Introduzione alle Biblioteche Digitali Audio/Video

Gestione del video

- Perchè è importante poter gestire biblioteche digitali di audiovisivi
- Caratteristiche specifiche dell'audio/video
- Applicazioni delle biblioteche digitali audio/video
- Alcuni esempi di biblioteche digitali audio/video

The importance of video

Video can be considered today the primary information and communication channel, due to

- Richness in information contained
- Appeal

 Video libraries will become essential in many application fields

- Personal information
- Distance learning
- Telemedicine
- •

Video characteristics

High video production vs print production

- TV stations produce 50 Million hours of video per year (25,000 TB)
- Newspapers and periodicals produce less than 200 TB of data per year

Storage and transmission problems

Video is usually compressed

Richness in content

Difficulties in automatic extraction of content description

Services of A/V Digital Libraries

- Digital Video Libraries are more complex than traditional DLs; they require the integration of several specialized technologies
- They offer the same services of text digital libraries

Specific characteristics of Indexing and retrieval services

- Indexing based on the integration of different technologies for the automatic feature extraction
- Integration of manual and automatic indexing
- Retrieval based on different video features

Characteristics of an Audio/Video DL

The need of A/V DLs

Nowadays, video is present in many situations

- TV broadcasting
- Professional applications, such as medicine, journalism, advertising, education, training, surveillance, etc.
- Movies
- Historical videos
- Personal videos

The combination of audio and video is a very powerful communication channel

approximately 50% of what is seen and heard simultaneously is retained

Advantages of A/V DLs

Most of the video material produced is used only once, due to the difficulty to archive it, to protect and to retrieve.

A large video library of distributed and network searchable videos would enable

- Preservation of precious and expensive material
- Reduction of production costs for new videos, through the reuse of existing material
- Diffusion of knowledge

In general, it will enable the access to information that could have been lost.

A/V vs traditional DLs [1/2]

Library creation

- Traditional DLs, contain text documents
 - Library creation requires automatic acquisition of text, extraction of document content, and indexing
 - This process is well known and many different techniques have been developed
- Video is extremely rich in "content" but
 - the indexing of video content is difficult, expensive, and extremely dependent from the user and the application
 - → A possible approach consists in an appropriate integration of automatic content extraction (e.g. speech recognition, image analysis, etc.) and manual indexing

A/V vs traditional DLs [2/2]

Library exploration

- Traditional DLs, contain text documents
 - Library exploration requires simple interfaces to formulate queries on free text and document metadata.
- Video libraries should permit
 - ➔ To formulate queries on many different "dimensions"
 - Text, as extracted from speech and captions
 - Images extracted as key frames
 - Motion information
 - Other features automatically extracted
 - Metadata provided manually

Applications of Audio/Video DL

Who may use A/V DLs?



We consider four main categories

- Large companies
 - → Large corporations that may use Digital Video for their internal business, for advertising, promotion, etc.
- Media and entertainment

➔ The most traditional area. Video is one of the key assets.

- Education
 - → Video recording of courses
 - ➔ Video used as course material
- Others
 - Health and medicine
 - → Government
 - → Surveillance
 - → Etc.

Large companies

Audio/video digital libraries are used for

- Sales
- Product launches
- Marketing
- Relation with investors
- Product design (acquisition and analysis of customer's needs)
- Support for online sales
- Video archives for internal use
- Special services for customers, such as web access to specialized video archives, e.g.
 - News
 - Economic information
 - Products
 - → Materials
 - → Etc.

