

Informatica**U**manistica

# Lezione 4

## Gestione del video

*Pasquale Savino*

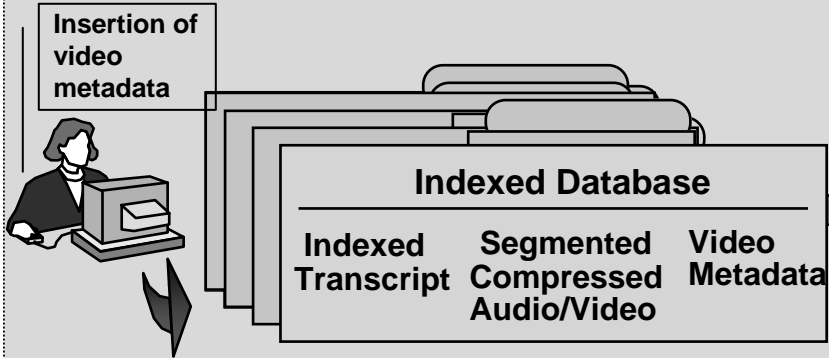
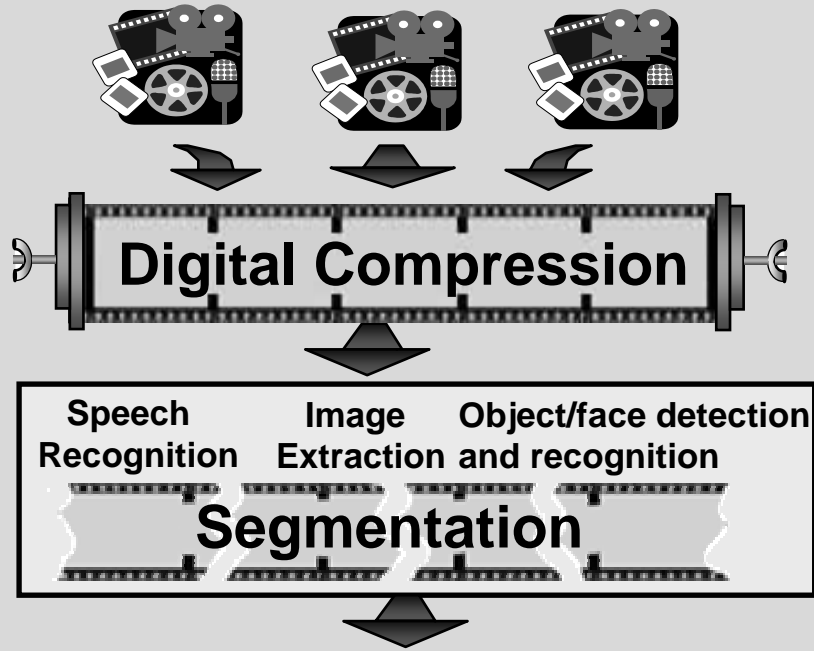
*ISTI - CNR*



UNIVERSITÀ DI PISA

# Library Creation

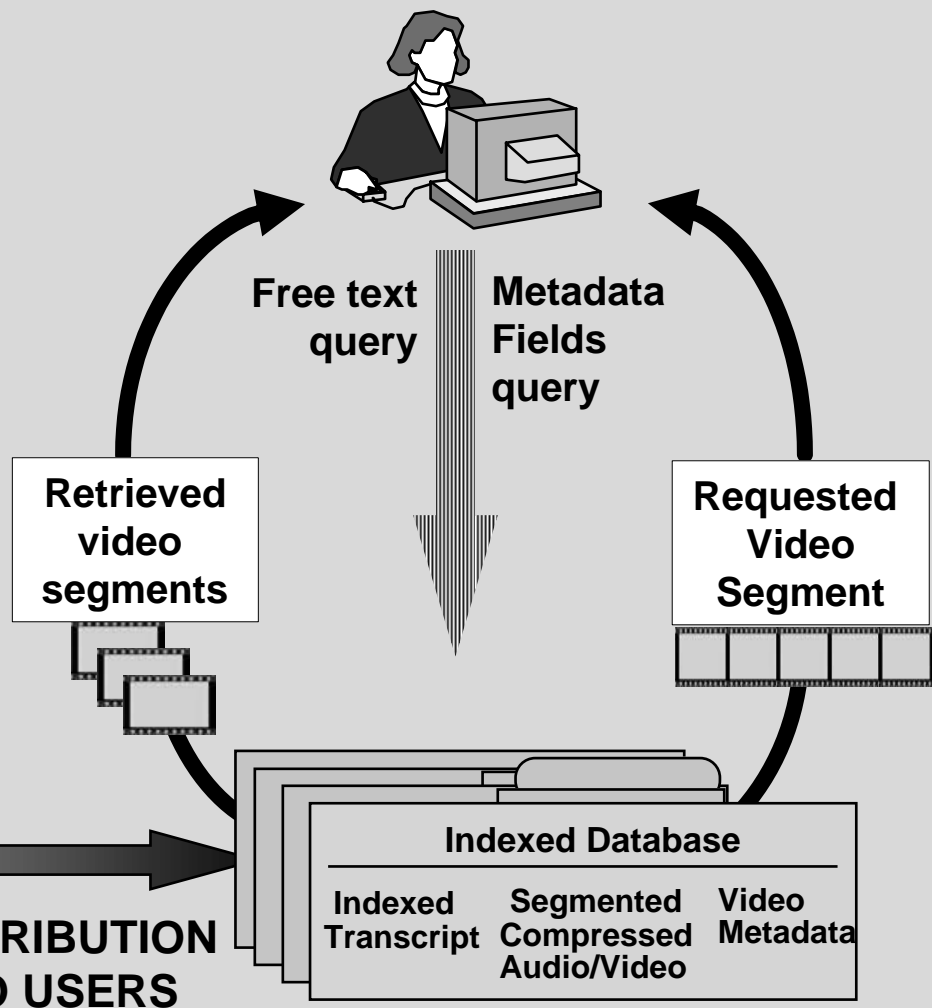
## Offline



**DISTRIBUTION TO USERS**

# Library Exploration

## Online



# Tipi di dati presenti nel video

- ◆ **Un video digitale è composto di una sequenza di immagini (frame) che vengono visualizzate con un determinata frequenza, più (opzionalmente) una traccia audio.**
  
- ◆ **Il video può essere analizzato da diversi punti di vista, che possono essere anche combinati tra loro**
  - La parte audio può essere analizzata individuando le parti che contengono
    - **Parlato**
    - **Suono/musica**
  - Si può analizzare la sequenza dei frame, quindi l'evoluzione temporale del video e la sua organizzazione in shot e sequenze.
  - Si possono considerare i singoli frame come immagini
  
- ◆ **Da tutte queste componenti è possibile estrarre informazione da utilizzare per l'indicizzazione e per la ricerca del video.**

# Caratteristiche del video digitale

- ◆ **Il video può essere descritto come una sequenza di frame visualizzati con una determinata frequenza, che varia a seconda della qualità del video e del formato di rappresentazione utilizzato.**
  - NTSC 30 frames/sec, PAL 25 f/s, HDTV 60 f/s
  - Nella maggior parte dei casi vi è una piccola variazione tra un frame ed il successivo.
- ◆ **Ogni frame è un'immagine con una risoluzione che dipende dalla qualità del video e dal formato**
  - 768 x 576 PAL, 720 x 480 NTSC
- ◆ **Da questi dati si deduce che un video non compresso occupa molto spazio su disco e richiede una elevata banda di comunicazione durante la trasmissione**
  - Per esempio, un secondo di video PAL non compresso richiede  
 $768 \times 576 \times 16 \times 25 \sim 172$  MByte

# Memorizzazione e trasmissione di video digitale [1/3]

- ◆ **La dimensione dei file video impone, in molti casi, di comprimere il video sia per memorizzarlo che per trasmetterlo.**
  
- ◆ **Nel caso del video si possono ottenere livelli di compressione molto alti, a causa delle caratteristiche del video stesso:**
  - Correlazione spaziale: in un singolo frame è possibile individuare una correlazione tra pixel adiacenti (come avviene nelle immagini), per cui è possibile utilizzare le tecniche di compressione delle immagini
  - Correlazione temporale: normalmente ogni frame differisce poco dai frame che lo seguono e lo precedono. Molto spesso il valore di un pixel rimane invariato per diversi frame successivi.
  - Una parte significativa del video non viene percepita da chi guarda il video.

# Memorizzazione e trasmissione di video digitale [2/3]

- ◆ **Le tecniche di compressione si possono dividere in due grandi categorie**
  - **Lossless compression** (compressione senza perdita) che permette di comprimere e decomprimere il video senza alcun degrado della qualità
    - **La *lossless compression* permette di ottenere livelli di compressione bassi**
    - **Si ottiene ad esempio se ogni frame viene compresso in JPEG senza perdita. Questa tecnica di compressione si chiama MJPEG (Motion JPEG). In questo caso si sfrutta solo la correlazione spaziale.**
    - **Altri esempi di metodi di compressione senza perdita sono run-length coding e Huffman coding**

# Memorizzazione e trasmissione di video digitale [3/3]

- **Lossy compression**, (compressione con perdita). In questo caso l'intero ciclo di compressione/decompressione introduce produce un video diverso dall'originale. L'obiettivo consiste nell'individuare delle tecniche di compressione che non introducono un degrado percepibile
  - ➔ **La compressione con perdita permette di ottenere livelli di compressione molto alti**
  - ➔ **Esempi di standard di compressione con perdita sono quelli della famiglia MPEG (MPEG1, MPEG2)**

# MPEG

## ◆ MPEG (Moving Pictures Experts Groups)

- MPEG1 ha un bit-rate di 1.5Mb/sec
  - Viene utilizzato per la compressione di video di qualità VHS. È stato sviluppato per la memorizzazione del video su CD-ROM.
  - Nota: Il bit-rate è il numero di bit al secondo che devono essere elaborati per riprodurre il filmato
- MPEG2 è stato sviluppato per comprimere video di qualità televisiva
  - Bit rate: 2Mbps o superiori
  - Usato per DVD, cable TV, etc.
- MPEG4 è un formato basato sulla compressione degli oggetti presenti nel filmato ed è un formato che permette di gestire più stream (flussi) video
  - Supporta bit-rate variabili, da <64 kbps fino a 4Mbps e più (nel futuro)

