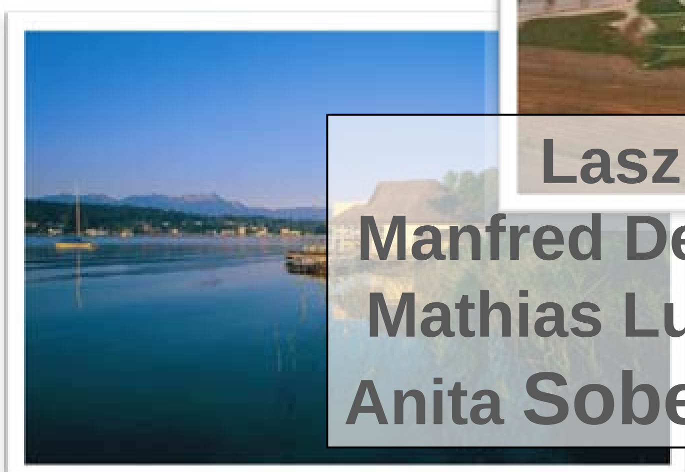
ALPEN-ADRIA
UNIVERSITÄT
KLAGENFURT

Institute of Information Technology
Distributed Multimedia Systems

Multimedia, Quo vadis?
Multimedia, What's your future?

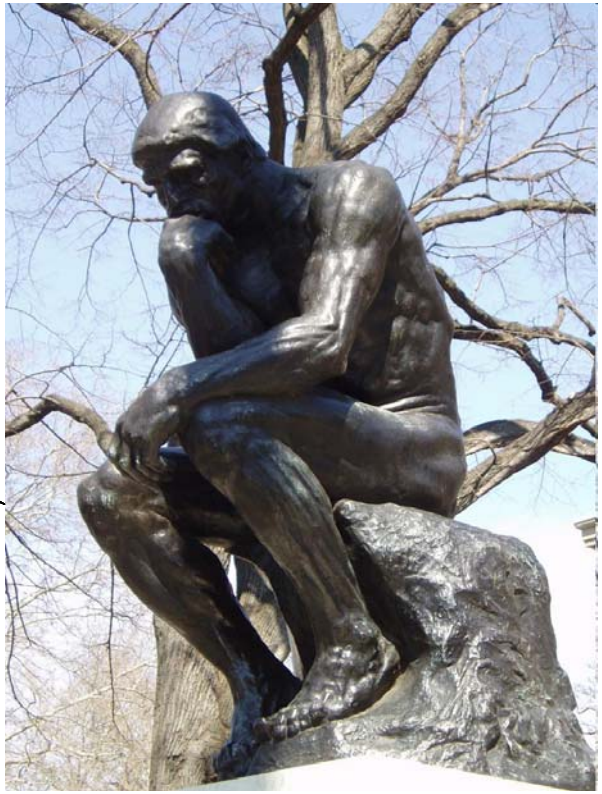


Laszlo Böszörmenyi
Manfred Del Fabro, Oliver Lampl
Mathias Lux, Klaus Schöffmann
Anita Sobe, Christian Spielvogel

