



# Video Information Retrieval: What's New Now

---

Gloria Rohmann, NYU Libraries  
DLF Spring Forum  
San Diego, April 14, 2005

# The problem: I know it's in there somewhere...

- "Gist" (what it's about)
- Genre
- Style
- Scenes
- People
- Objects
- Dialogue
- Soundtrack





# Video IR: What works?

---

- “Conventional” methods: catalogs, databases and analog previewing
- Why digitize?
- Discovering video structure
- Automatic and manual indexing
- Data models & user interfaces
- Prospects for the future: mobile and web services



# Conventional Methods: Browse and Search

---

- Structured databases:
  - AV cataloging (AACR2, MARC 21)
  - Shot lists
  - Asset management systems
- “Pathfinders” (librarians, archivists)
- Embedded markers: hints, chapters, scenes (DVD)
- Video logging systems
- Hardware browse/skim: FF, slow-mo, etc.

Result List

Search History

Recall / Request

Help

Patron  
Functions

LOGOFF

# Video Search in Libraries

<b>Publisher</b>	Alexandria, VA : PBS Home Video ; Atlanta : Turner Home Entertainment [distributor], 1995, c1987.
<b>Description</b>	7 videocassettes (120 min. each) : sd., col. with b&w sequences ; 1/2 in.
<b>Notes</b>	PBS Home Video ; Turner Home Entertainment: B5424 (set) ; B3268-3274 (vols.) Vols. 1-6 have subtitle: America's civil rights years, 1954 to 1965; v. 7-14 have subtitle: America at the racial crossroads, 1965 to 1985.
<b>Contents</b>	Vol. 1. Awakenings, 1954-1956 and Fighting back, 1957-1962 -- v. 2. Ain't scared of your jail, 1960-1961 and No easy walk, 1961-1963 -- v. 3. Mississippi: Is this America?, 1962-1964 and Bridge to freedom, 1965. -- v. 4. The time has come, 1964-1966 and Two societies, 1965-1968 -- v. 5. Power! 1966-1968 and The promised land, 1967-1968 -- v. 6. Ain't gonna shuffle no more, 1964-1972 and A nation of laws, 1968-1971 -- v. 7 The keys to the kingdom, 1974-1980 and Back to the movement, 1979-mid 1980s.
<b>Performer note</b>	Julian Bond.
<b>Credits</b>	Vols. 7-14 produced, directed, and written by Madison Davis Lacy, Jr. and James A. DeVinney. Vols. 1-6 tell the story of America's civil rights years from 1954 to 1965; v. 7-14 examines the new America from 1966 to 1985, from community power to the human alienation of urban poverty.
<b>Technical details</b>	VHS.
<b><u>Subject (LCSH)</u></b>	<a href="#">African Americans -- Civil rights -- United States -- History -- 20th century.</a> <a href="#">African Americans -- Suffrage -- History -- 20th century.</a> <a href="#">Segregation -- United States -- History.</a> <a href="#">African Americans -- History -- 20th century</a> <a href="#">Civil rights demonstrations -- United States -- History -- 20th century</a> <a href="#">Civil rights -- United States</a> <a href="#">United States -- Race relations -- History -- 20th century</a>
<b>Other author(s)</b>	<a href="#">Vecchione, Judith</a> <a href="#">Hampton, Henry, 1940-</a> <a href="#">Bond, Julian, 1940-</a>

505:  
Contents

520: Summary

650: Subject  
headings



# Enhanced metadata: shot lists, transcripts: Open University video collection

*Ever wondered about.... chocolate? / produced by Lisa MacHale ;...*

Relevance: ● ● ● ● ●

**Author:** Open University EWFood/10

**Other Author(s):** Coxon, Alan

Sykes, Kathy

Stanes, Sara Jayne

Horrocks, Sally

Chinn, Carl

Broomfield, Margaret

Barron, Robert

Coady, Chantal

Bilsborough, Tony

Peters, David

Valender, Alan

Hetherington, Marion

Kilkast, David

MacHale, Lisa

**Title:** Ever wondered about.... *chocolate?* / produced by Lisa MacHale ; contributors Alan Coxon, Kathy Sykes, Sara Jayne Stanes, Sally Horrocks, Carl Chinn, Margaret Broomfield, Robert Barron, Chantal Coady, Tony Bilsborough, David Peters, Alan Valender, Marion Hetherington, David Kilkast

**Primary Material:** Visual Material

**Physical Description:** Videocassette

**Subject(s):** Food Great Britain History.

Quakers History.

**Publisher:** BBC/Open University

**Description:** Video, 28min., 50sec. VHS

**Summary:** This programme looks at Britain's passion for *chocolate* and its history.

**Date of transmission:** 2002

**Notes:** Open University course EWFood

Footage description available

Script (ref only)

**Linked resources:** [Website](#)



Next

Location

Full View

Contents/Footage

*Ever wondered about... chocolate? / produced by Lisa Mastale;*

Relevance: ● ● ● ● ●

**Contents:** Footage - Opening credits. *Chocolate* factory workers. Alan Coxon and Kathy Sykes preparing food. Man biting into *chocolate* bar (0'00"-0'50") Alan opening fridge and walking over to Kathy at table. Kathy grating orange. Alan showing ingredients for cheesecake. Cooking *chocolate*. Alan and Kathy breaking *chocolate* and smelling it. Breaking *chocolate*. Kathy tasting *chocolate* (0'51"-2'00") Boy emerges from shop and holds up Lion bar. Boy emerges from shop and holds up Dairy Milk bar. Girl emerges from shop and holds up Galaxy bar. Girl emerges from shop and holds up Mars Bar. Box of chocolates. B/w 1800s prints - "Drink Cadbury's cocoa" and "Frys pure concentrated cocoa". Sara Jayne Stavnes speaking. Old fashioned confectioners. man behind window opening box of Frys chocolates (2'01"-2'50") Sara sneaking. Frys *chocolate*. Print of