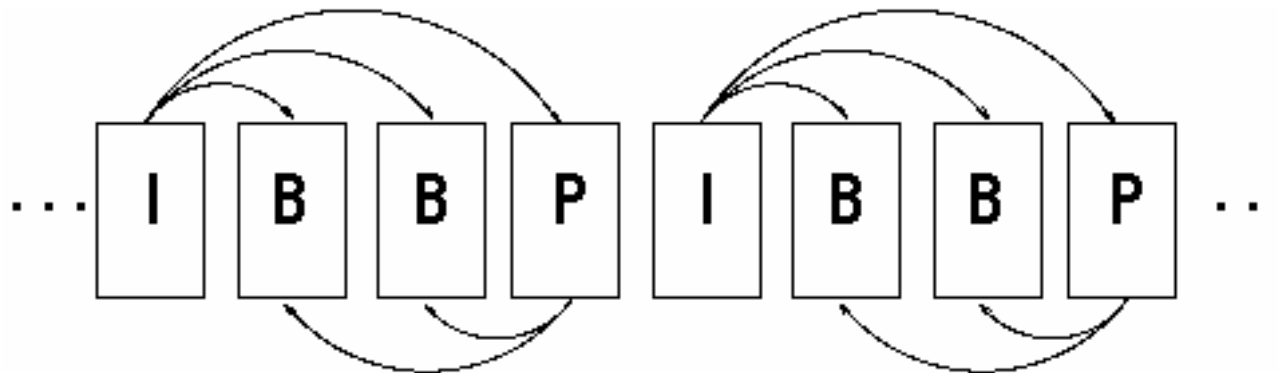


# MPEG-1 [1/2]

- ◆ **Compression based on intra-frame and inter-frame encoding**
- ◆ **Intra-frame coding**
  - Each frame is subject to compression
  - Uses DCT compression schema
- ◆ **Inter-frame coding**
  - Exploits temporal redundancy
  - Predictive coding
    - **current picture is modeled as a transformation of picture at some previous time**
  - Interpolative coding
    - **Uses past and future pictures for reference**

# MPEG-1 [2/2]

- ◆ **MPEG uses three types of frame coding**
  - I frames: intra-frame coding
    - **Moderate compression**
    - **Access points for random access**
  - P frames: predictive-coded frames
    - **Coded with reference to I or P frames**
  - B frames: bi-directionally predictive coded
    - **Coded using previous/next I and P frames**
    - **High compression**



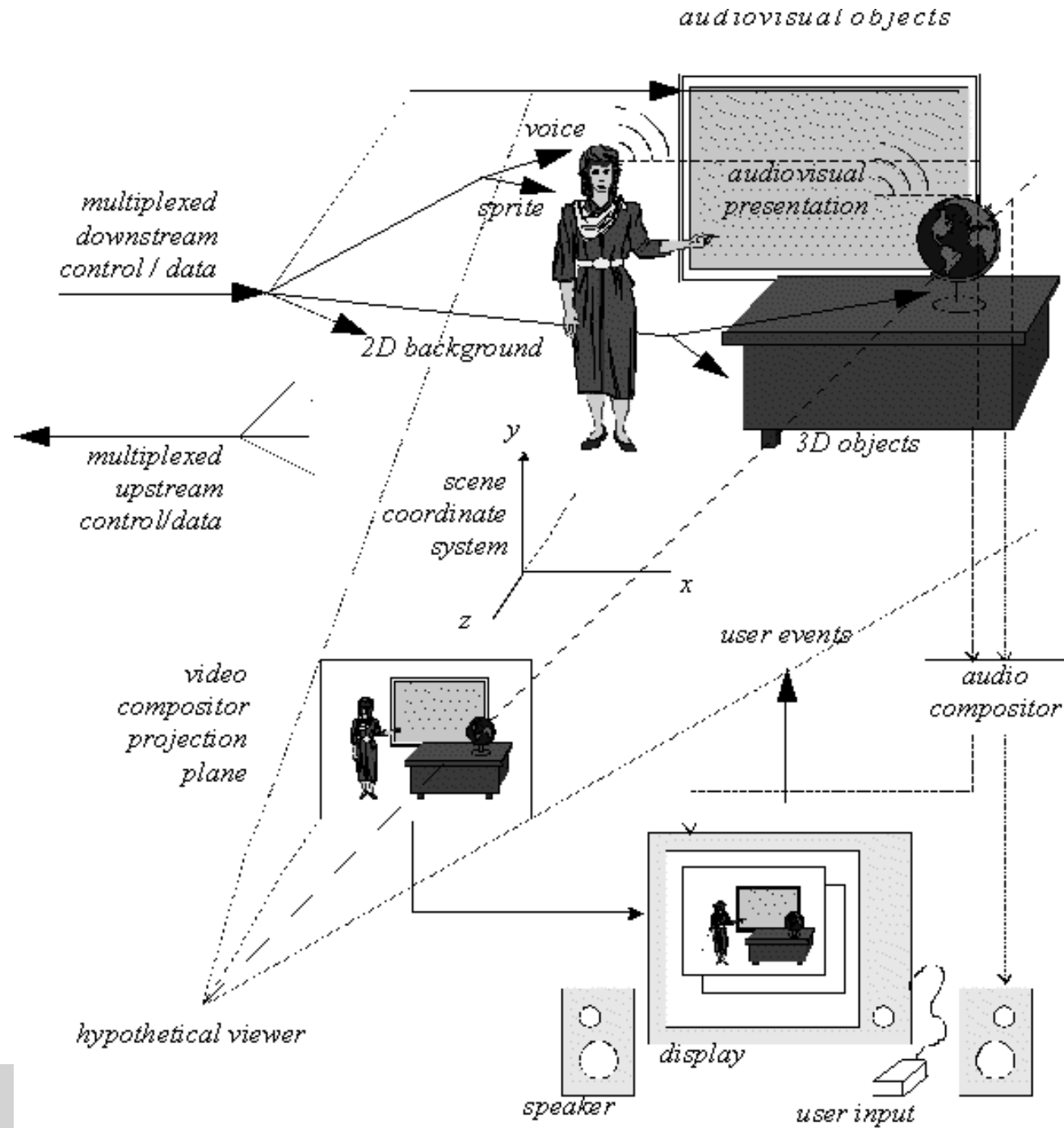
# MPEG-4 [1/4]

- ◆ **Scalability of bit rate vs quality**
- ◆ **Better Audio/Video compression than MPEG-1**
- ◆ **Content based coding**
- ◆ **Support for efficient streaming**

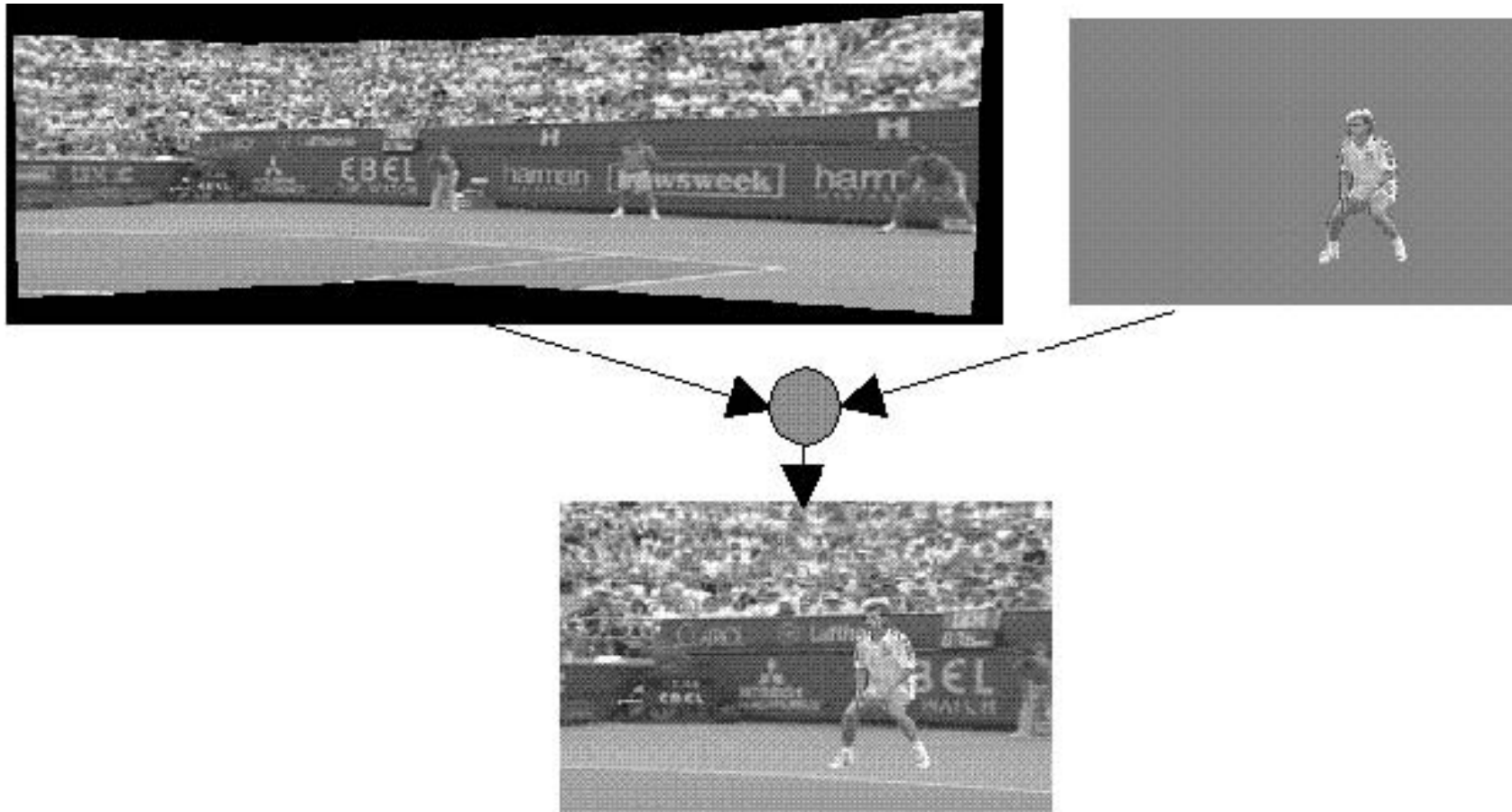
# MPEG-4 [2/4]

- ◆ **Content based coding**
  - Reusability of object coding
  - Adaptation (different coding for different objects)
  - High quality for interesting parts
  - Possibility of scene composition
    - **Integration of natural and synthetic content**
    - **Tele-presence**