Media & Entertainment [1/3]

Broadcasting companies

- Many broadcasters are creating and distributing video programs on the web. A video archive is very helpful to them to add a new service for accessing old video material.
- Examples:
 - → ABC News
 - ➔ Mediaset
 - → RAI

- Archive of old programs
- Archive of daily programs
- Additional material w.r.t. tv programs

Media & Entertainment [2/3]

Video archives

- Many national and private organizations own old video material. The digitalization and archiving of this material is beneficial for content owners (for example, they can promote the use of their material) and for users belonging to different categories: e.g. professional users (that need the material to produce their video programs) or researchers or general public.
- Examples:

→ Istituto Luce

Media & Entertainment [3/3]

\diamond

Movie production companies

- Many large movie production companies own a large amount of video material, composed of the films and of related material, such as cuts not used in the final film version, interview, video trials, etc. This material is very helpful for many purposes, from the production of DVD version of the film up to the critical study of the video. Providing access to the general public of this material is also a powerful promotion and advertising channel.
- Examples:
 - → <u>MGM</u>
 - → 20th Century Fox

Education

Digital video used for different purposes

- Promotion and advertising
 - Online preview of training content
- Store and distribute the video courses
 - Remote access of the courses
 - Keep track of classroom discussion
- Used as course material
 - Delivery of video clips to students, either online or in the classroom
 - From remote sites provide students and teachers with on-demand, searchable access to whole programs and video clips
 - Free search and access to the video library con be used by students to find answers to specific questions, to study in depth some topics, etc.
- Production of new courses
 - Improve the course production procedures, allowing teachers and producers to remotely access the video library
- Examples:
 - Princeton University
 - → Harward Business School
 - University of Arizona

Other Applications [1/2]

Health and medicine

- Health and social care info to the general public
- Information to physicians for special purpose medical procedures
- Training

Other Applications [2/2]

Government

 Enhancement of the governmental decision making process, by recording and archiving of public meetings and discussion.

Surveillance

- A large amount of video is produced for surveillance purposes.
 - Required automatic video analysis
 - Archiving for successive search

The characteristics of Digital video

Types of data managed

A digital video is composed of a sequence of frames plus possibly an audio track.

- In general, it is possible to view an audio/video document from different perspectives
 - The audio part can be separated into
 - → Speech
 - → Sound
 - Sequence of frames (video shot and sequence)
 - Single frames as images

 From all of them is possible to extract information that can be used for indexing and retrieval purposes

Digital video characteristics

Sequence of frames with a certain frame rate

- NTSC 30 frames/sec, PAL 25 f/s, HDTV 60 f/s
- Minimal change between frames
- Single frames resolution
 - 768 x 576 PAL, 720 x 480 NTSC

Uncompressed video requires high storage space and bandwidth

 For example, one second of uncompressed PAL video requires 768 x 576 x 16 x25 ~ 172 MByte

Digital video storage and transmission [1/3]

- The high storage requirements of video imposes the adoption of compression techniques.
- High compression rates are possible with video signals, due to the following reasons:
 - Spatial correlation: correlation among neighboring pixels
 - Temporal correlation: correlation among pixels in different frames
 - A significant part of video data is not perceived

Digital video storage and transmission [2/3]

Compression can be divided in two broad categories

- Lossless compression, that allows one to compress decompress video without any degradation
 - Lossless compression provides low compression factors
 - An example of lossless compression is MJEPG, where each frame is compressed using the JPEG format
 - Examples of lossless coding techniques are run-length coding, Huffman coding

Digital video storage and transmission [3/3]

- Lossy compression, where the complete cycle of compression and decompression introduces some degradation of the original video
 - Lossy compression allows to obtain high compression factors
 - Examples are the MPEG compression family (MPEG1, MPEG2)
 - Example of lossy coding is DPCM
 - DPCM compares adjacent pixels and stores only their difference

MPEG

MPEG (Moving Pictures Experts Groups)