Footage - Opening credits Chocolate factory workers.  
Alan Coxon and Kathy Sykes preparing food.  
Man biting into chocolate bar (0'00-0'50")  
Alan opening fridge and walking over to Kathy at table.  
Kathy grating orange. Alan showing ingredients for  
cheesecake.  
Cooking chocolate. Alan and Kathy breaking chocolate  
and smelling it.  
Breaking chocolate.Kathy tasting chocolate (0'51"-2'00")

and Kathy preparing cheesecake in kitchen. Cocoa powder (14'44"-15'46")

Next



# Browse and skim: DVD's Digital Advantages

Pause, FF, rewind

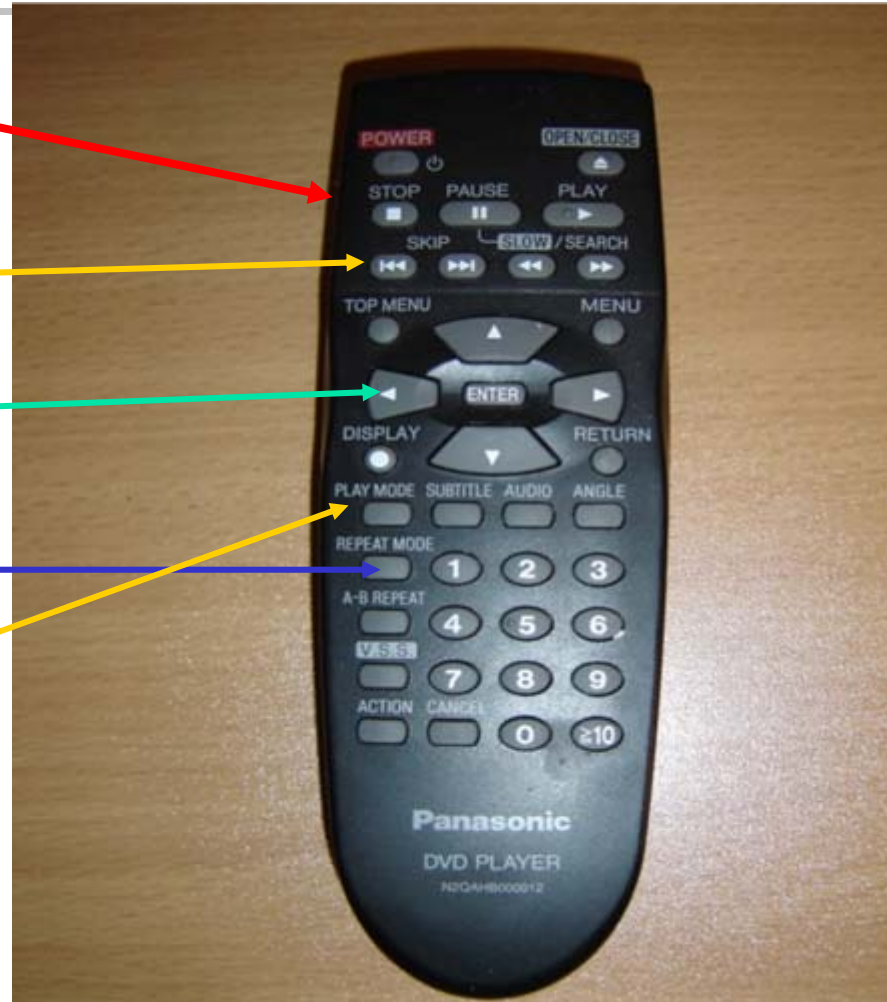
Navigate

Frame-by-frame  
menus, chapters or  
tracks

Insert markers, repeat  
play

Change audio, subtitle  
languages, show  
closed captioning

Shuttle/scrub onscreen



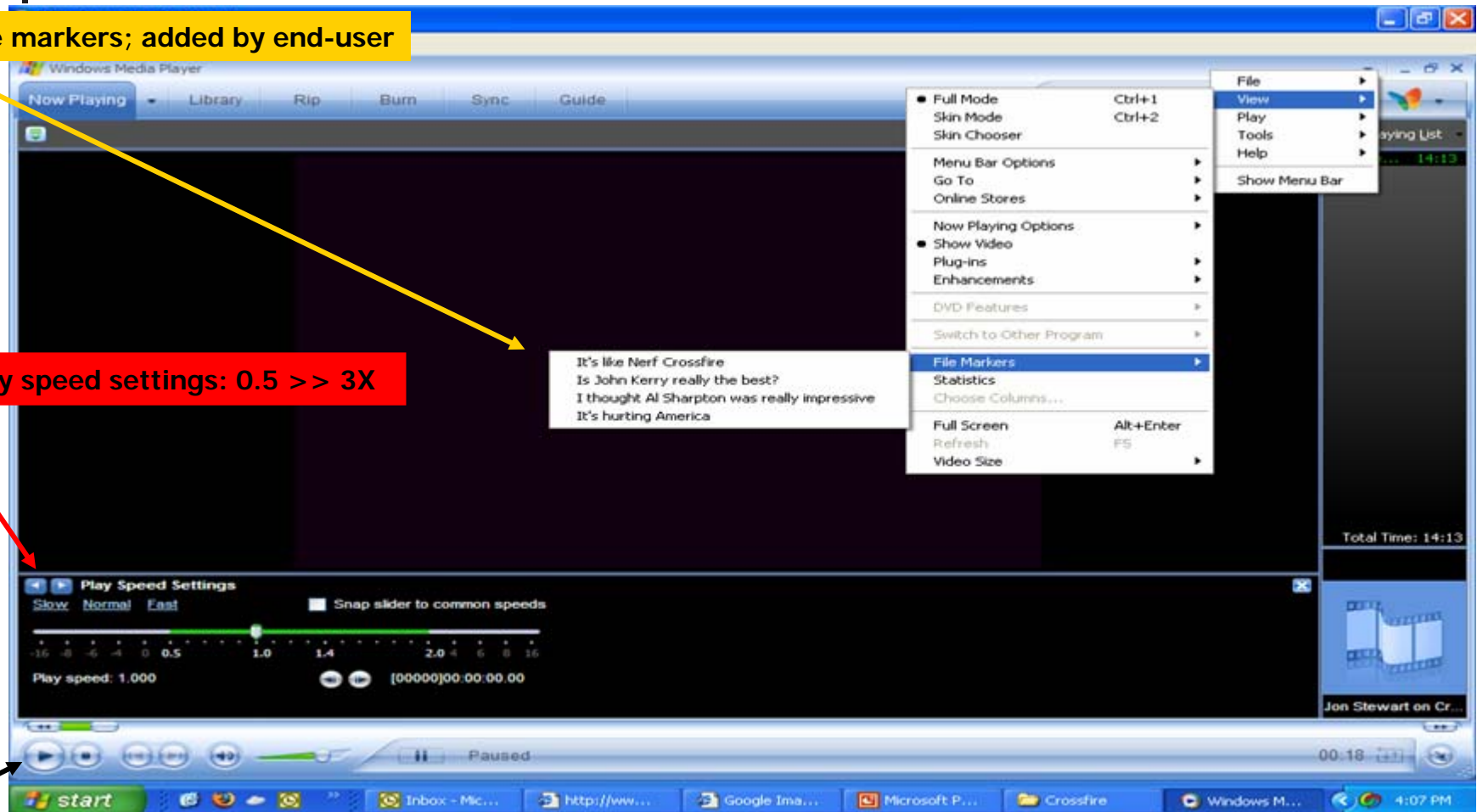


# Media Player Example

DVD player clones; can be enhanced with SDKs

File markers; added by end-user

Play speed settings: 0.5 >> 3X



Start, stop, pause, rewind to beginning, FF to end, advance by frame



# What Is Video?

---

Authored video has:

- Series of still images @25-30 fps
- Structure: frames >> shots >> scenes
- MODALITIES
  - (Audio tracks)
  - (Text: captioning, subtitles, etc.)
  - (Graphics: logos, running tickers etc.)
- Production metadata: timestamp, datestamp, flash on/off



# Why Digitize Video?

---

**Store and deliver over networks**

**Allow analysis by computers**

**Allow auto & manual indexing**

**USING:**

- Image processing
- Signal processing
- Information visualization



# Why Compress Video?

---

- 1 frame (@TV brightness) = 0.9 megabytes (MB) of storage
- At 29 fps, each second = 26.1 MB of storage
- 30 minute film = 53 gigabytes (GB) of storage

**OBJECT: Make file smaller; retain as much information as possible**



# Encoding Formats

---

All use some kind of compression; similar encoding methods—many CODECS—some “lossy,” others “lossless”

- AVI: audio-video interleave or interactive
- QuickTime
- MPEG family: MPEG-1, 2, 4
- H261: for video conferencing
- New: H264; JPEG 2000



# CODECS

---

- *Compressor/Decompressor, or Coder/Decoder*
- Produce and work with encoding formats.
- Central to compression and encoding; perform signal and image processing tasks
- Examples: Cinepak, Indeo, Windows Media Video.
- MPEG-4: DivX, Xvid, 3ivX implementations of certain compression recommendations of MPEG-4.