# MPEG-4 [3/4]



# MPEG-4 [4/4]

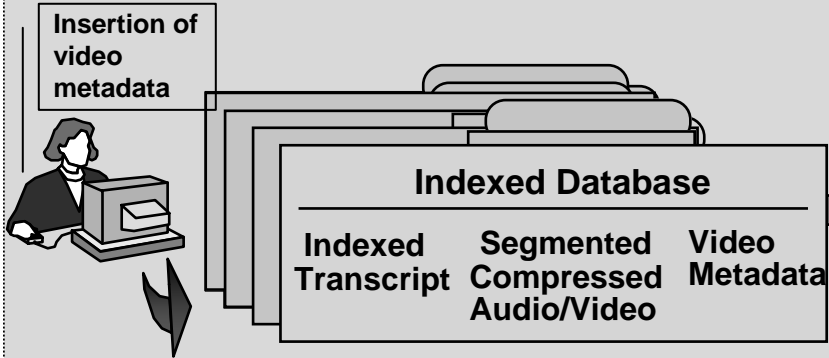
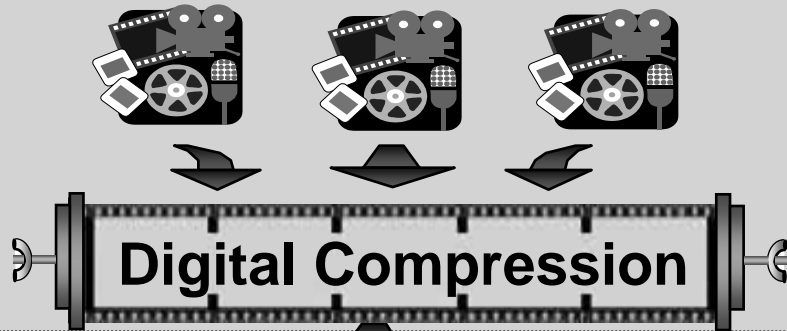


# Digital Video representation

- ◆ **Video is composed of a sequence of frames**
- ◆ **Video is separated into shots**
  - A shot is a sequence of frames separated by a transition
  - Transitions between shots are given by
    - Camera break
    - Dissolve
    - Wipe
    - Fade-in, fade-out
- ◆ **A video can be separated into sequences, that are semantically meaningful groups of shots, possibly non consecutive**

# Library Creation

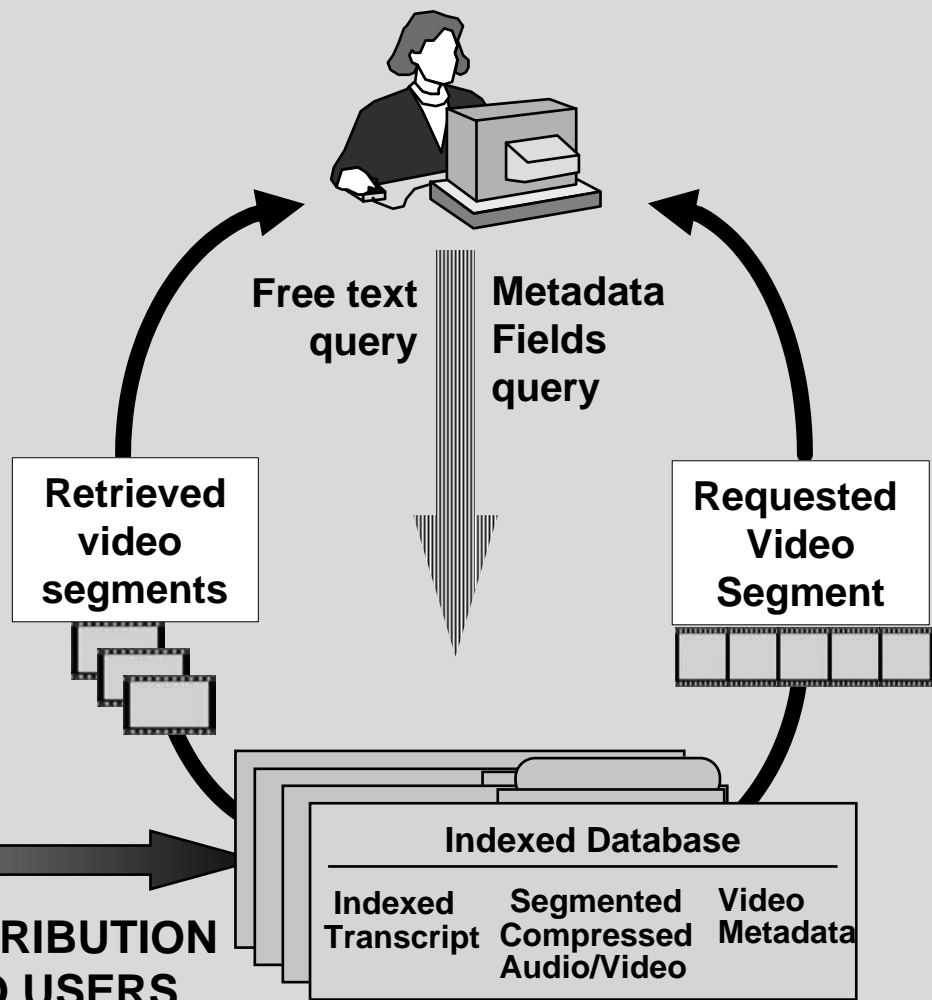
## Offline



**DISTRIBUTION TO USERS**

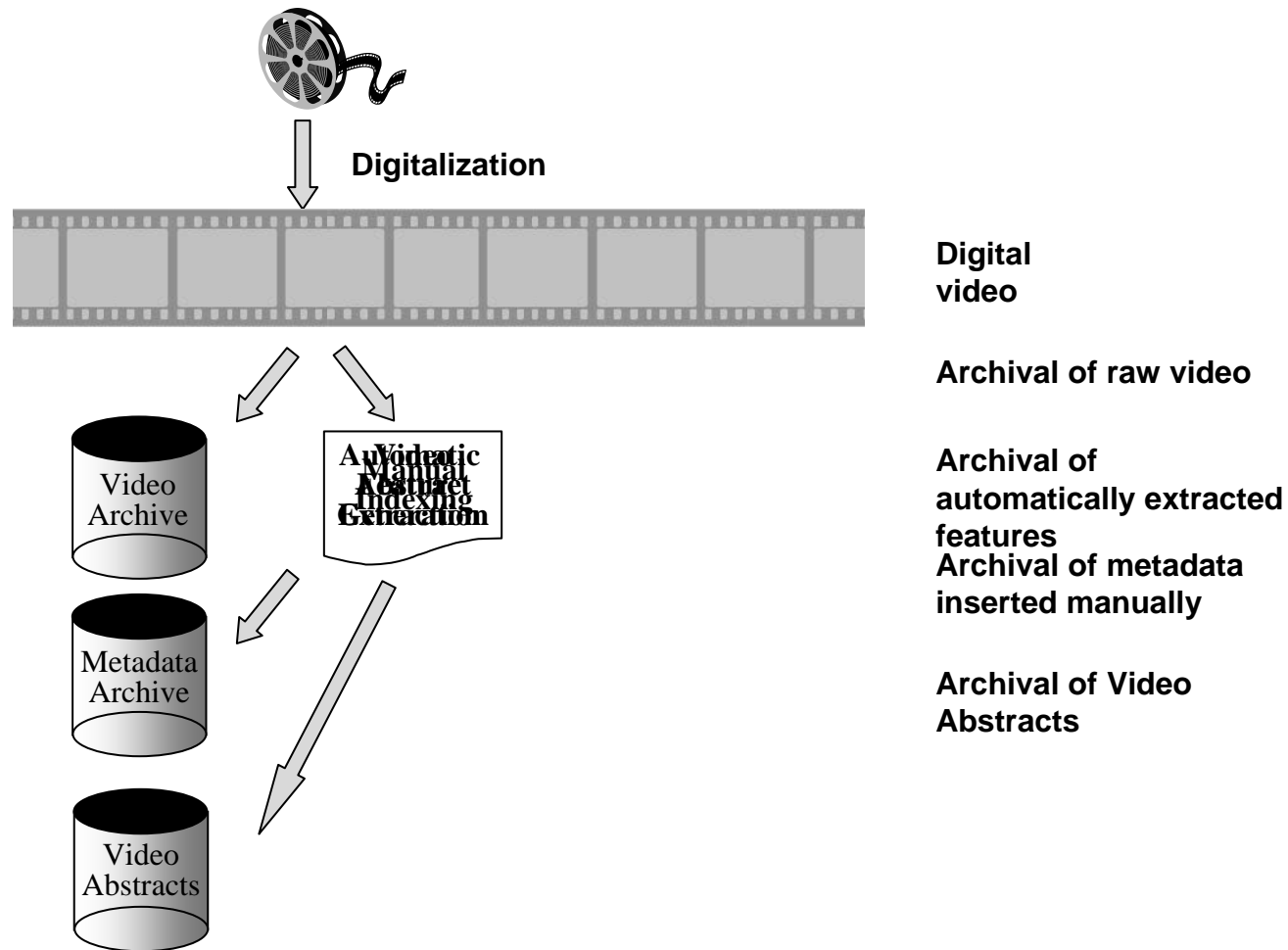
# Library Exploration

## Online





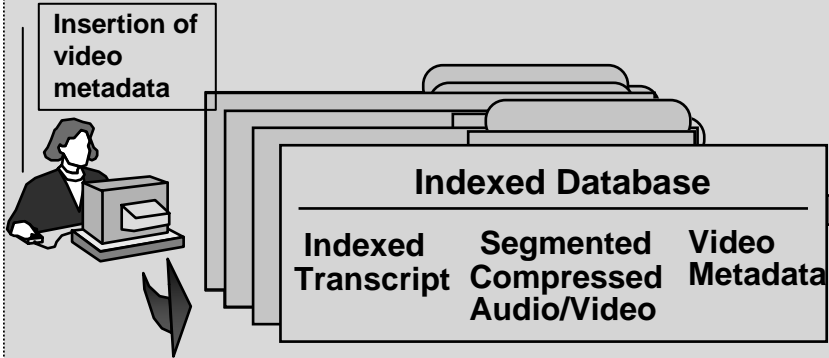
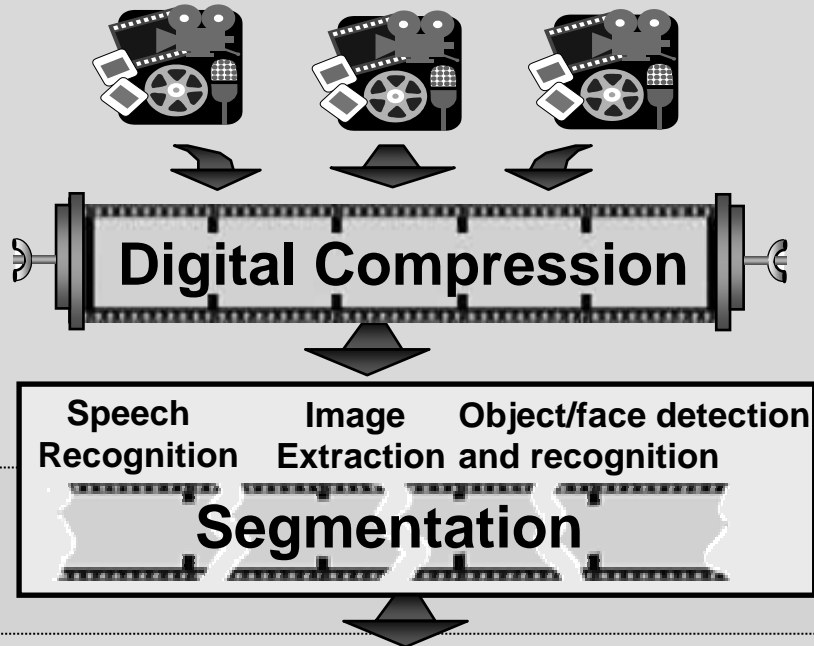
# Data flow of the Digital Library creation