MPEG1 has a bit-rate up to 1.5Mb/sec

Designed for storage and retrieval of VHS quality video on CD-ROM

MPEG2 Designed for broadcast video quality

→ Bit rate: 2Mbps or higher

- → Used for DVD, cable TV, etc.
- MPEG4 is object-based, multi stream

Variable bit-rates, from <64 kbps, up to 4Mbps and more (in the future)</p>

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

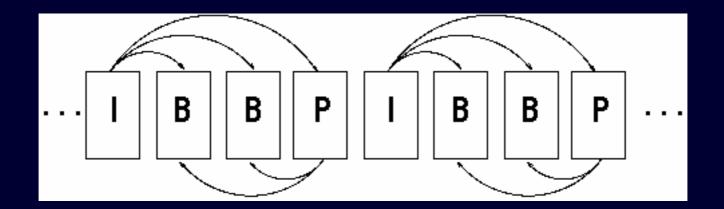
Inter-frame coding

- 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



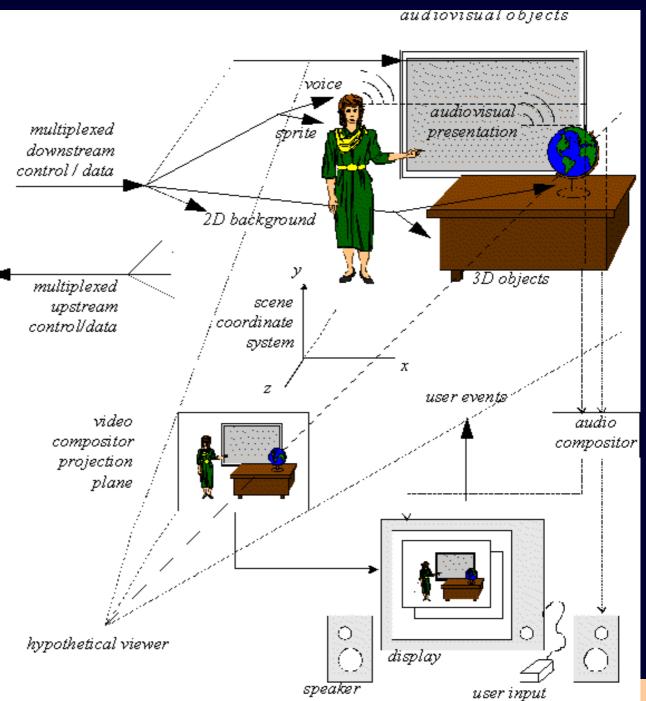
MPEG-4 [1/4]

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

MPEG-4 [2/4]