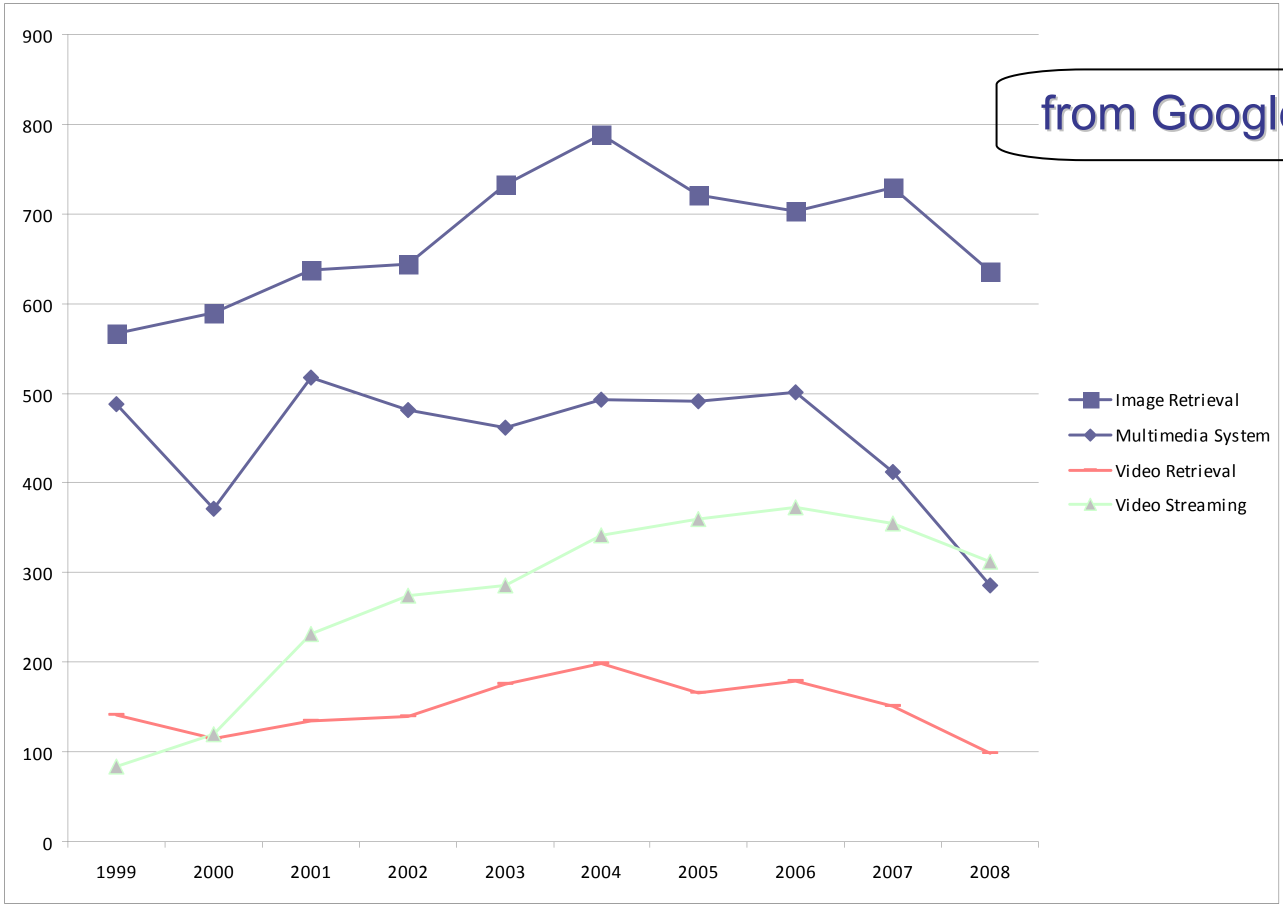
Is Multimedia Research still relevant?

- Is compression relevant?
- Streaming?
- Peer-to-peer multimedia delivery?
- Metadata?
- Social tagging?
- The semantic gap?
- Computer vision?
- Questions on Quality-of-Service?
- Multimedia engineering?
- Is multimedia pervasive as graphics? Should it be?