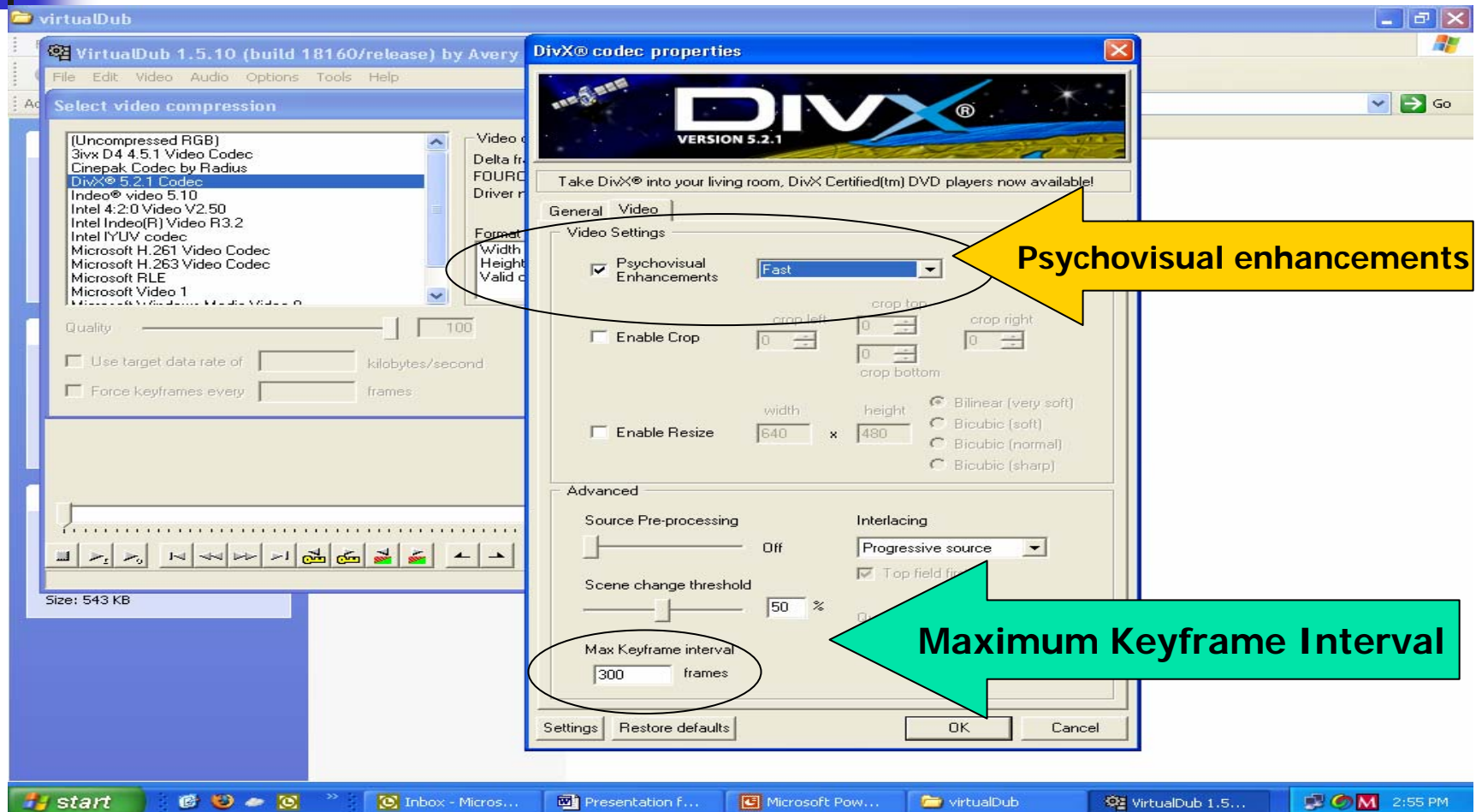
# Human Psychovisual System

---

- Movement creates “temporal aliasing”: human eye/brain fills in the gaps
- Blurring produced by camera shutter softens edges
- Modeled by CODECS and algorithms
- Goal: acceptable facsimile of moving scene



# Configuring CODECS for analysis



# What looks best to you?





# Encoding Methods: predictive

---

- **Sampling:** value of function @ regular intervals (example: brightness of pixels)
- **Quantization:** frequency of sampling (1 in 10 vs. 1 in 100 frames)
- **Discrete cosine transforms (DCT)** an *array* of data (not just one pixel) is transformed into another set of values.
- **Inter-frame vs. Intra-frame** encoding



# Building Video Indexes

---

Same as any indexing process...decide:

- What to index: granularity
- How to index: modalities (images, audio, etc.)
- Which features?
- Discover spatial and temporal structure:  
deconstructing the authoring process
- Construct data models for access

# Video Structure

Video  
Scene  
Shot  
Frame





## Using Encoding to Discover Structure

---

**This link is to a QT movie not included in this presentation.**



# Shot Boundary Detection

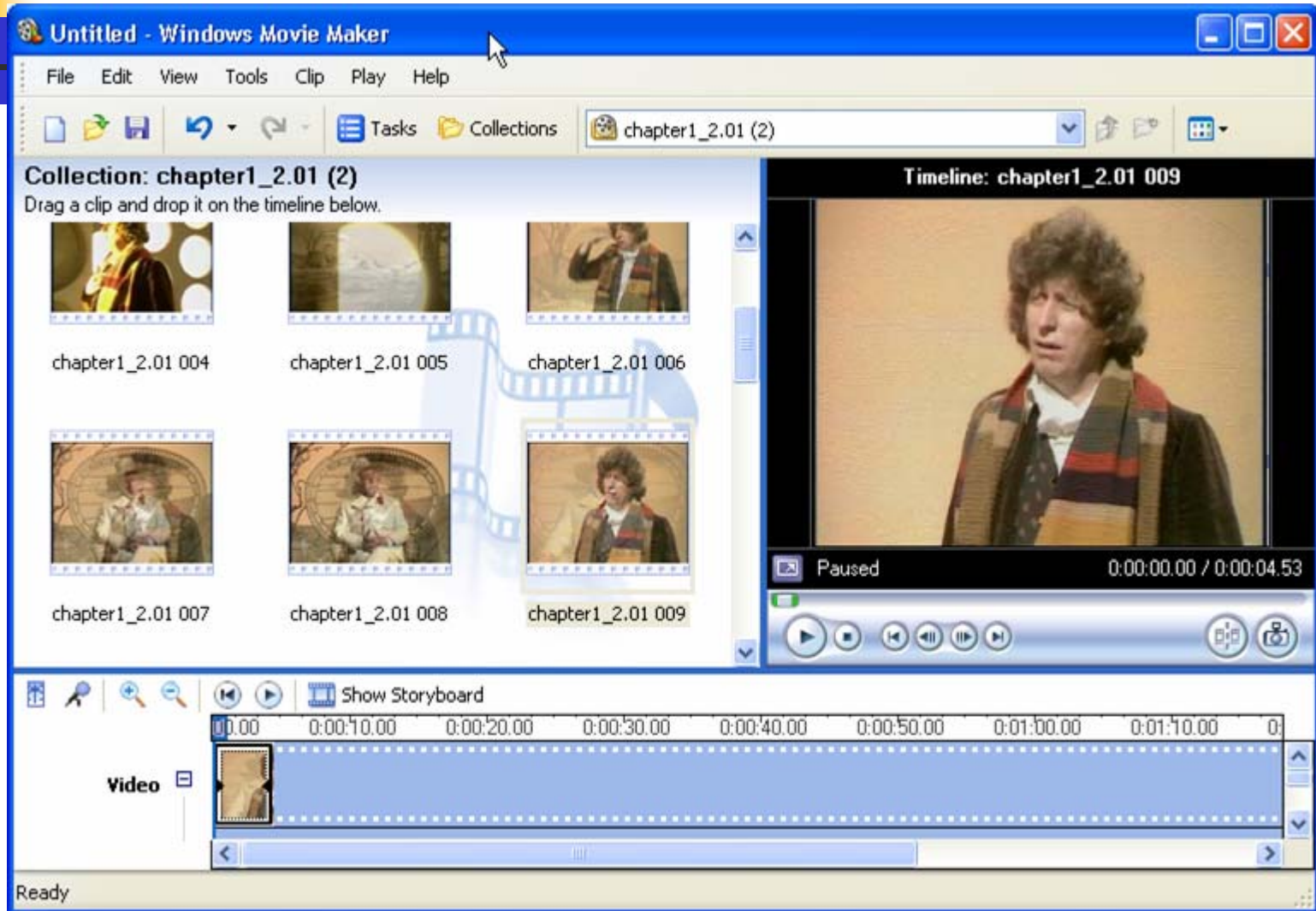
---

Algorithms that compare the similarities between nearby frames. When the similarities fall below a pre-determined level, the limit of a “shot” is automatically defined:

- Edge detection
- Compare color histograms
- Compare motion vectors



# Clip Creation with NLEs





# Spatial & Temporal Segmentation

---

1. Use shot boundary detection and keyframes to define shots & choose representative frames
2. Use CBIR (Content-based Image Retrieval) techniques to reveal features in representative frames  
(shapes, colors, textures)



# CBIR Techniques

---

Images (frames) have no inherent semantic meaning: only arrays of pixel intensities

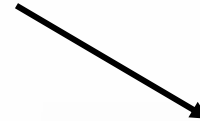
- Color Retrieval: compare histograms
- Texture Retrieval: relative brightness of pixel pairs
- Shape Retrieval: Humans recognize objects primarily by their shape
- Retrieval by position within the image

# MPEG-4: Content-based encoding



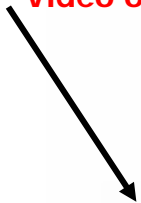
(a)

Video object plane (VOP)



(b)

Video object plane (VOP)



(c)



Background encoded only once

(d)

AMOS:

# Tracking Objects Beyond the Frame



<http://www.ctr.columbia.edu/~dzhong/rtrack/demo.htm>



# "Are We Doing Multimedia?"\*

## Multimodal Indexing

---

Ramesh Jain: "To solve multimedia problems, we should use as much context as we can."

- Visual (frames, shots, scenes)
- Audio (soundtrack: speech recognition)
- Text (closed captions, subtitles)
- Context—hyperlinks, etc.

***\*IEEE Multimedia. Oct-Nov. 2003***

[http://jain.faculty.gatech.edu/media\\_vision/doing\\_mm.pdf](http://jain.faculty.gatech.edu/media_vision/doing_mm.pdf)

# Multimodal Indexing

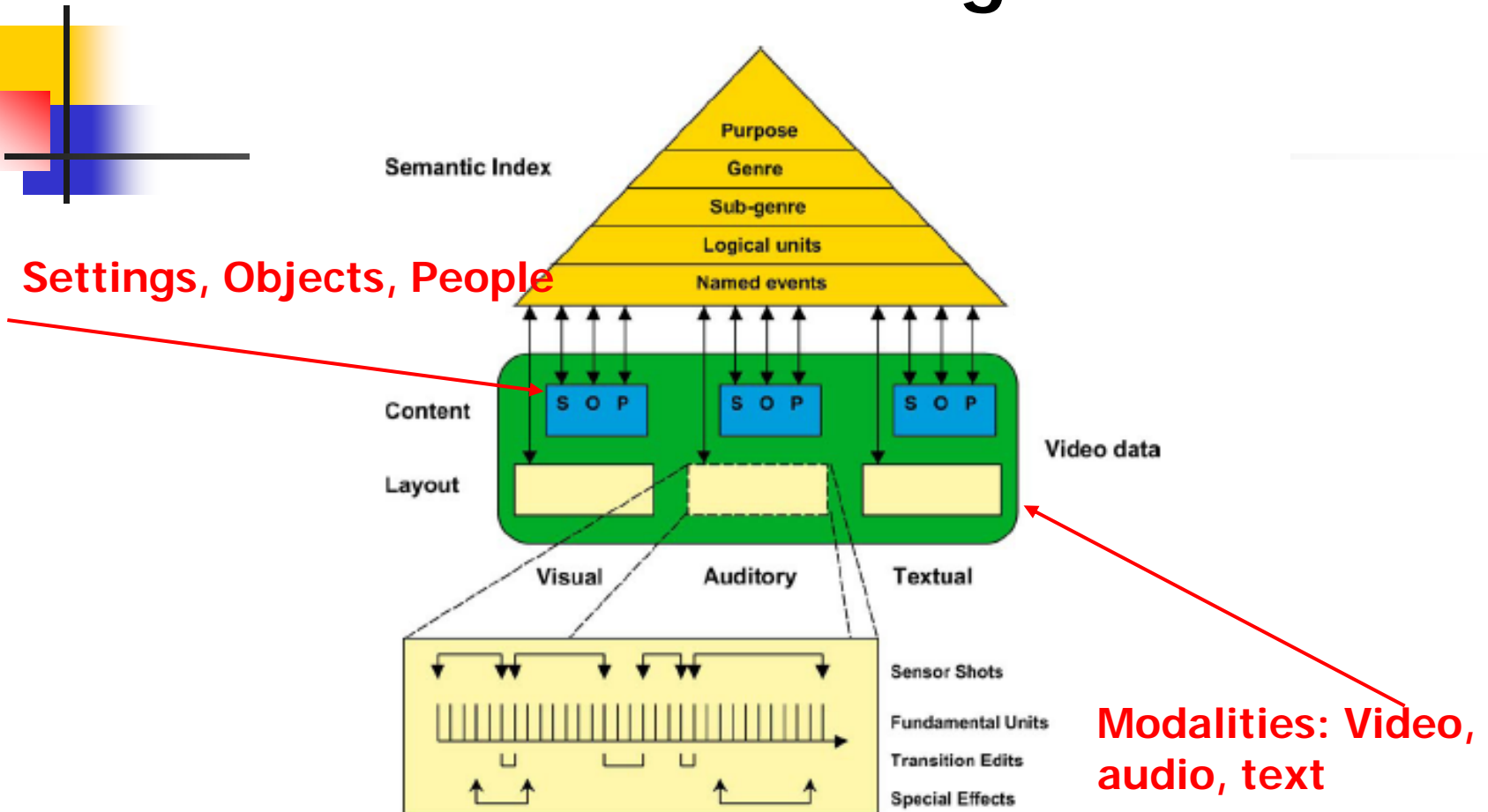


Figure 1. A unifying framework for multimodal video indexing based on an author's perspective. The letters S, O, P stand for setting, objects, and people. An example layout of the auditory modality is highlighted, the same holds for the others.