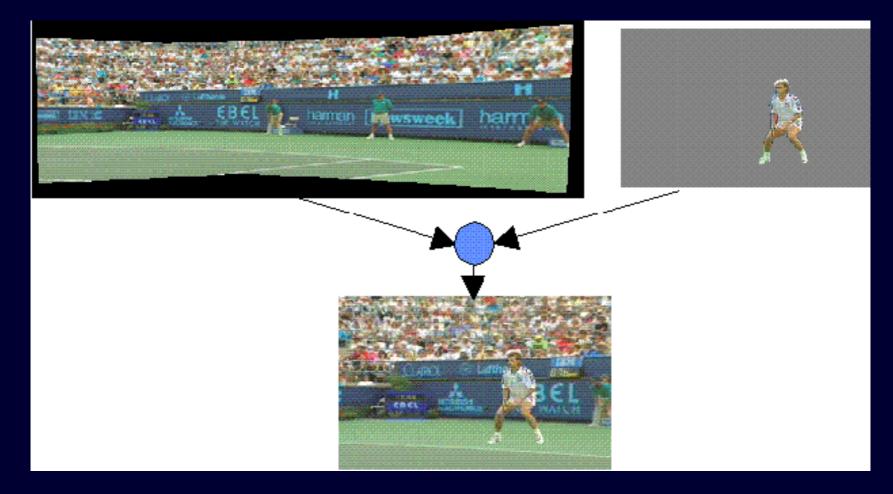
Content based coding

- 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



Pasquale Savino – ISTI-CNR





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

 \bullet

Operations of an A/V Digital Library

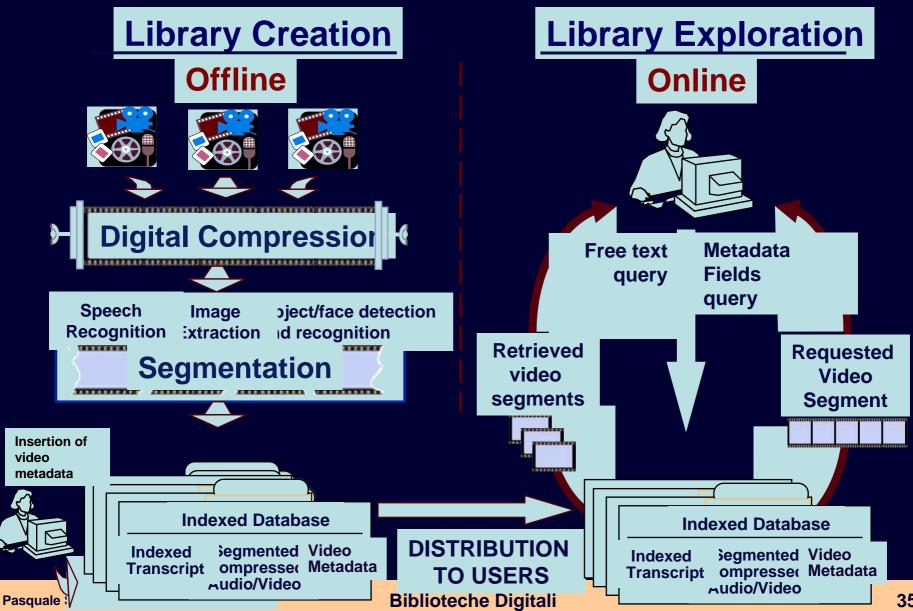


Video storage

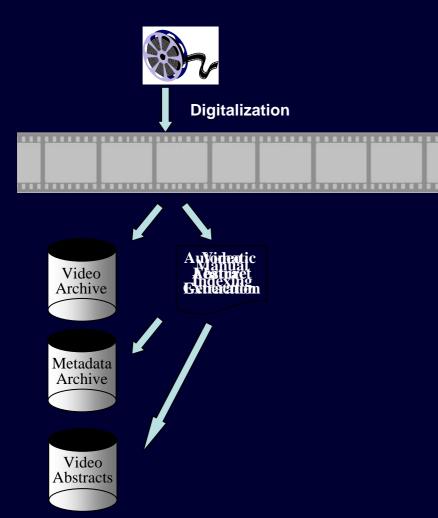
Content-based search

Video access (visualization and copy)

Summary of all phases & operations



Data flow of the Digital Library creation



Digital video

Archival of raw video

Archival of automatically extracted features Archival of metadata inserted manually

Archival of Video Abstracts



Automatic and manual indexing of Audio/Video documents



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

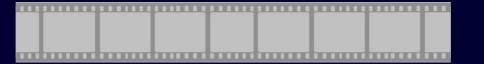
Features that can be extracted from each shot

 e.g. object track, camera movement, recognition of specific objects, recognition of faces, text captions, key frames

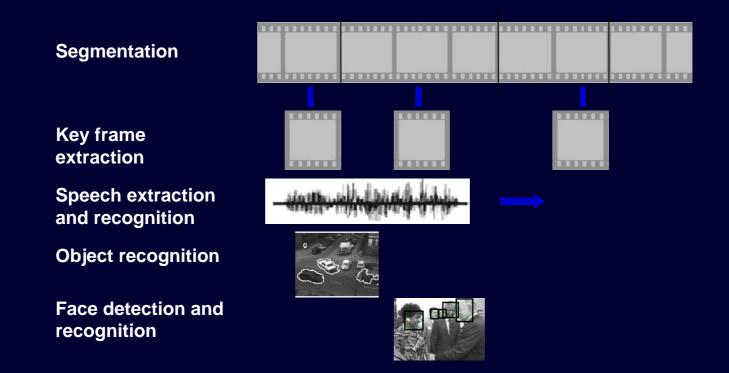
Features that can be extracted from each frame

 these are typical image features, such as color distribution, texture, object's shapes, etc.