# Segmentazione del video

# Library Creation

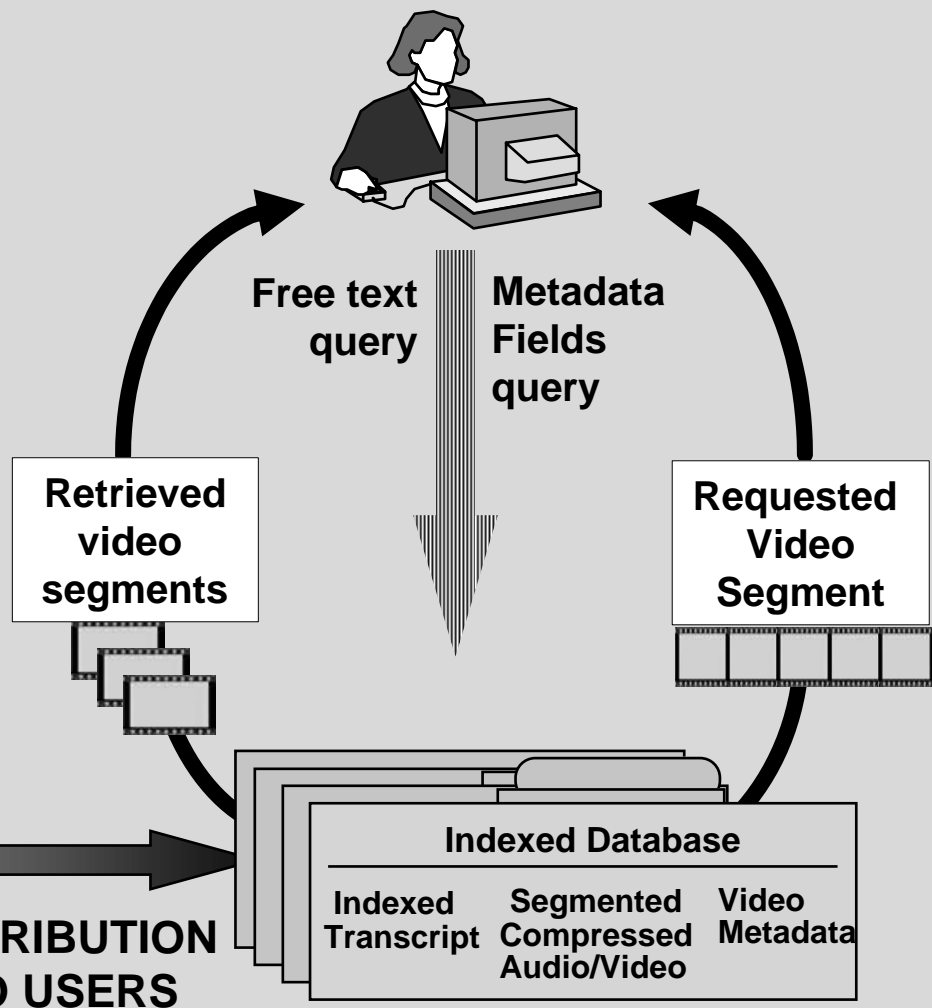
## Offline



**DISTRIBUTION TO USERS**

# Library Exploration

## Online



# Video segmentation

- ◆ **Segmentation is needed in order to identify the index units for video content. These units are *generic clips* which correspond to individual camera shots.**
- ◆ **A *generic clip*, which is the basic indexing unit, is defined as a single uninterrupted camera shot.**
- ◆ **Video partitioning consists in detecting boundaries between consecutive camera shots.**
- ◆ **The type of transitions between shots are**
  - camera break (the simplest to be detected), dissolve, wipe, fade-in, fade-out

# Segmentazione del video

- ◆ **La segmentazione del video è la decomposizione temporale del contenuto visuale in unità più piccole.**
- ◆ **I segmenti del video sono normalmente, dal più grande al più piccolo, noti come *sequenze*, *scene*, *shots* e fotogrammi.**
- ◆ **Lo shot è formato da una serie di fotogrammi che identificano una singola azione della telecamera.**
- ◆ **La scena è una sequenza di shots che appartengono ad uno stesso contesto, ad esempio la scena di un inseguimento d'auto sempre presente in un film di Hollywood.**
- ◆ **In fine, una sequenza è un segmento video composto da più scene correlate semanticamente, ad esempio un episodio all'interno di un film.**

# Segmentazione del video (cont.)

- ◆ Il processo di segmentazione si basa sul partizionamento delle sequenze video in shots che sono più facili da identificare.
- ◆ Tipicamente gli shot vengono di solito rilevati automaticamente e vengono rappresentati da uno o più keyframes.
- ◆ Gli algoritmi per l'estrazione degli shots sono utilizzati anche da programmi per l'editing video per facilitare la navigazione nel video e si basano sulla determinare la transizione temporale da uno shot all'altro.
- ◆ Esistono due tipi di transizioni, quelle improvvise (*cuts*) e quelle per graduali, come la dissolvenza.
- ◆ Un caso particolare di dissolvenza sono la *fade-in* e la *fade-out* che sono delle transizioni del video dal buio o verso il buio rispettivamente.

# Identificazione delle transizioni

- ◆ **Le transizioni vengono rilevate analizzando le differenze di due fotogrammi consecutivi.**
- ◆ **Queste differenze possono essere rilevate osservando la variazione di colore dei singoli pixel, se molti pixel dei due fotogrammi sono cambiati allora esiste una buona probabilità di aver individuato una transizione.**
- ◆ **Gli svantaggi di questa tecnica sono due: è molto sensibile ai movimenti della cinepresa, e bisogna lavorare sul video già in forma decompressa. A volte, invece di analizzare singoli pixel, si analizzano alcune statistiche di blocchi di pixel (colore medio, varianza del colore, etc.).**

## Identificazione delle transizioni (cont.)

- ◆ **Un altro tipo di approccio analizza le differenze dell'istogramma dei colori di due fotogrammi successivi. Se la differenza tra i due istogrammi supera una certa soglia predefinita, il punto di passaggio tra i due fotogrammi viene marcato come transizione.**
- ◆ **Algoritmi più sofisticati analizzano la variazione su un tempo più lungo (ossia un numero di fotogrammi maggiore di due). In modo da rilevare anche variazioni lente di una scena (dissolvenze).**



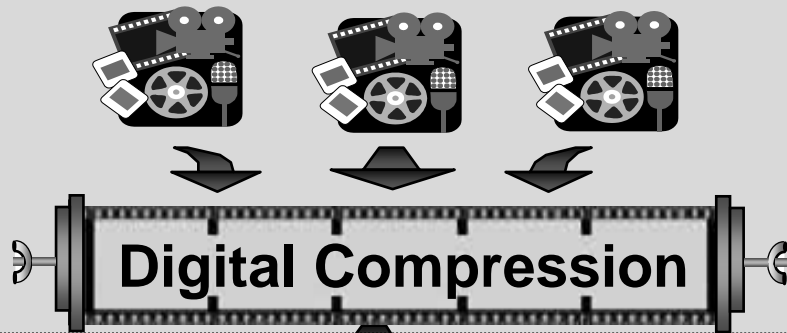
# Identificazione delle scene

- ◆ **L'identificazione delle scene, vale a dire il rilevamento di transizioni nel contenuto audiovisivo dal punto di vista semantico, invece che fisico (come quello degli shots), è molto più difficile da ottenere ed ancora oggetto di ricerca. Le soluzioni adottate richiedono un alto livello di analisi del contenuto audiovisivo, e si basano su tre tipi di strategie:**
  - l'identificazione di manifestazioni locali (dal punto di vista temporale) basata sulle regole cinematografiche di produzione, che possano far pensare ad una transizione più macroscopica. Come ad esempio, effetti di transizione, comparsa/scomparsa di musica dalla colonna sonora.
  - il raggruppamento secondo vincoli temporali: si basa sull'idea che gruppi di contenuti correlati semanticamente tendono ad essere localizzati temporalmente. Quindi solo gli shots che cadono all'interno di una predefinita soglia temporale vengono eventualmente aggregati in un'unica scena.
  - l'uso di modelli a priori sulla base del soggetto: si affidano alla conoscenza a priori del tipo di contenuto: notizie, sport, etc.

# Automatic and manual indexing of Audio/Video documents

# Library Creation

## Offline



Speech Recognition    Image Extraction    Object/face detection and recognition

## Segmentation



Insertion of video metadata



## Indexed Database

Indexed Transcript    Segmented Compressed Audio/Video    Video Metadata

**DISTRIBUTION TO USERS**

# Library Exploration

## Online

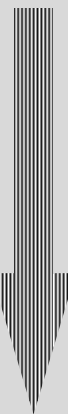


Free text query

Metadata Fields query

Retrieved video segments

Requested Video Segment



## Indexed Database

Indexed Transcript    Segmented Compressed Audio/Video    Video Metadata

# Indicizzazione

- ◆ **Integrazione tra indicizzazione automatica e manuale**
  
- ◆ **Indicizzazione automatica**
  - Individuazione keyframes
  - Analisi del movimento
  - Riconoscimento automatico di oggetti
  - Riconoscimento di scritte
  - Riconoscimento di volti
  - Riconoscimento del parlato
  
- ◆ **Indicizzazione manuale**
  - Aggiunta di attributi alle diverse parti del video (video completo, scene, singoli frame)

# What is the purpose of video indexing

- ◆ **The indexing process provides a “description” of video content that can be used to support the retrieval process**
  
- ◆ **Three main categories of video descriptions**
  - Keywords describing the entire video
  - Visual properties
  - Semantic information

# Automatic vs manual indexing

- ◆ **The goal is to provide a completely automatic indexing**
  - Fast
  - Reliable (user independent, error reduction)
  
- ◆ **In many cases this is not possible**
  - Complexity of the task (e.g. semantic interpretation of a shot content)
  - Information is not available in the video (e.g. creation date, place where the movie was recorded)

# Information that cannot be extracted automatically

## ◆ **Background information, e.g.**

- Creation date
- Author
- Names of the actors
- Ecc.