Rodin

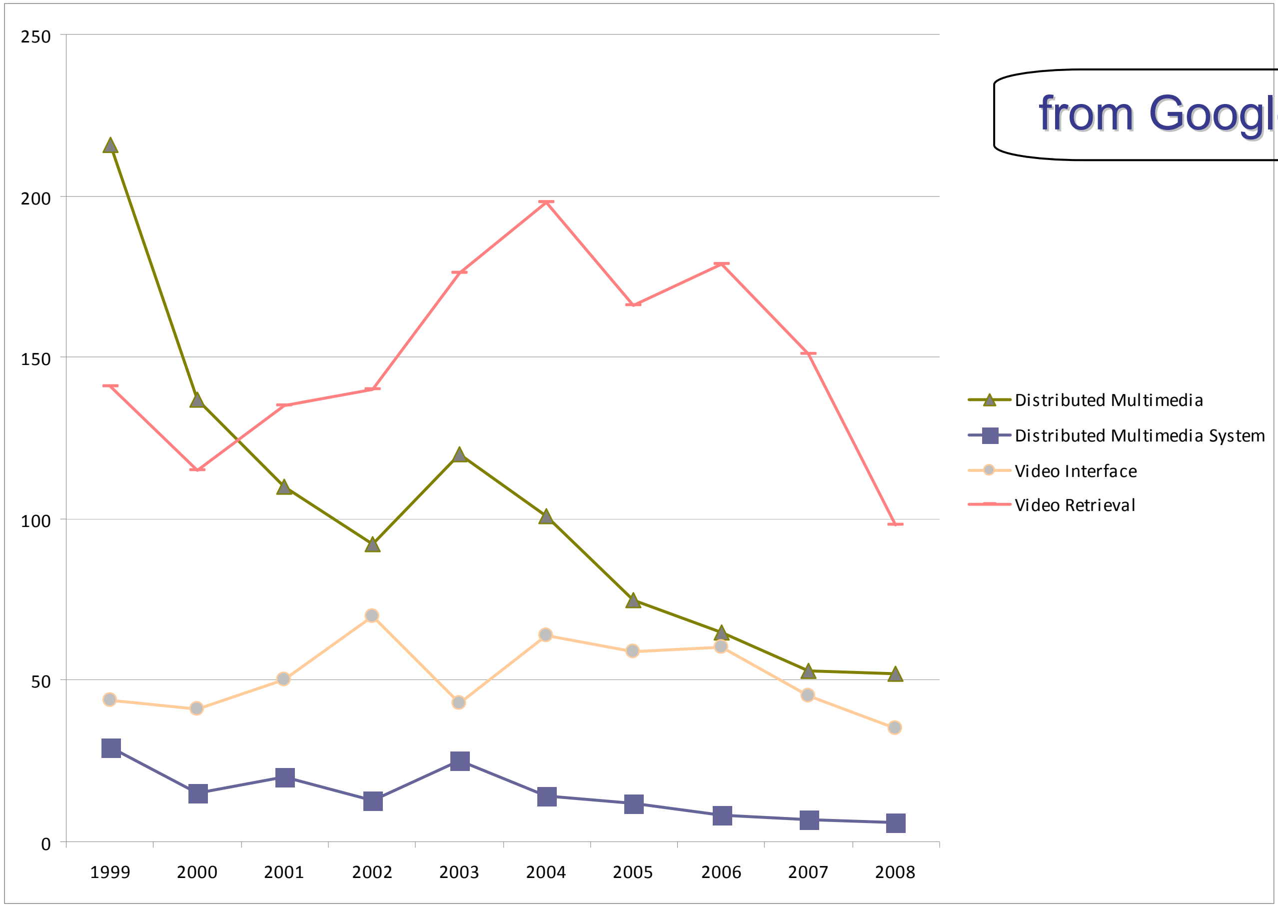


Number of research papers in MM (1)