Automatic feature extraction



Digital Video



Video segmentation [1/2]

- 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

Video segmentation [2/2]

Detection of camera break

- Pair-wise Pixel comparison Given two consecutive frames, corresponding pixels are compared and the number of pixels changed is determined. For monochromatic images, a pixel is judged as changed if the difference between its intensity values in the two frames exceeds a given threshold T.
- Histogram comparison This method uses a comparison of some feature of the images. For example it may use the histogram of intensity levels. The principle behind this approach is that two frames having an unchanging background will show little difference in their respective histograms. This method is less sensitive to object motion because it ignores the spatial changes in a frame.
- Motion continuity Motion can be represented quantitatively by assigning a field of motion vectors to the pixels of an image. The motion vectors are computed by dividing each frame into blocks and determining where each block is located in the successive frame. A correlation between the two frames can be computed; a low correlation between two consecutive frames is interpreted as a camera break.

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

Speech recognition

The purpose of speech recognition is the generation of a transcript to be used as a support for retrieval

Main functionality required

- Speaker independent
- Multiple languages
- Operating also with low quality audio

Does not require perfect recognition

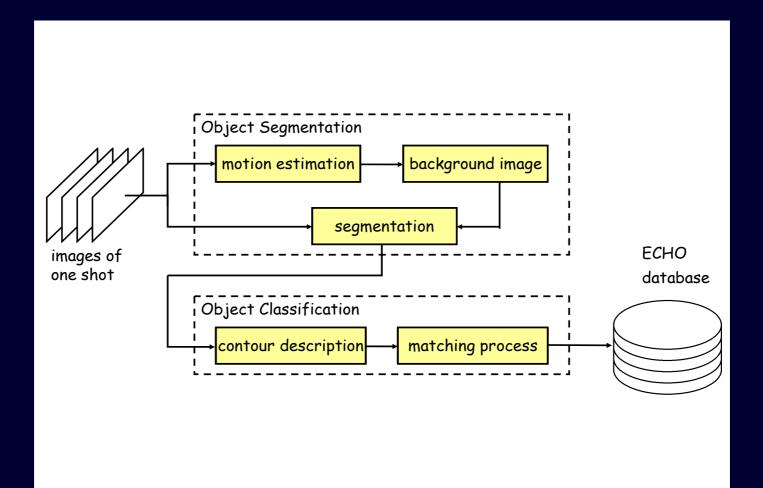
 Retrieval quality is acceptable for W.E.R. up to 30-40%

Object detection and recognition

- The system for moving-object recognition consists of two components, a segmentation module and a classification module.
- For each shot in the video, a background panorama image is constructed. The foreground objects in this background image are removed by means of temporal filtering (median).

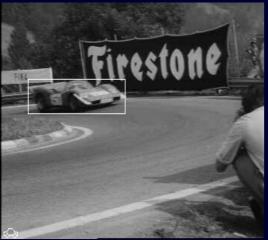
The object is segmented by comparing each frame of the video to the background image.