## ◆ **Semantic information**

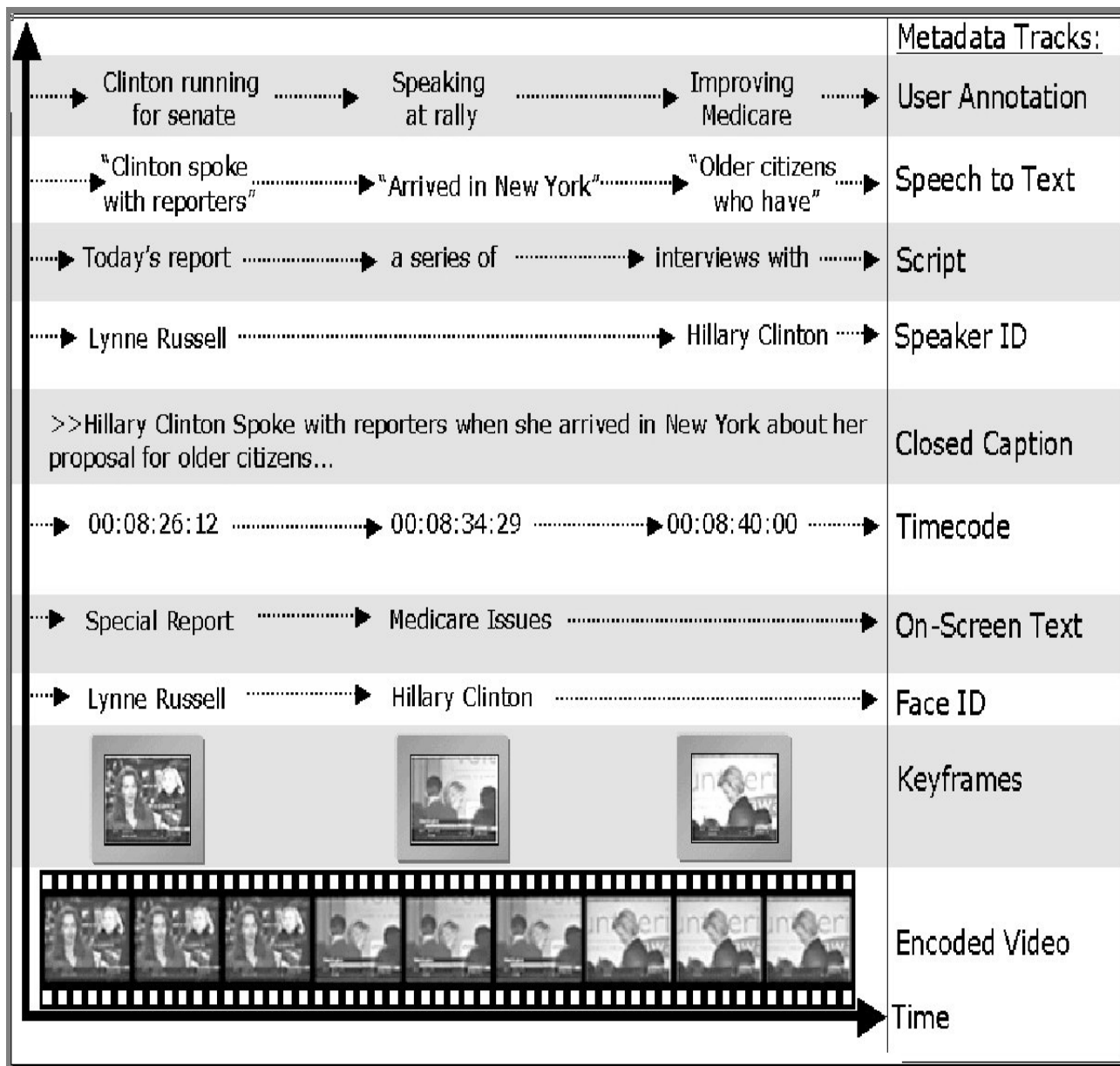
- Relations among different shots
- Interpretation of the meaning of a shot
- Interpretation of the meaning of a frame

## ◆ **All this type of information must be provided manually, possibly by using a specific tool**

# Information that can be extracted automatically

- ◆ **Features that can be extracted from the entire video,**
  - e.g. frame rate, resolution, b&w or color video, etc.
- ◆ **Features that are associated to the audio part**
  - e.g. the transcript of the speech
- ◆ **Features that can be extracted from each shot**
  - e.g. object track, camera movement, recognition of specific objects, recognition of faces, text captions, key frames
- ◆ **Features that can be extracted from each frame**
  - these are typical image features, such as color distribution, texture, object's shapes, etc.





**Video is indexed by using different tracks, which are time-synchronized**

# Il concetto di Keyframe

- ◆ In realtà realizzare un sistema che permetta di cercare su tutti i fotogrammi di un filmato è poco vantaggioso e dispendioso. Difatti un video realizzato nello standard televisivo europeo PAL è composto da 25 fotogrammi al secondo.
- ◆ Quindi un'ora di video in PAL è composto da  $25 \times 60 \times 60 = 90.000$  fotogrammi, una quantità enorme se si pensa che un sistema di archiviazione può contenere migliaia di ore di video, inoltre di molti di questi fotogrammi sono simili tra loro.
- ◆ Per ovviare a questo problema si utilizzano solo un sotto insieme dei fotogrammi del video, i keyframes (fotogrammi chiave). Un keyframe è un fotogramma particolare che rappresenta adeguatamente una scena di un video.
- ◆ I keyframes vengono generati durante la fase di identificazione delle scene discussa nel seguente paragrafo.

# Il concetto di Keyframe (cont)



# Image indexing

- ◆ **Image indexing is performed on key frames**
- ◆ **Image indexing is difficult, since the concept of image similarity is not precise**
- ◆ **Two different indexing approaches**
  - Based on image visual features
    - **Color, texture, object's shape, etc**
  - Based on a semantic information

# Analisi del movimento

- ◆ **Considerare il video come una semplice sequenza di immagini può essere riduttivo, infatti i fotogrammi di un video sono tra loro correlati temporalmente.**
- ◆ **L'estrazione di features del moto forniscono un modo efficace per effettuare ricerche attraverso la dimensione temporale. Queste features permettono di descrivere in modo sintetico informazioni sul movimento di oggetti nel video o della cinepresa.**
- ◆ **Un esempio tipico è la feature che descrive la quantità di moto un presente in una certa scena. Dato che la quantità di moto è semplicemente uno scalare (vale a dire un numero) è possibile ridurre il numero di fotogrammi su cui effettuare la ricerca usando la similarità per immagini.**
- ◆ **Ad esempio volendo cercare delle scene di calcio da programmi televisivi può essere utile ridurre l'insieme di fotogrammi da cercare a quelli che posseggono una quantità di moto superiore ad una certa soglia. Dopodichè è possibile selezionare le immagini che contengono ad esempio un prato verde.**

# Riconoscimento di scritte all'interno di immagini

- ◆ **Il riconoscimento di scritte all'interno di un'immagine non è un'operazione complessa. In una prima fase si identificano le regioni che contengono testo, isolando parti dell'immagine che hanno elementi peculiari dei caratteri tipografici.**
- ◆ **In seguito i segmenti individuati vengono elaborati in modo da accentuare i caratteri rispetto al resto, aumentando ad esempio il contrasto.**
- ◆ **In fine la parte estratta viene analizzata da un programma OCR (Optical character recognition) che estrae il testo.**

## Riconoscimento di scritte all'interno di immagini (cont.)



OCR

IN DIRETTA

L'ADDIO DELL'ITALIA AI  
CADUTI DI NASSIRIYA

SKY TG24

11.36

**Identificazione**

# Identificazione e riconoscimento di volti

- ◆ **Per ottenere al riconoscimento di un volto è necessario prima indentificarlo.**
- ◆ **Esistono decine di metodi di identificazione di volti, la maggior parte si basano sui colori o sulla ricerca di schemi predefiniti come ad esempio occhi-naso-bocca. I più promettenti utilizzano tecniche che si basano sull'approccio delle reti neurali.**
- ◆ **I problemi da affrontare nell'identificazione di un volto sono svariati. Un volto può essere illuminato in modo insufficiente, trovarsi di profilo, ci possono essere altri volti nella stessa inquadratura o esserci altri elementi di disturbo come animali o cose. Sono quindi frequenti falsi positivi e falsi negativi.**

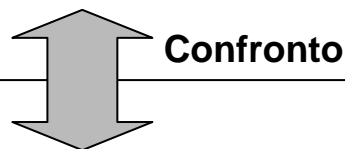
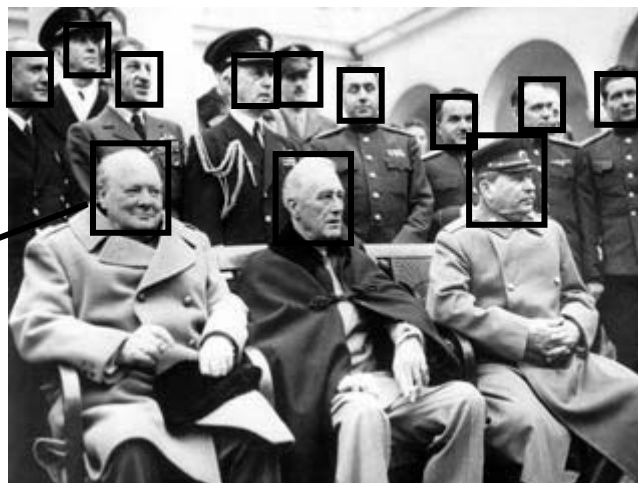


# Identificazione e riconoscimento di volti

- ◆ **Un volto, una volta identificato (tipicamente viene racchiuso in un rettangolo), può essere inviato ad una procedura per il suo riconoscimento.**
- ◆ **Per questa fase è necessario avere uno o più modelli della persona da riconoscere, non è pensabile realizzare un sistema che riesca a riconoscere chiunque. I modelli possono essere utilizzati per creare un base di conoscenza con la quale è possibile annotare un video con i personaggi riconosciuti.**