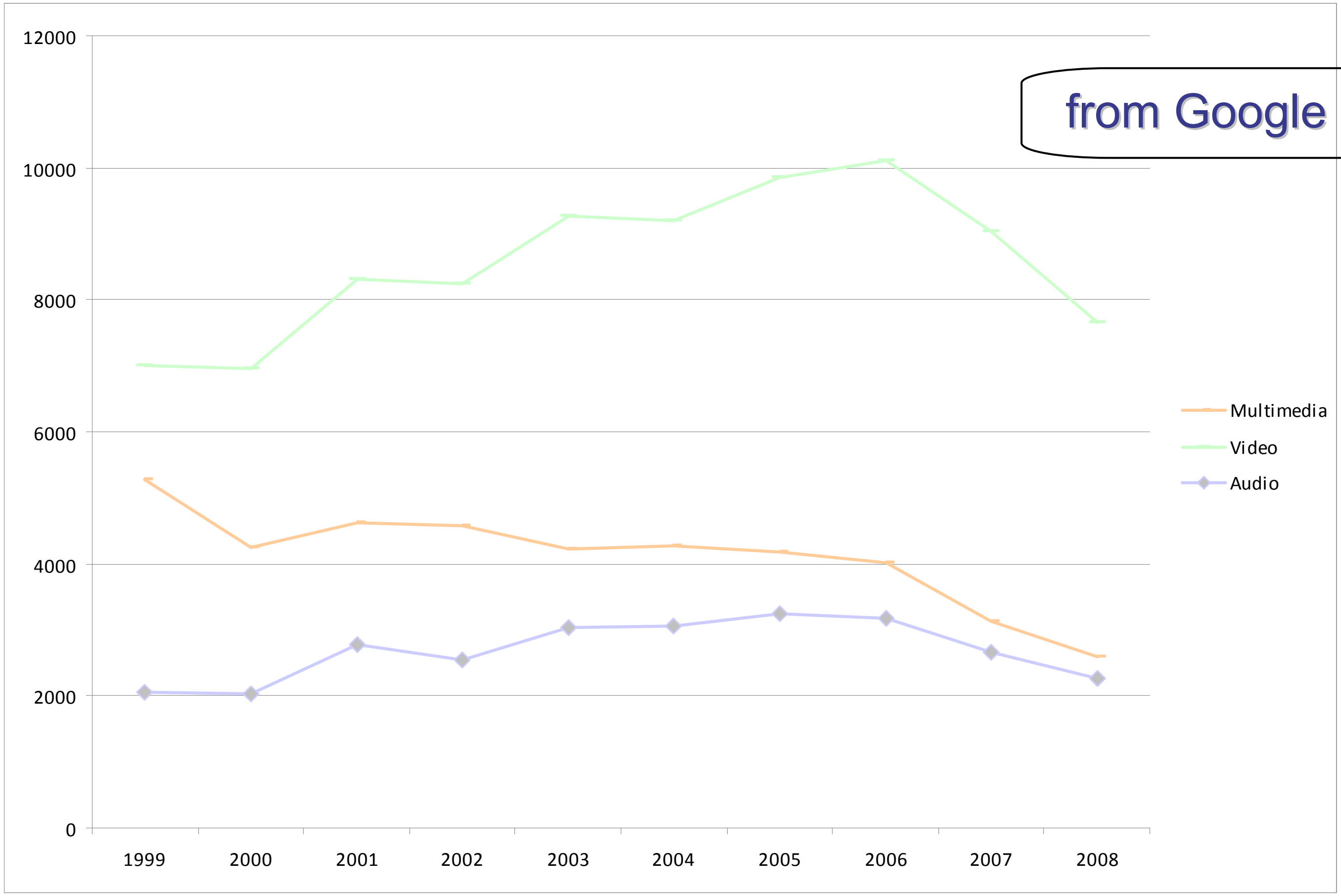
from Google Scholar

Number of research papers in MM (2)



from Google Scholar

Number of research papers in MM (3)



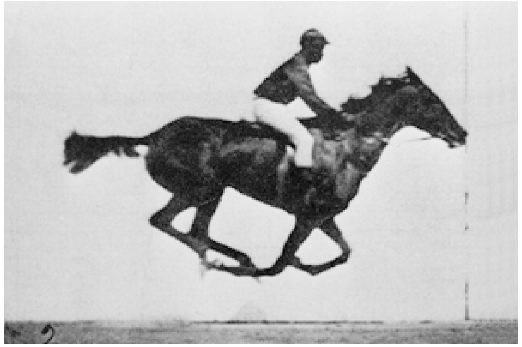
from Google Scholar

Is Research still relevant?

- Where are the questions (ideals) of (multimedia) research coming from?
  - ➢ Industry? Ivory tower? Society?
- Is pure academic research legitimate?
- Is pure industry oriented research legitimate?
- The canonical research paper ??
- Do we steadily reinvent the wheel?

Géricault

Muybridge



Are the issues of semantic gap still relevant?

- Pictures vs. matrix of pixels
- Automatic annotation – limitations (comp. vision)
- Manual annotation – laborious and error prone
- Caliph (Common And Light-Weight Photo Annotation)
  - Manual creation of semantic descriptors
- Emir (Experimental Metadata Based Image Retrieval)
  - Retrieval tool for photos annotated with Caliph
- How long does it take to annotate photos?
  - ~ 235 photos / day (15 – 1.7 minutes / photo)

Is broadcasting still relevant?

