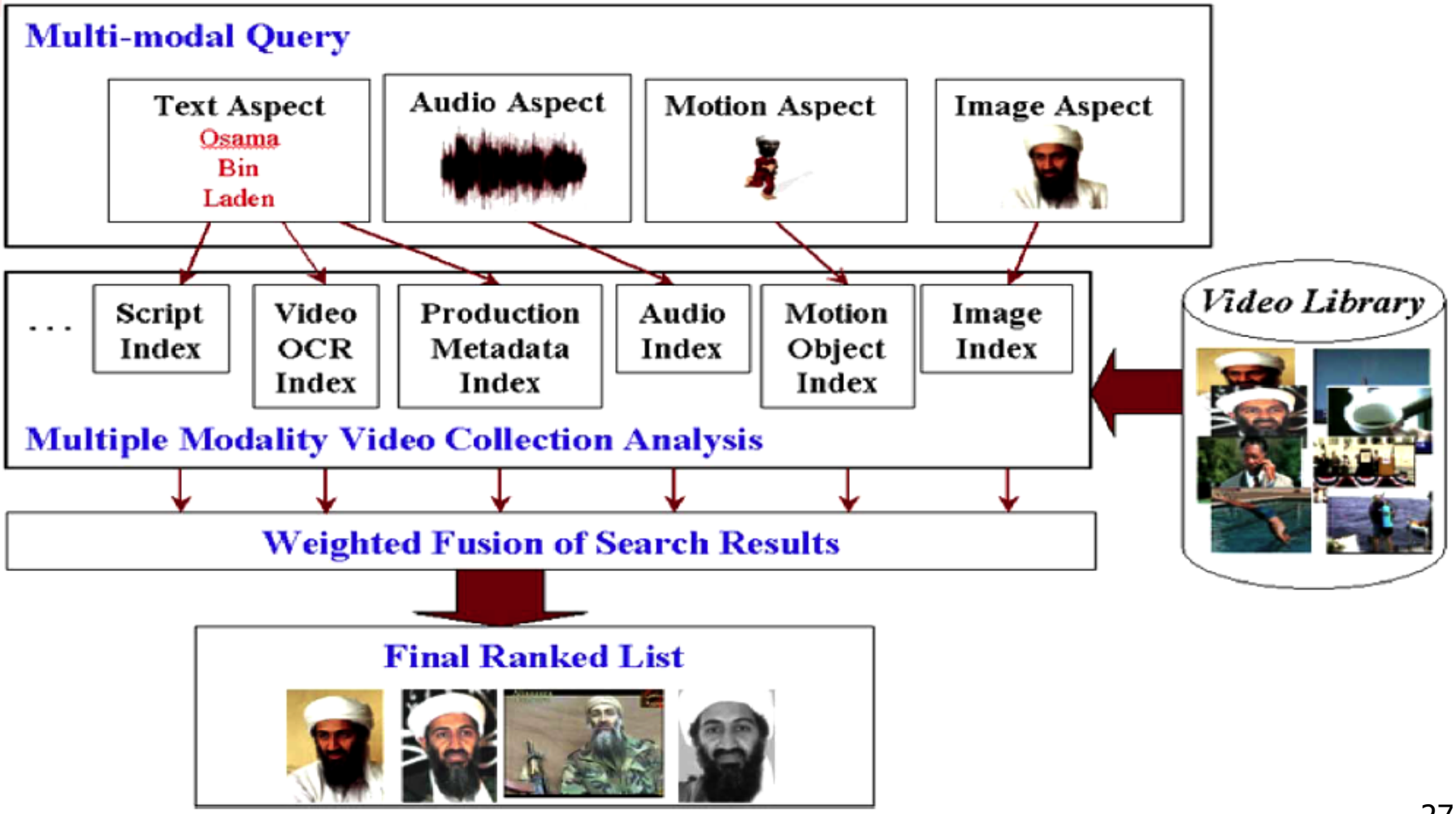
Vari tipi di audio:

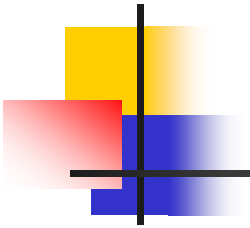
- Parlato:
  - E' possibile utilizzare tecniche di speech recognition per trasformare l'audio in testo
- Suono:
  - Un qualunque segnale audio con frequenze nel range dell'udito umano (e.g., suoni prodotti da animali…)
- Musica:
  - Si tiene conto dei diversi strumenti musicali utilizzati, dei vari tipi di suoni prodotti, degli effetti musicali, ecc.
- Per eseguire le query, si può operare con le Query by example: viene fornito/sintetizzato un file audio chiedendo di recuperare audio simili, oppure si opera con Query by humming: l'utente accenna (vocalmente o anche fischiando…) la melodia da cercare

Rappresentazione e similitudine

- Relativamente alla rappresentazione e similitudine, uno spazio delle feature può essere creato utilizzando, ad esempio istogrammi ricavati dalla rappresentazione spettrale del segnale
- La similitudine percettiva in tal caso avviene calcolando la distanza (Euclidea, di Mahalonobis, ecc.) tra punti multidimensionali come nel caso delle immagini

Aspetti percettivi diversi possono essere combinati





Image, video e audio retrieval

- A. Del Bimbo, Visual Information Retrieval, Morgan Kaufmann Publishers, Inc. San Francisco, California", 1999
- Forsyth, Ponce, Computer Vision, a Modern Approach 2003
- Long et al., Fundamentals of Content-based Image Retrieval, in: D. D. Feng, W. C. Siu, H. J. Zhang (Ed.),Multimedia Information Retrieval & Management-Technological Fundamentals and Applications, Springer-Verlag, New York(2003)
- Foote et al., An Overview of Audio Information Retrieval, ACM Multimedia Systems, 1998
- Hemant M. Kakde, Range Searching using Kd Tree, 2005