# Identificazione e riconoscimento di volti

winston  
churchill



**Basi di dati di winston churchill**

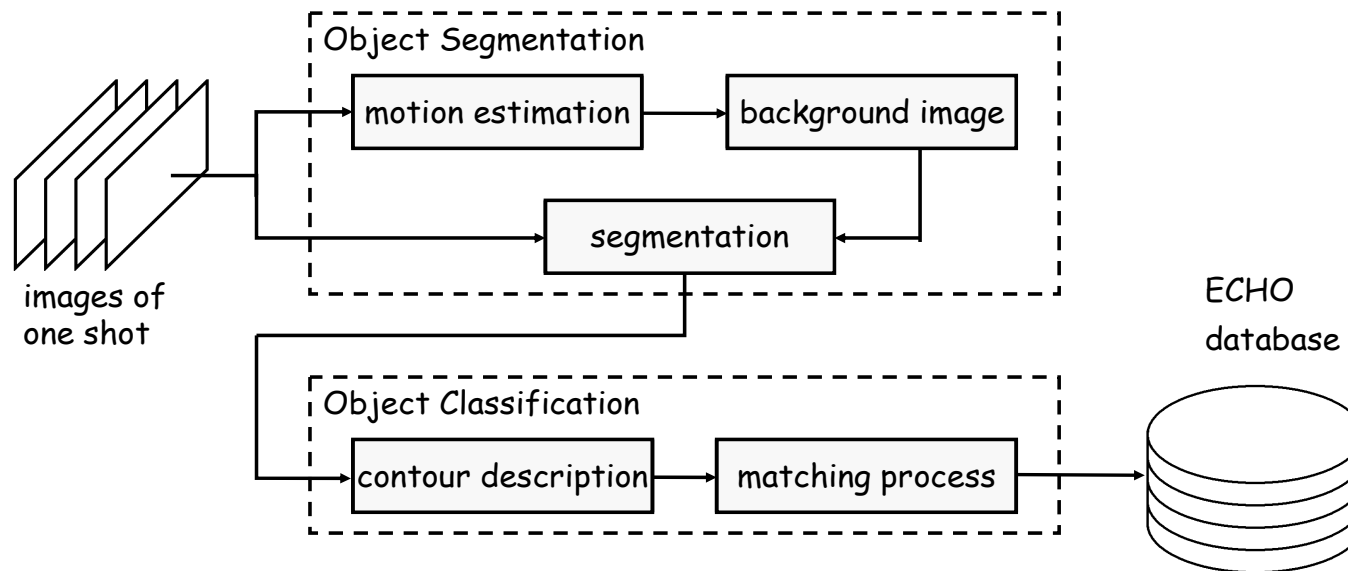
# Oggetti

- ◆ **Il riconoscimento del contenuto di un video è senz'altro la sfida più importante che vede impegnati molti ricercatori sia del mondo accademico che industriale.**
- ◆ **Il miglioramento di queste tecniche di riconoscimento potrebbe un giorno colmare quello che è chiamato in inglese semantic gap, ossia il divario semantico, che rappresenta oggi il più grosso ostacolo nella ricerca su dai multimediali. Il semantic gap è in pratica la differenza tra quello che l'utente percepisce e quello il sistema automatico riconosce.**

## Oggetti (cont)

- ◆ **Oggi si riescono a riconoscere automaticamente abbastanza bene scritte all'interno di un'immagine (e quindi in un video), e si riesce ad identificare un certo insieme di oggetti bene definiti come, automobili, animali, volti, etc.**
- ◆ **Attenzione però a non confondere il concetto di identificazione con riconoscimento. Il primo implica semplicemente l'individuazione all'interno di un'immagine di un'area in cui probabilmente è contenuto un oggetto noto (ad esempio un volto); il secondo, oltre l'identificazione del volto implica anche, appunto, il riconoscimento della persona.**

# Phases of Object detection and recognition



# Identificazione di oggetti

- ◆ **L'identificazione di oggetti è una problematica ancora più ampia della precedente.**
- ◆ **In generale nel caso di oggetti si parla semplicemente riconoscimento piuttosto che identificazione in quanto, quasi sempre, il secondo implica necessariamente il primo. Se devo identificare un cavallo necessariamente lo avrò anche riconosciuto come tale. Questa distinzione è però necessaria per oggetti particolari come le scritte e i volti.**
- ◆ **Quando si cerca di individuare diversi oggetti in un'immagine generalmente si procede ripartendo l'immagine in segmenti significativi che sono poi confrontanti con una base di conoscenza costituita da modelli vari, ad esempio aeroplani, automobili, biciclette, etc.**

# Esempio: Automobili



# Esempio: persone





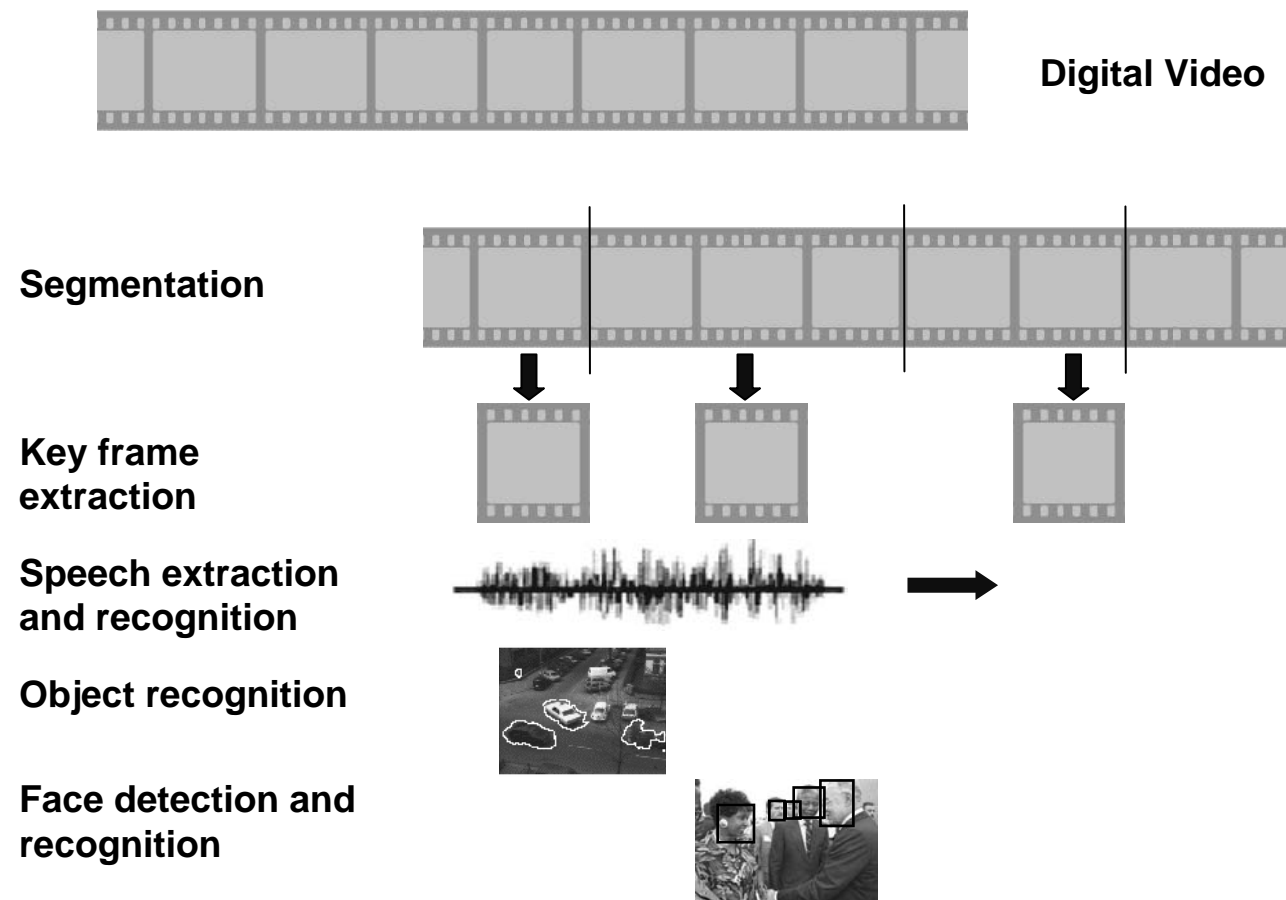
# Estrazione del parlato

- ◆ **La ricerca di parole o frasi all'interno della trascrizione del parlato può essere sorprendentemente efficace nel recupero di informazioni in documenti audiovisivi. L'utilità del testo estratto dipenderà naturalmente dal tipo di documento audiovisivo.**
- ◆ **Ad esempio il parlato di documentario o di un telegiornale è fortemente correlato con quello che è mostrato nel video. Viceversa, in un film il parlato caratterizza poco una scena dal punto di vista semantico, in quanto i personaggi in un film normalmente (come d'altra parte succede nella realtà) non commentano le scene in cui sono coinvolti.**
- ◆ **Inoltre, mentre in un documentario c'è un solo narratore che parla e che scandisce bene le parole, in un film non succede altrettanto e spesso più persone sono coinvolte nella conversazione parlano insieme.**

# Estrazione del parlato (cont.)

- ◆ **L'estrazione del parlato è un processo particolarmente complesso, che si basa su principi simili a quelli utilizzati dai programmi di dettatura.**
- ◆ **A differenza di questi ultimi però l'estrazione del parlato deve fronteggiare diversi problemi come l'ambiente (sovrapposizione di rumori, suoni, musica o riverberi), la qualità di registrazione, le caratteristiche parlatore (sesso, età, proprietà di pronuncia, stato emotivo).**
- ◆ **Ovviamente il risultato di questi algoritmi è affetto da errori, però è stato dimostrato che anche con una percentuale di errori del 50%, l'efficacia della ricerca sul testo è soddisfacente. Il riconoscimento del parlato utilizza i seguenti tre fasi di riconoscimento:**
  - *Riconoscimento acustico.* Utilizza un modello che descrive il suono dei singoli fonemi del parlato.
  - *Riconoscimento delle parole.* Utilizza un modello del lessico che descrive quali sequenze di fonemi rappresentano parole valide.
  - *Riconoscimento di frasi.* Utilizza un modello linguistico il quale determina la probabilità che una specifica parola sia stata pronunciata, sulla base della lingua del parlatore.