Phases of Object detection and recognition



Example: Cars









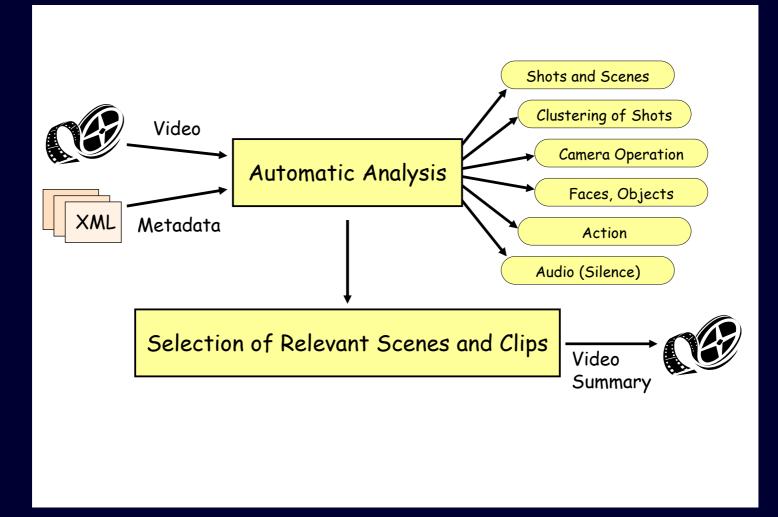




Video abstract generation [1/2]

- 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]



Representation of video content

- Metadata are used to represent the video content in order
 - To support video retrieval and navigation, as well as video management and processing

 Simple attribute values can be used as metadata to represent the video content (e.g. Dublin Core)

 or complex representations can be used to describe content information extracted from videos (e.g. MPEG-7)

 Metadata can also provide a description of video structure



Retrieval functionality



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

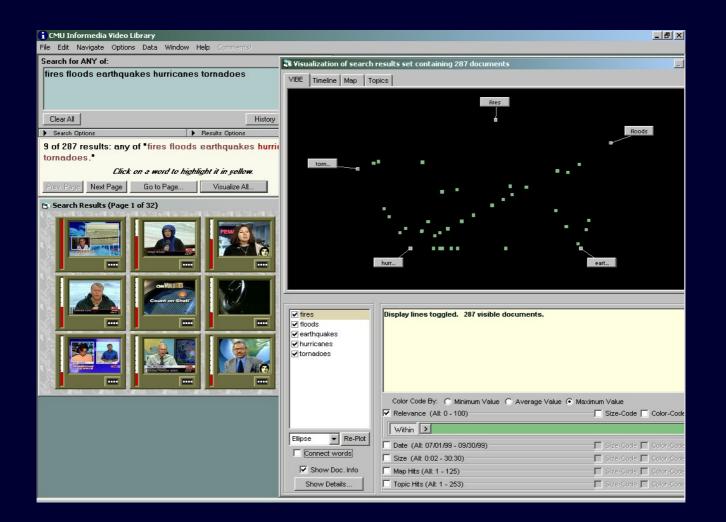
- Metadata associated to the entire video
 - E.g. find b&w videos produced before II world war by Istituto Luce
- Metadata associated to video shots
 - → E.g. find a shot where the audio transcript contains the words "Attentato Banca Nazionale dell'Agricoltura"
- Metadata associated to single frames
 > 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

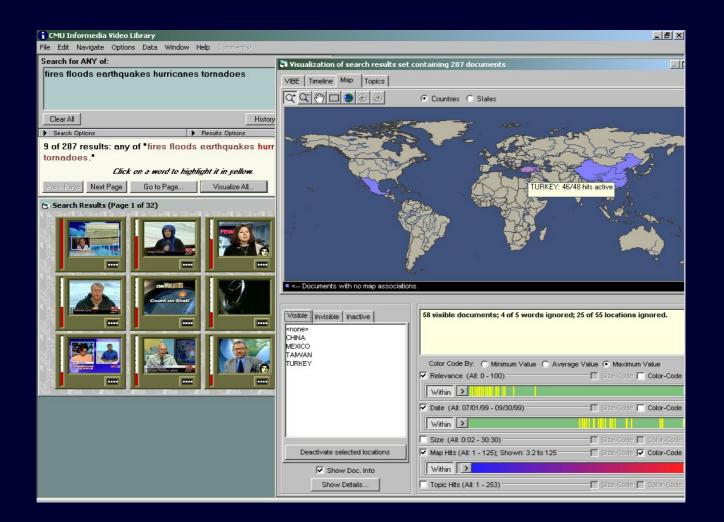
- Retrieval is based on an approximate match between the query and the retrieved videos. This is mainly the case when imprecise query elements are used (e.g. free text, images)
- Retrieved videos are returned to the user in decreasing relevance order, possibly indicating the degree of relevance of the 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