- Broadcasting?
  - ➢ Sharing + Authorized messages
- Classical broadcasting
  - ➢ Authority (Zeus) speaks to everybody
  - ➢ Multiple Gods (Pallas Athena, Aphrodite)
  - ➢ Personal interaction – (Odysseus, Moses)
- “French revolution” of broadcasting
  - ➢ Everybody may be broadcaster and receiver
  - ➢ Authority is replaced by popularity (must be “sold”)
  - ➢ Complex intertwining of data and metadata
  - ➢ Events, like “iron man”; motorway company: hundreds of cameras, thousands of sensors

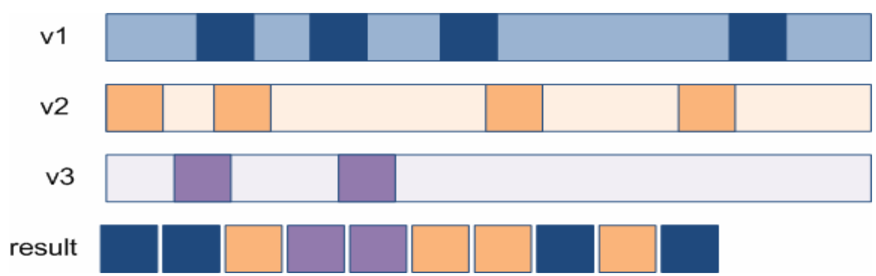
Are metada still relevant?

- Metadata? – Data about "other" data
  - Which are the data, which are the "others"?
  - True hypermedia required (bi-directional references)
- Two-Phase delivery – „Restaurant Model"
  - Offering Content (appetizer and menu card)
    - Preview of Content
  - Delivering Content
    - Preparation of „real" content delivery

- Rationales
  - It is better to show "something" than nothing
  - It is better to show images and movies than

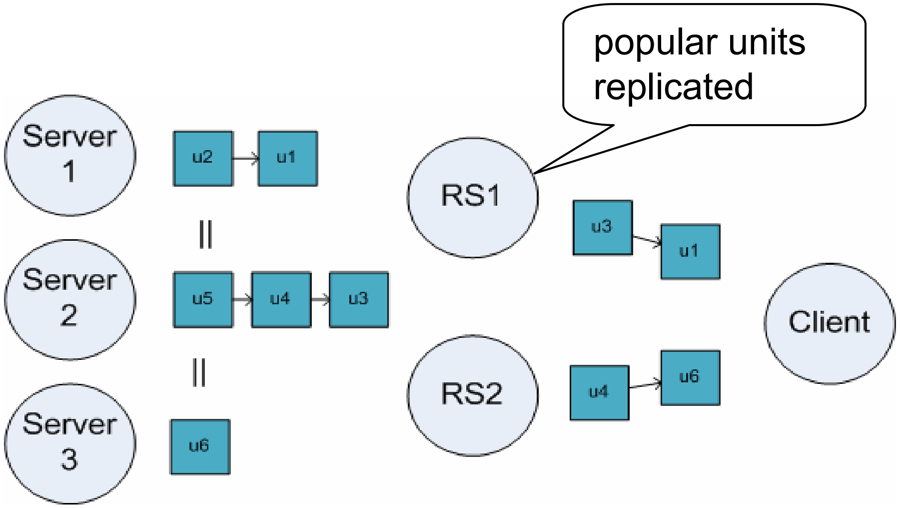
Is compression still relevant?

- Compression?
  - Remove (spatial/temporal) redundancy
  - Were H 2.64/AVC, SVC the last words?
  - What is with redundant information?
  - What is with irrelevant/boring information?
- Laws of Zip and Pareto
  - 20% of movies popular
  - 20% of scenes of these
  - 4% (or less) scenes
  - 2 OoM saving potential
  - Is this still compression? (we need additional info)



Is streaming still relevant?

- Streaming?
  - Reduce startup delay + enhance sharing
  - Enforces sequential access (long spaghetti in time)
- Non-sequential media
  - Composition/ Decomposition of units
  - QoS constrains
  - Parallel/Sequential access



  - Is this still streaming? (we need additional info)

Is MM searching relevant? (1)

- I am looking for a scene in a movie
  - I know neither the movie nor the scene exactly
  - But a colleague told me that this is exactly for me
- Desirable
  - A video system that finds me, what I am looking for
- Probable
  1. The system offers me 100 hours movies that could fit – however they do not
  2. The system noticed what kind of movies I seem to like and do not stop recommend me movies

Is MM searching relevant? (2)