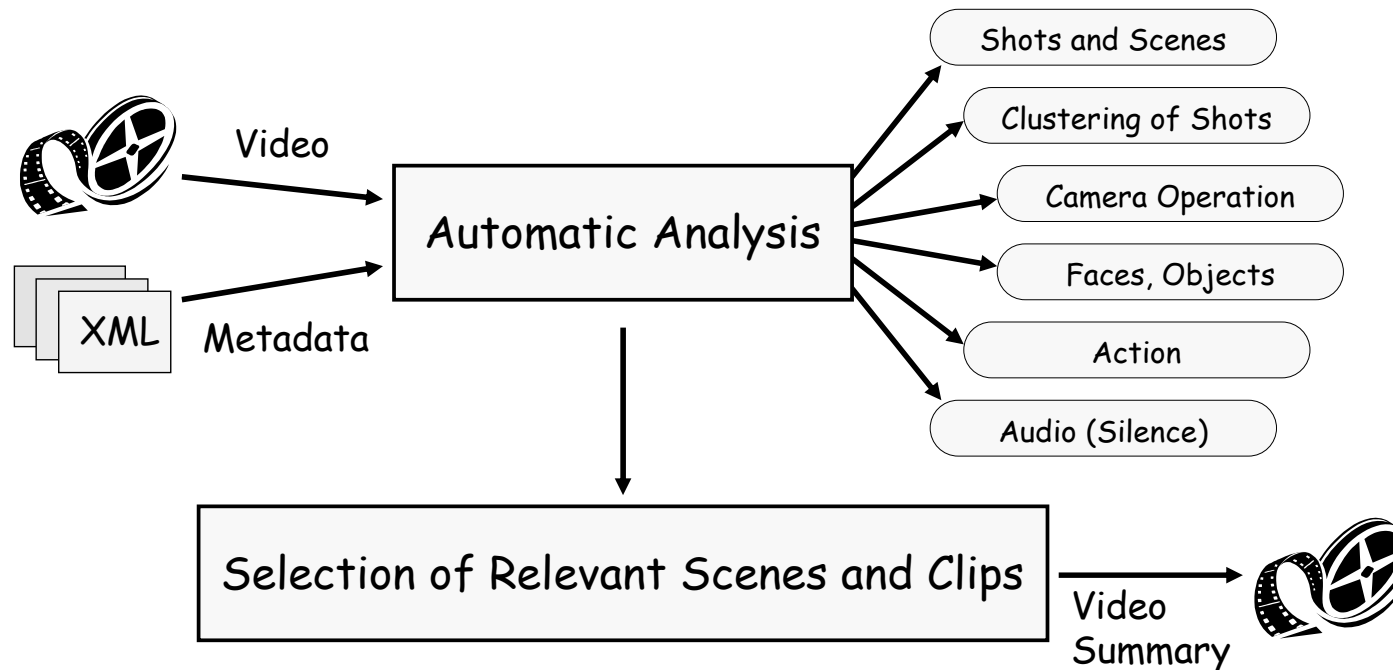
# Automatic feature extraction



# Video abstract generation [1/2]

- ◆ **A video abstract is a part of a much longer video, which preserves the essential message of the original video.**
- ◆ **A video abstract does not change the presentation medium.**
- ◆ **The users get a quick overview of a much longer video.**
- ◆ **The video abstracting application will:**
  - Select relevant clips
  - Order these clips
  - Define a transition between two clips
  - Modify the audio track

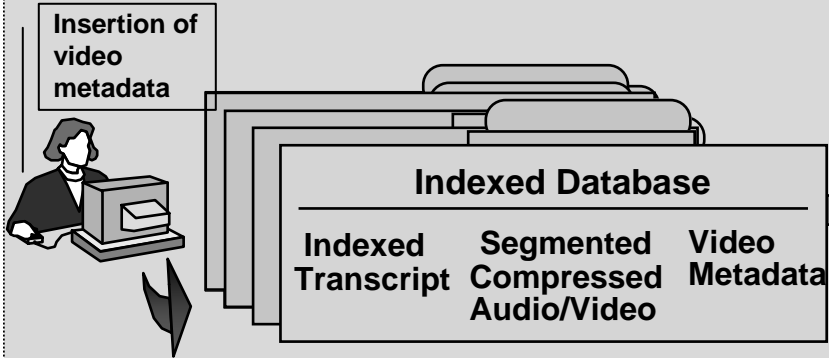
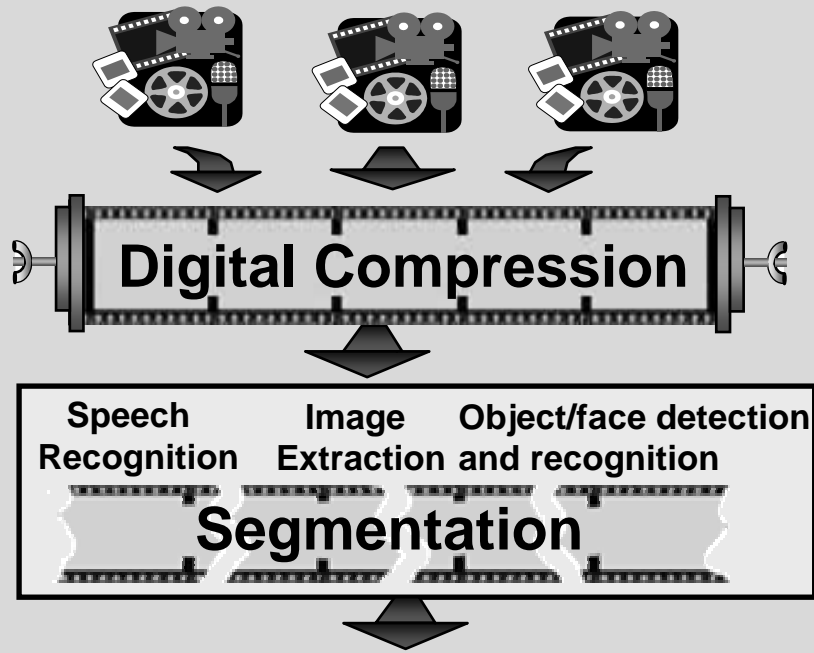
# Video abstract generation [2/2]



# Retrieval of Audio/Video documents

# Library Creation

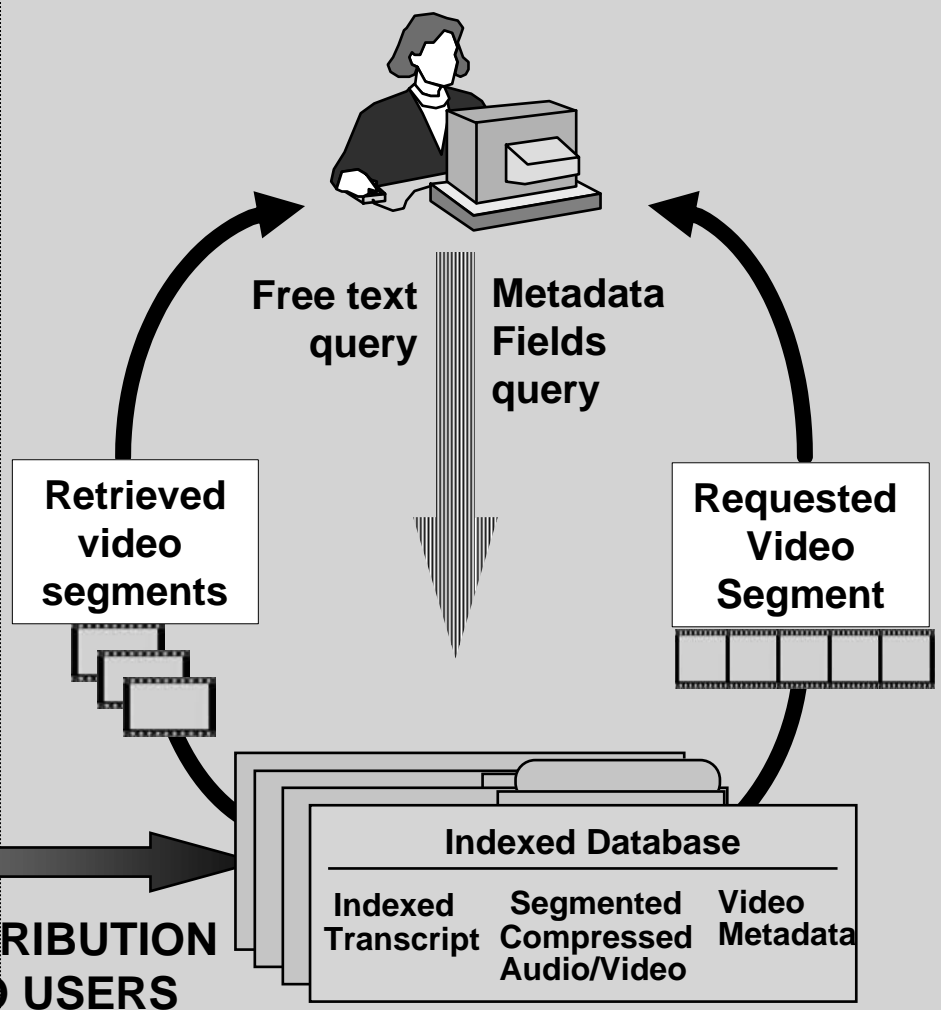
## Offline



**DISTRIBUTION TO USERS**

# Library Exploration

## Online



# Retrieval functionality

- ◆ **Retrieval is based on queries expressed on metadata values.**
- ◆ **Both automatically extracted metadata, as well as metadata associated manually to the video can be used.**
- ◆ **The user may not distinguish between these metadata types; system behavior may be different**



# Type of queries

## ◆ Queries can be expressed on

- Metadata associated to the entire video
  - **E.g.** *find b&w videos produced before II world war by Istituto Luce*
- Metadata associated to video shots
  - **E.g.** *find a shot where the audio transcript contains the words “Attentato Banca Nazionale dell’Agricoltura”*
- Metadata associated to single frames
  - **E.g.** *find a video that contains a frame similar to this image [the image is provided as an example]*
- Any combination of the previous cases

# Retrieval characteristics

- ◆ **Retrieval is based on an approximate match between the query and the retrieved videos. This is mainly the case when imprecise query elements are used (e.g. free text, images)**
- ◆ **Retrieved videos are returned to the user in decreasing relevance order, possibly indicating the degree of relevance of the retrieved items.**
- ◆ **Due to the imprecision of the method (i.e. some of the retrieved items are not relevant for the user and some relevant items are not retrieved), it is helpful to have a query refinement and a relevance feedback mechanism.**

# Informedia – an example

The screenshot displays the CMU Informedia Video Library interface. The main window is titled "CMU Informedia Video Library" and contains a search bar with the query "fires floods earthquakes hurricanes tornadoes". Below the search bar, there are buttons for "Clear All" and "History". The search options are set to "ANY of". The results show "9 of 287 results: any of 'fires floods earthquakes hurricanes tornadoes.'" and a prompt to "Click on a word to highlight it in yellow." Below this, there are navigation buttons: "Prev. Page", "Next Page", "Go to Page...", and "Visualize All...".

The "Search Results (Page 1 of 32)" section shows a grid of video thumbnails. The right-hand side of the interface features a "Visualization of search results set containing 287 documents" window. This window has tabs for "VIBE", "Timeline", "Map", and "Topics". The "VIBE" tab is active, showing a scatter plot of points representing documents. Labels for "fires", "floods", "torn...", "hurr...", and "eart..." are placed near clusters of points. Below the plot, there is a list of checked terms: "fires", "floods", "earthquakes", "hurricanes", and "tornadoes".