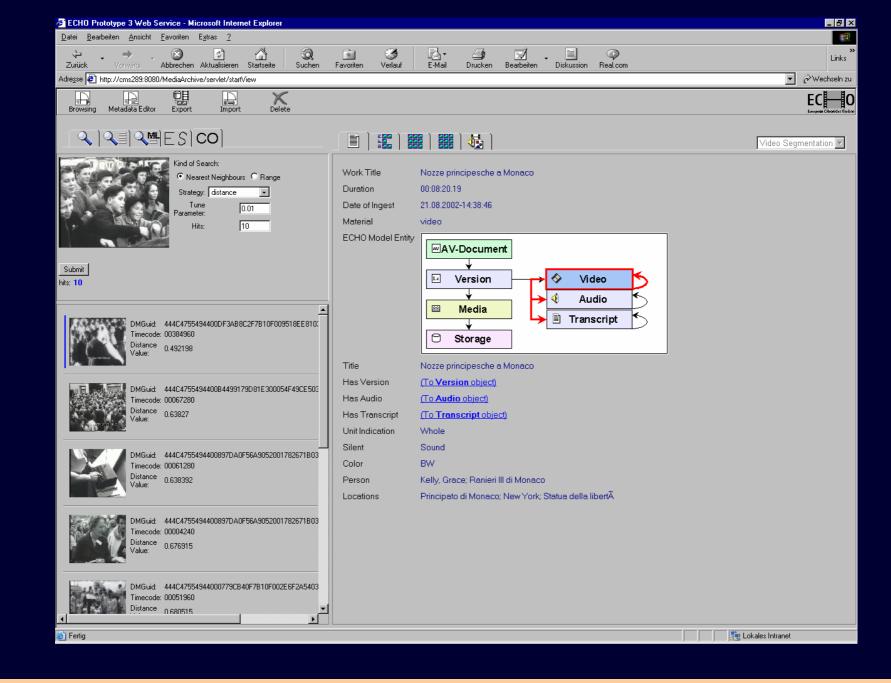


Informedia – an example



ECHO Retrieval Interface

ECHO Retrieval Web Service - Microsoft Internet Explorer				- 원	 Integration of Clients
<u>Datei Bearbeiten Ansicht Favoriten Extras 2</u>				<u> 19</u>	and Services in GUI
Browsing Player Cataloging Vetadata Edit Transcript Export Dele				Default	Views on the material
Label: Value: In Titles × Submit In Content ✓	Title	Nella sede provvisoria del R programma di governo.	teichstag il can	celliere Hitler espone il suo	Attribute Search Retrieval interface
In Dates 1933	Series Title Series Number English	Giornale Luce B0242 In the Reichstag, Hitler expla	Genre	Newsreel	Search for category level of material (Work, Expression,
sorted by	Abstract Themes	2.2 - Le Guerre Mondiali - 19 1945 Eventi Principali		Jogram.	Manifestation, Item)
 Francoforte sul Meno (Germania). I primi lavori stradali della progettata rete tedesca 00:02:04.23 IPR • 01.10.1933 	Description Language Producer Name	IT FOX Movietone	Production Date	01.01.1933	Detailed view on an item corresponding to the full ECHO data model
Nella sede provvisoria del Reichstag il cancelliere Hitler espone il suo programma c	Producer Nationality	USA	1		(structures, links,)
00:00:50.10 IPR • 01.01.1933		Whole			List of retrieved items
Studi sulla popolazione mondiale. 00:01:00.07 IPR • 08.09.1954 IPR •	Silent Audio Language	Sound e DE	Color	BW	
	Collocation Provider	B024203 IL	Storage ID	B024203	
🖉 Fertig				🗐 Lokales Intranet	59



Pasquale Savino – ISTI-CNR