3. The system finds the required scene, then
   a) I have to wait 5 minutes to start
   b) After 1 minute I get blurred pictures for a while
   c) After further minutes the sound is breaking down etc.
4. The system finds the required scene, but
   a) I do not have the proper decoder
   b) My screen has the wrong resolution
   c) The system talks to me in Japanese
5. System finds the required scene, but police comes and takes me into jail because of watching movies illegally

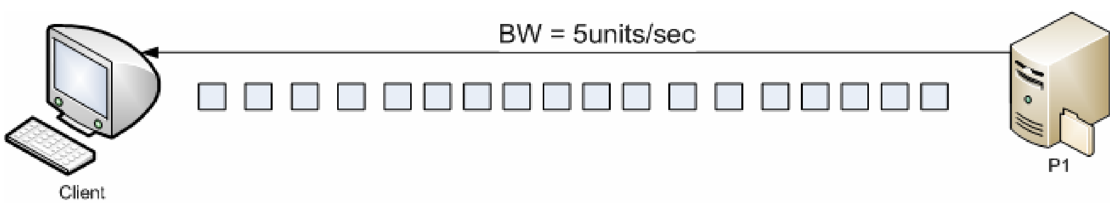
A few "interdisciplinary" answers

1. SOMA (Self-Organizing Multimedia Architecture), relies on cooperation among sensor - delivery - user
2. Video Explorer, relies on combination of browsing - retrieval - summarization
3. Proxy-to-Proxy (X2X), relies on cooperation of proxies - videos - client requests
4. MMC# and QoS-aware framework, combines language - adaptation - patterns

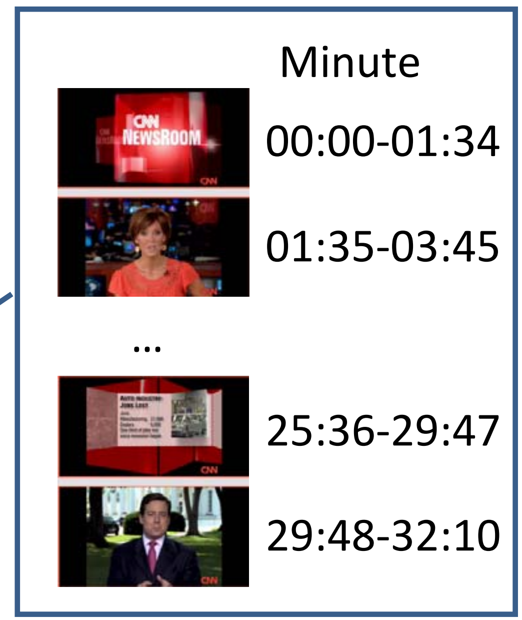
1.1. Video Calculus (1)

- Operators on *units*
  - Sequential: $u_1 \vdash u_2$
  - Sequential + QoS: $u_1 \leftarrow_Q u_2$
  - Parallel: $u_1 \,||\, u_2$
- Example: sequential download