At the bottom right of the visualization window, there are settings for "Color Code By" (Minimum Value, Average Value, Maximum Value) and "Display lines toggled. 287 visible documents." Below this, there are checkboxes for "Relevance (All: 0 - 100)", "Size-Code", and "Color-Code". A "Within" field with a right arrow is also present. At the bottom, there are checkboxes for "Date (All: 07/01/99 - 09/30/99)", "Size (All: 0.02 - 30.30)", "Map Hits (All: 1 - 125)", and "Topic Hits (All: 1 - 253)", each with "Size-Code" and "Color-Code" options.

# Informedia – an example

The screenshot displays the CMU Informedia Video Library interface. The main window is titled "CMU Informedia Video Library" and contains a search bar with the query "fires floods earthquakes hurricanes tornadoes". Below the search bar, there are buttons for "Clear All" and "History". The search results section shows "9 of 287 results: any of 'fires floods earthquakes hurr tornadoes.'" and a list of video thumbnails. A "Visualization of search results set containing 287 documents" window is open, showing a world map with a tooltip for "TURKEY: 46/48 hits active". The map visualization includes tabs for "VIBE", "Timeline", "Map", and "Topics", and a legend for "Countries" and "States". The bottom right panel contains filters for "Visible", "Invisible", and "Inactive" documents, with a list of visible locations: "CHINA", "MEXICO", "TAINWAN", and "TURKEY". The filter panel also includes options for "Color Code By" (Minimum Value, Average Value, Maximum Value) and various search criteria like "Relevance", "Date", "Size", "Map Hits", and "Topic Hits".

# ECHO Retrieval Interface

**ECHO Retrieval Web Service - Microsoft Internet Explorer**

File Bearbeiten Ansicht Favoriten Extras ?

Browsing Player Cataloging Metadata Edit Transcript Export Delete

Label: In Titles \* Value: Submit

In Content

In Dates 1933

Collocation-ID

in archive: IL show 500 hits

sorted by

01.06.1933

00:02:04.23 IPR 01.10.1933

00:00:50.10 IPR 01.01.1933

00:01:00.07 IPR 08.09.1954

Title Nella sede provvisoria del Reichstag il cancelliere Hitler espone il suo programma di governo.

Series Title Giornale Luce

Series Number B0242 Genre Newsreel

English Abstract In the Reichstag, Hitler explain his political program.

Themes 2.2 - Le Guerre Mondiali - 1920-1945 Eventi Principali

Description Language IT

Producer Name FOX Movietone Production Date 01.01.1933

Producer Nationality USA

Kind Whole

Silent Sound Color BW

Audio Language DE

Collocation B024203

Provider IL Storage ID B024203

Fertig Lokales Intranet

Integration of Clients and Services in GUI

Views on the material

Attribute Search Retrieval interface

Search for category level of material (Work, Expression, Manifestation, Item)

Detailed view on an item corresponding to the full ECHO data model (structures, links, ...)

List of retrieved items

**ECHO Prototype 3 Web Service - Microsoft Internet Explorer**

File Bearbeiten Ansicht Favoriten Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links

Adresse http://cms289:8080/MediaArchive/Servlet/startView Wechseln zu

Browsing Metadata Editor Export Import Delete

ES | CO Video Segmentation

Kind of Search:  
 Nearest Neighbours  Range  
 Strategy: distance  
 Tune Parameter: 0.01  
 Hits: 10

Submit  
 hits: 10

DMGuid: 444C4755494400DF3AB8C2F7B10F009518EE810C  
 Timecode: 00384960  
 Distance Value: 0.492198

DMGuid: 444C4755494400B4499179D81E300054F49CE50C  
 Timecode: 00067280  
 Distance Value: 0.63827

DMGuid: 444C4755494400897DA0F56A9052001782671B03  
 Timecode: 00061280  
 Distance Value: 0.638392

DMGuid: 444C4755494400897DA0F56A9052001782671B03  
 Timecode: 00004240  
 Distance Value: 0.676915

DMGuid: 444C47554944000779CB40F7B10F002E6F2A5403  
 Timecode: 00051960  
 Distance Value: 0.680515

Work Title: Nozze principesche a Monaco  
 Duration: 00:08:20.19  
 Date of Ingest: 21.08.2002-14:38:46  
 Material: video  
 ECHO Model Entity

```

graph TD
    AV[AV-Document] --> V[Version]
    V --> M[Media]
    M --> S[Storage]
    M --> Video[Video]
    M --> Audio[Audio]
    M --> Transcript[Transcript]
    Video --> Video
    Audio --> Audio
    Transcript --> Transcript
  
```

Title: Nozze principesche a Monaco  
 Has Version: [\(To Version object\)](#)  
 Has Audio: [\(To Audio object\)](#)  
 Has Transcript: [\(To Transcript object\)](#)  
 Unit Indication: Whole  
 Silent: Sound  
 Color: BW  
 Person: Kelly, Grace; Ranieri III di Monaco  
 Locations: Principato di Monaco; New York; Statua della libertÃ

Fertig Lokales Intranet

ECHO Prototype 2 Web Service - Microsoft Internet Explorer

File Edit View Favorites Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links

Adresse http://cms289.8080/MediaArchive/servlet/startView Wechseln zu

Browsing Metadata Editor Export Import Delete ECHO European Content Online

Find \* Submit

as of unlimited show 10 hits of type

in archive \* sorted by Relevance in ascending order Hits: 10

Zuckerfabrik Frauenfeld

Zirkus Pilatus

Zirkus Knie: Training im Winterquartier

Zibelemärit

Zentrum des schönen Buches in Ascona

Zeltlager im Schnee

Zeitmessung und Uhrenindustrie

ZEVENDE NEDERLANDSE KATHOLIEKENDAG

Yoga

00:06:59.19...00:06:59.21 (00:00:00.02)

00:06:59.22...00:07:02.09 (00:00:02.12)

00:07:02.10...00:07:04.05 (00:00:01.20)

00:07:04.07...00:07:06.06 (00:00:01.24)

00:07:06.07...00:07:07.13 (00:00:01.06)

00:07:07.15...00:07:08.21 (00:00:01.06)

00:07:08.24...00:07:12.10 (00:00:03.11)

00:07:12.13...00:07:15.02 (00:00:02.14)

Fertig Lokales Intranet

**ECHO Prototype 3 Web Service - Microsoft Internet Explorer**

File Edit View Favorites Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links

Adresse <http://cms289.8080/MediaArchive/servlet/startView> Wechseln zu

Browsing Metadata Editor Export Import Delete

ES CO Video Segmentation

Find:  Submit

Archive: IL INA Memoriv NAA Themes: 1 - Post War 2 - The World Wars 3 - Sports in the 20th Century 4 - Daily Life 5 - Youth Culture in Europe

show 100 hits Hits:

- [mec001](#)
- [mec001](#)
- [RAMSIS.I](#)
- [Vw Käfer](#)
- [Zuckerfabrik Frauenfeld](#)
- [Zirkus Pilatus](#)
- [Zirkus Knie: Training im Winterquartier](#)
- [Zibelemärit](#)
- [Zentrum des schönen Buches in Ascona](#)
- [Zelllager im Schnee](#)

Fertig Lokales Intranet



ECHO Prototype 3 Web Service - Microsoft Internet Explorer


Adresse <http://cms289:8080/MediaArchive/servlet/startView>


Find:  Submit


as of:  show:  hits of type:


in archive:  sorted by:  in:  order


Hits:

- 

DMGid: 444C4755494400B4499179D81E300054F49CE503  
 Timecode: 00069680  
 Distance Value: 0.811999
- 

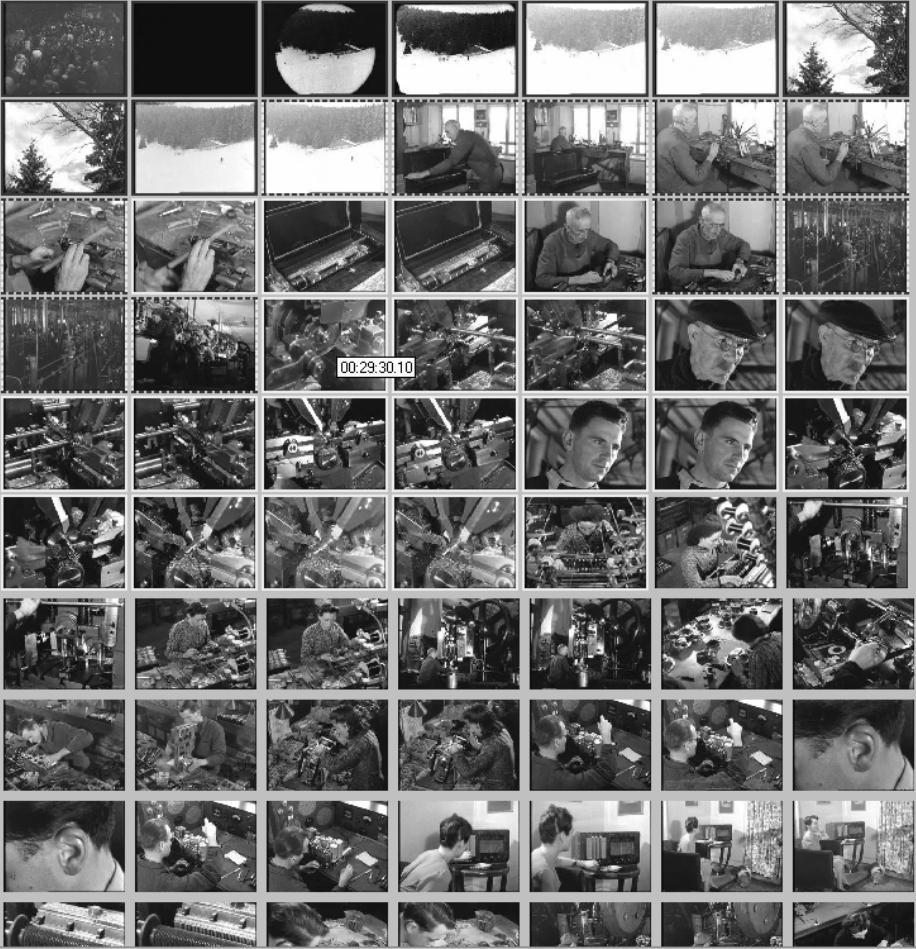
DMGid: 444C4755494400897DA0F56A9052001782671B03  
 Timecode: 00074480  
 Distance Value: 0.870983
- 

DMGid: 444C4755494400897DA0F56A9052001782671B03  
 Timecode: 00239920  
 Distance Value: 0.874066
- 

DMGid: 444C4755494400BD2E9CCC65CE0E00C65F457CC  
 Timecode: 02318000  
 Distance Value: 0.904751
- 

DMGid: 444C4755494400897DA0F56A9052001782671B03  
 Timecode: 00025800  
 Distance Value: 0.811999

Video Segmentation



Fertig Lokales Intranet