# Virage VideoLogger™

The screenshot displays the Virage VideoLogger software interface with several panels and annotations:

- Monitor (Live):** Shows a video feed of a man in a suit. Below the feed, a text box contains the caption "LOOKS AT JOYCE CAROLE OATES' NEW". A red arrow points to the SMPTE timecode "00:22:45:23".
- Clips (20 total, 1 selected):** A table listing video clips with columns for In SMPTE, Out SMPTE, Slug, and Description. An arrow points to the "Mark & annotate clips" text.
- Keyframes (296 total, 5 selected):** A grid of keyframe thumbnails. An arrow points to the "Keyframes" text.
- Speech (68 total, 4 selected):** A table listing extracted speech segments with columns for In SMPTE, Out SMPTE, and Speech. An arrow points to the "Text or audio extracted automatically" text.

In SMPTE	Out SMPTE	Slug	Description
00:00:00:00	00:00:10:18		Tape intro
00:00:00:00	00:02:04:10		... ADDITIONAL SKILLS.
00:00:10:18	00:00:17:01		the months after graduation from happen with those numbers coming a
00:00:17:01	00:00:45:01		Richard Smith
00:00:45:01	00:01:02:02	ROOKIES	
00:01:02:02	00:01:07:19		speech
00:01:07:19	00:01:27:19		

In SMPTE	Out SMPTE	Speech
00:01:15:00	00:01:20:00	if higher here's ready to hit the London stage reports say the House reveals
00:01:20:00	00:01:25:03	far as signing up for the lead in the Philadelphia Story anwhirl with play on st...
00:01:25:03	00:01:30:16	her co-star in the flag saw the flight the Academy Awards wrapped up yes...
00:01:30:16	00:01:35:20	F. who he was Altai the hon gloves well hidden now on the lives
00:01:35:20	00:01:37:28	out of this type of them who has
00:11:49:01	00:11:54:05	on a proposal to end affirmative action in Florida higher education shields i...
00:11:54:05	00:11:59:08	wrath of the week in politics if essays Taylor Fleming looks a Joyce Carol O...
00:11:59:08	00:12:05:20	new novel on Marilyn Monroe film follows our summary of the news office F...
00:12:05:20	00:12:10:23	were signs today the economy may be slowing the Labor Department said ...
00:12:10:23	00:12:15:27	to in up tenths of from American
00:12:15:27	00:12:21:06	businesses got a jobs if the worst showing in years the
00:12:21:06	00:12:26:15	Commerce Department said factory orders fell for in the most since
00:12:27:23	00:12:33:05	on Wall Street the news fears of inflation drove the major market indexes
00:12:33:05	00:12:39:00	to new gains the Nasdaq index: rose points or more than to
00:12:39:00	00:12:44:28	finish at the Dow Jones Industrial Average was up on points
00:12:45:08	00:12:50:10	to close at Clinton today urged the
00:12:56:28	00:13:02:00	sure Russia becomes fully part of your spoke in off in Germany after
00:13:02:08	00:13:07:15	receiving the Charlemagne Prize for from voting European uniform travels t...

Text or audio extracted automatically



# Building Video Indexes: Structured modeling

---

Use statistical classification methods to predict relationship *between shots*:

- Pattern recognition
- Hidden Markov Models
- SVM (support vector machines)
- Neural networks
- Relevance feedback via machine learning



# Annotation: Metadata Schemes

---

- MPEG-7
- MPEG-21
- METS
- SMIL

# IBM MPEG-7 Annotation Tool

File Tools Lexicon View Help

Shot Annotation

Key Frame:



Events:

- ☐ Water\_Skiing
- ☐ Boat\_Sailing
- ☐ Person\_Speaking
- ☐ Landing
- ☐ Take-Off/Launch
- ☐ Explosion
- ☐ Multiple\_Shots
- ☐ Blank

Static Scene:

- ☐ Nature\_(Low-level)
  - ☐ Sky
  - ☐ Cloud
  - ☐ Water
  - ☐ Snow
  - ☐ Greenery
  - ☐ Rock
  - ☐ Land
- ☐ Nature(High-level)
  - ☐ Mountain
  - ☐ Beach
  - ☐ Field
  - ☐ Forest
  - ☐ Canyon
  - ☐ Desert
  - ☐ Waterfall
- ☐ Man-Made
  - ☐ Road
  - ☐ Cityscape

Key Objects:

- ☐ Animal
  - ☐ Deer
  - ☐ Bird
  - ☐ Duck
- ☐ Human
  - ☐ Person(w/front)
  - ☐ Person(w/side\_
  - ☐ Person(w/o\_fac
  - ☐ People(countab
  - ☐ People(uncoun
- ☐ Man-Made\_Structu
  - ☐ Building
  - ☐ Dams
  - ☐ Statue
  - ☐ Swimming\_Poo
- ☐ Man-Made\_Object
  - ☐ Whiteboard
  - ☐ Slide\_Presenta
  - ☐ Robot

Keywords:

Clear

OK

Play

FF

FFF

Stop

Shot Information

Shot Number: 5

Start Frame: 1370

End Frame: 1437

**IBM MPEG-7**  
**Annotation Tool**  
 Copyright 2000

Frames in the Shot

Shots in the Video

Prev

Next



Indoors

Human

Person(w/side\_vie

People(countable)

Person\_Speaking

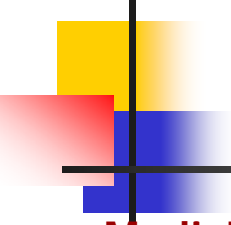
Indoors

Person(w/side\_vie

People(countable)

Person\_Speaking

# MPEG-7 Output from IBM Annotation Tool



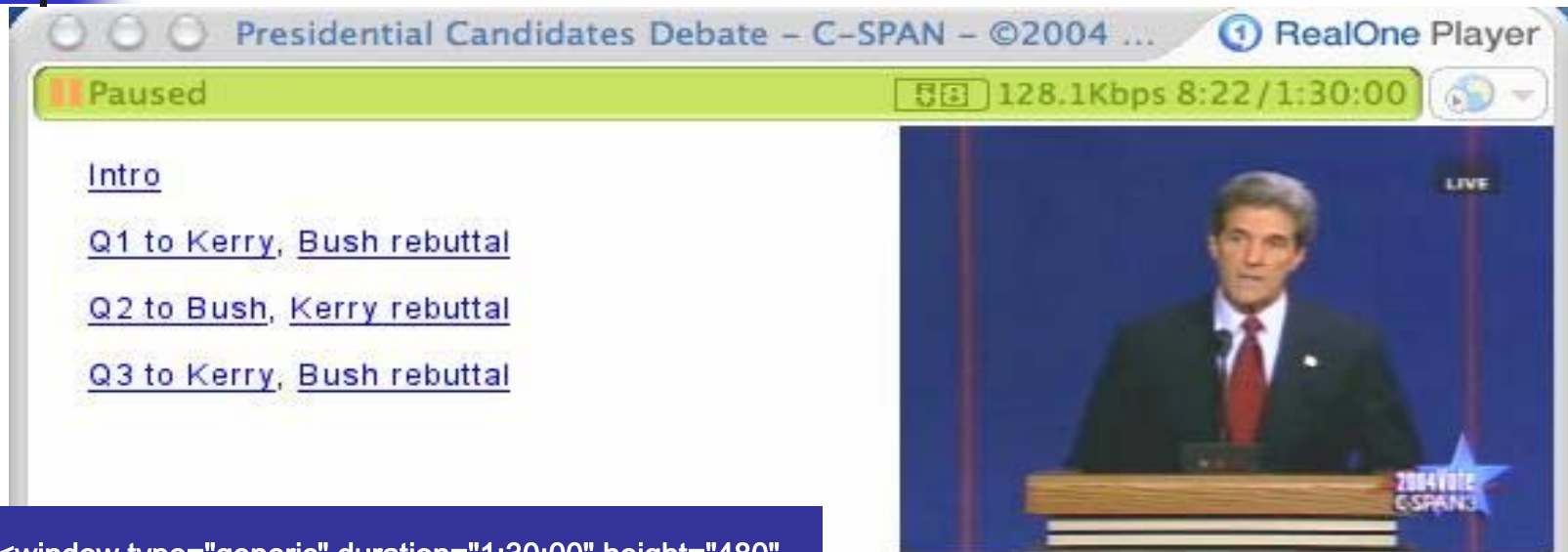
```
- <MediaTime>
  <MediaTimePoint>T00:00:27:20830F30000</MediaTimePoint>
  <MediaIncrDuration
mediaTimeUnit="PT1001N30000F">248</MediaIncrDuration>
  </MediaTime>
- <TemporalDecomposition>
- <VideoSegment>
- <MediaTime>
  <MediaTimePoint>T00:00:31:23953F30000</MediaTimePoint>
  </MediaTime>
- <SpatioTemporalDecomposition>
- <StillRegion>
- <TextAnnotation>
  <FreeTextAnnotation>Indoors</FreeTextAnnotation>
  </TextAnnotation>
- <SpatialLocator>
  <Box mpeg7:dim="2 2">14 15 351 238</Box>
  </SpatialLocator>
  </StillRegion>
```

Duration of shot in frames

Annotation

Location and dimension of spatial locator in pixels

# SMIL: Hypertext & Hypermedia



```
<window type="generic" duration="1:30:00" height="480"
width="320"
underline_hyperlinks="true" /> <font face="arial" size="2">
<ol>
<li><a href="command:seek(0:0)"
target="_player">Intro</a></li>
<br/> <li> <a href="command:seek(2:10)"
target="_player">Q1 to Kerry</a>,
<a href="command:seek(4:26)" target="_player">Bush
rebuttal</a> </li>
```

Q2 to Bush: Do you believe the election of Senator Kerry on November the 2nd would increase the chances of the U.S. being hit by another 9/11-type terrorist attack?



# Data Models for Video IR

---

- Based on records (DBMS)
- Semi-structured (video + XML or hypertext): MPEG-7, SMIL
- Based on context: Yahoo Video, SingingFish





# Scholarly “Primitives”\*

---

Low-level methods for higher-level research

- Discovering
- Annotating
- Comparing
- Referring
- Sampling
- Illustrating
- Representing

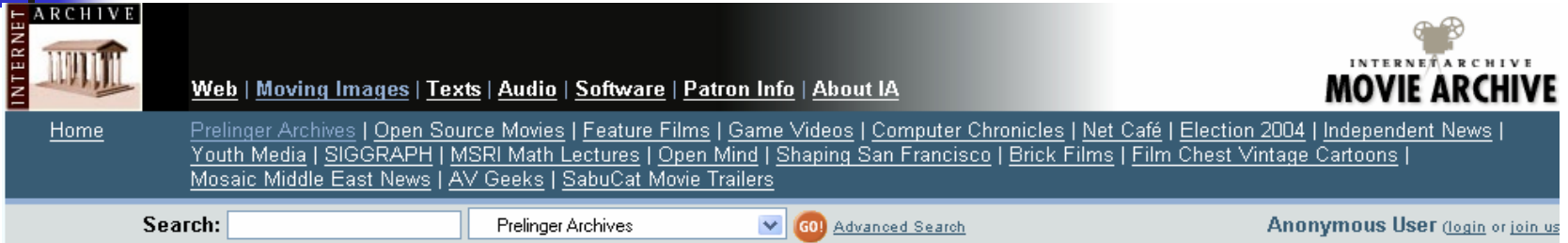
\*Unsworth, John. (2000) *Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?*



# User Interfaces for Video IR

---

- Discovering
- Annotating
- Comparing
- Referring
- Sampling
- Illustrating
- Representing
- Browse, query text
- Browse **surrogates**
- Interactive filtering: dynamic query based on visual aspects
- Interactive zooming
- Interactive distortion
- Compare results for feedback
- Annotate results



\_\_\_\_\_



Anonymous User ([login](#) or [join us](#))

[Return to Program Details](#)

Below are images for each minute in the program.





# IBM Research MARVel

---

MPEG-7 video search engine

- Manual annotations are used for machine learning
- Automatic multimodal indexing
  - Image processing--CBIR, histogram matching
  - Automatic speech recognition
- Structured modeling: clustering by comparing features



# Video Search on the Web: Yahoo

---

- Using existing (text) metadata
- Bought AltaVista media search engine
- Does not analyze content of media stream
- Horowitz: “Web pages are self-describing”
- Analyze the web page around the link
- Analyze the metadata included in video file
- Media RSS: publishers can add links to multimedia within feed





# Video Search on the Web: Google

---

- Using metadata **in the video stream**
- Almost all broadcast news video is closed captioned
- Google ingests video with closed captioning
- Transcripts are created linked to time-code
- Transcripts are indexed
- “Thumbnails” grabbed at time intervals
- Still text-based; “thumbnails” provide visual surrogate





social security

Search Video

[Preferences](#)

## Video

Results 1 - 10 of about 2137 for **social security** (0.05 seconds)[ABC 7 News](#)

... I'm Carolyn Johnson. >> And I'm Dan Noyes. Members of Congress from the Bay area got an earful today from people that were worried about **social security**. Reform of program is number one domestic issue for President Bush but it's also the concern for many seniors. More on the Town Hall meeting. >>...

KGO - ABC Network - Sat Feb 26 2005 at 6:00 PM PST - 30 minutes

[Fox Magazine](#)

... made possible by fox news channel and the U.S. Department of Education] Neil: the next time you hear a congressman or Senator bashing **social security** reform, ask them this question, If the system is so great, why aren't You in it? You have choices in your retirement plan. You can take it from a variety...

Fox News Channel - Sun Feb 20 2005 at 11:00 PM PST - 1 hour

[ABC World News Tonight](#)

The latest world and national news.

... On "world news tonight" -- President Bush says those who oppose his plans for **social security** are engaging in scare tactics. Democrats say there is reason to be scared. >>> Tonight, what the President's ideas are for people in their 40s and early 50s. >>> Tracking down Child pornographers. There...

KGO - ABC Network - Fri Feb 4 2005 at 5:30 PM PST - 30 minutes

[Forbes on Fox](#)

... you are always bitter. Here is David asman, Forbes Forks. David: "the Cost of Freedom" Continues with "Forbes on fox." Forget about **social security**. We need to change our tax code. Why that's the right battle for the President If he wants to come out on top. And put more money in all our pockets....

Fox News Channel - Mon Feb 14 2005 at 2:00 AM PST - 30 minutes





## About this show

[ABC 7 News](#)

next airs on Sat Mar 5  
2005 at 6:00 AM PST on  
KGO

[Upcoming episodes](#) of  
ABC 7 News  
for zip code 94043.  
[\[edit location\]](#)

[Google web search](#) for  
ABC 7 News

## ABC 7 News

KGO - ABC Network - Sat Feb 26 2005 at 6:00 PM PST - 30 minutes

Below is a summary of matches for **social security**.

Video is currently not available. [More information >](#)



at 9 minutes

>>> Good evening. I'm Carolyn Johnson. >> And I'm Dan Noyes. Members of Congress from the Bay area got an earful today from people that were worried about **social security**. Reform of program is number one domestic issue for President Bush but it's also the concern for many seniors.



at 9 minutes 30 seconds

More on the Town Hall meeting. >> Dan, the President's plan would allow working Americans to contribute a third to **social security** into personal retirement accounts. The President says that this plan would not affect current beneficiaries but many are feeling uneasy about it and at times, emotional. >> President Bush says his plans to overhaul **social security** will not affect the seniors but it didn't stop about 500 people from seeking reassurance.



at 10 minutes

>> I have a handicapped daughter. She is getting disability from my **security** and husband's. He's long deceased but I don't know what I would do without it. >> They hosted a Town Hall meeting. All democrats, they dispute the President's claim that it's in crisis and needs fixing now.



at 11 minutes

>> They say democrats don't have a plan because they don't want to change the current system. So the recommendation include taxes and incomes over \$90,000 and deferring benefits. >> The idea that in order to save it we have to destroy it. >> The truth is borrowing to fund by this scheme actually puts everybody's **social security** at risk.



# Evaluation of Video IR : The Open Video Project

---

## Testing:

- Users' perception of relevance
- Domain-specific vs. non-domain-specific users

## Findings:

- Leverage digital video medium
- Do not mimic analog or text searching



# Video IR: What's Needed

---

- Extensibility of textual descriptors
- Search based on visual features
- Visual representations of results
- Visual interfaces that interactively enable refinement of results
- Visual representations of relationship between results--hierarchies
- Ability to bookmark segments for retrieval