$$\delta = \frac{|U|}{BW} = \frac{100}{5} = 20sec$$

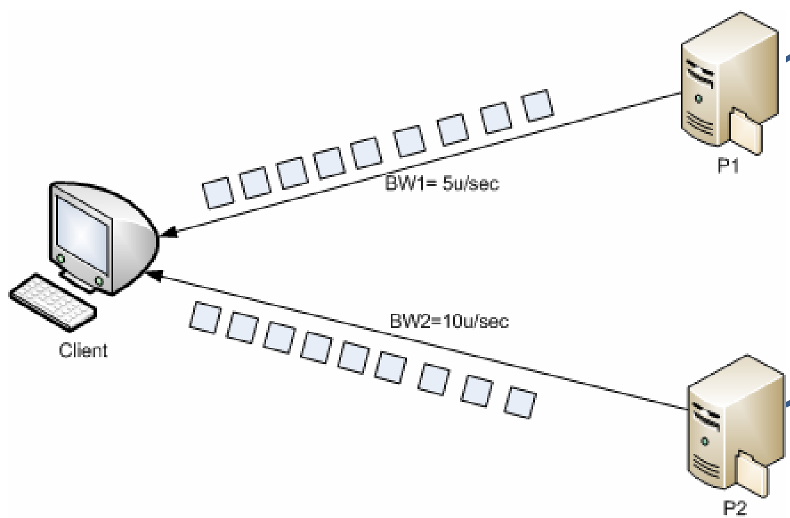


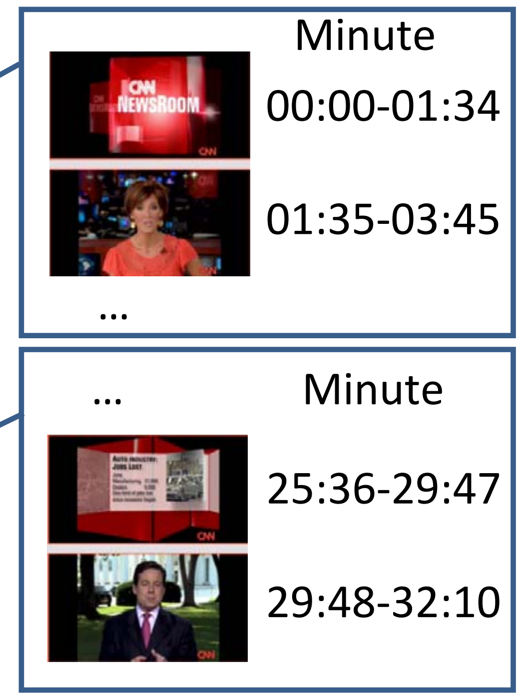
$$\left(u_1 \leftarrow u_2 \leftarrow u_3 \leftarrow \ldots \leftarrow u_{100}\right)_{BW}$$



1.2. Video Calculus (2)

- Parallel download





$$\delta = \max_{i=1}^{p}\left(\frac{|U_i|}{BW_i}\right) = \max\left(\frac{50}{10}, \frac{50}{5}\right) = 10\ sec$$

$$(u_1 \leftarrow u_2 \leftarrow u_3 \leftarrow \ldots \leftarrow u_{50})_{BW_1} || (u_{51} \leftarrow \ldots \leftarrow u_{100})_{BW_2}$$

1.3. Video Calculus (3)

- Streaming ~ Pipelining
  - Stages of the pipeline
    1. Download
    2. Playback

$$startup = \frac{|buffer|}{PB} = \frac{1}{5} = 0.2sec$$

$$\delta = startup + \sum_{i=0}^{p} \frac{|U|}{PB} = 0.2 + \frac{100}{5} = 20.2sec$$

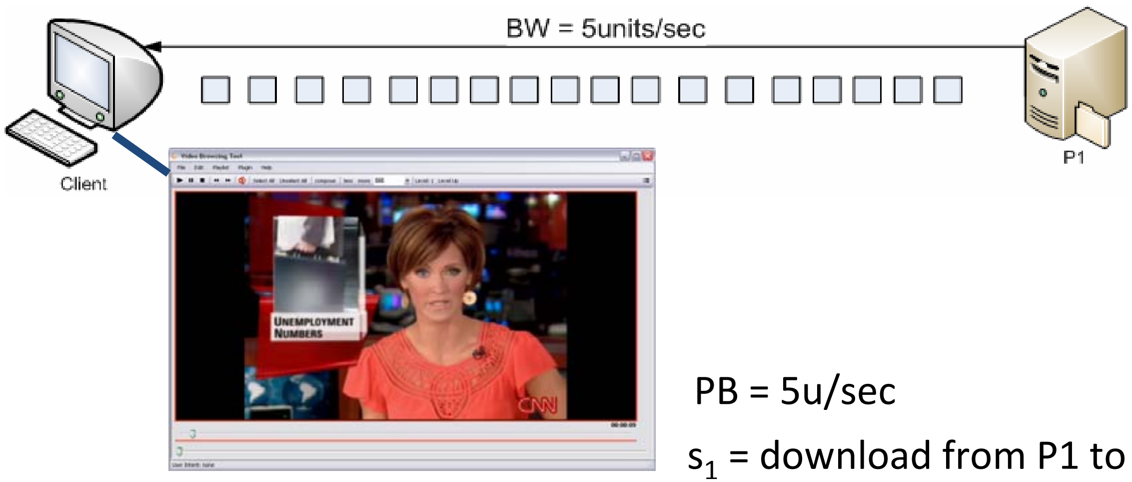
00:00-01:34

01:35-03:45

...

25:36-29:47

29:48-32:10



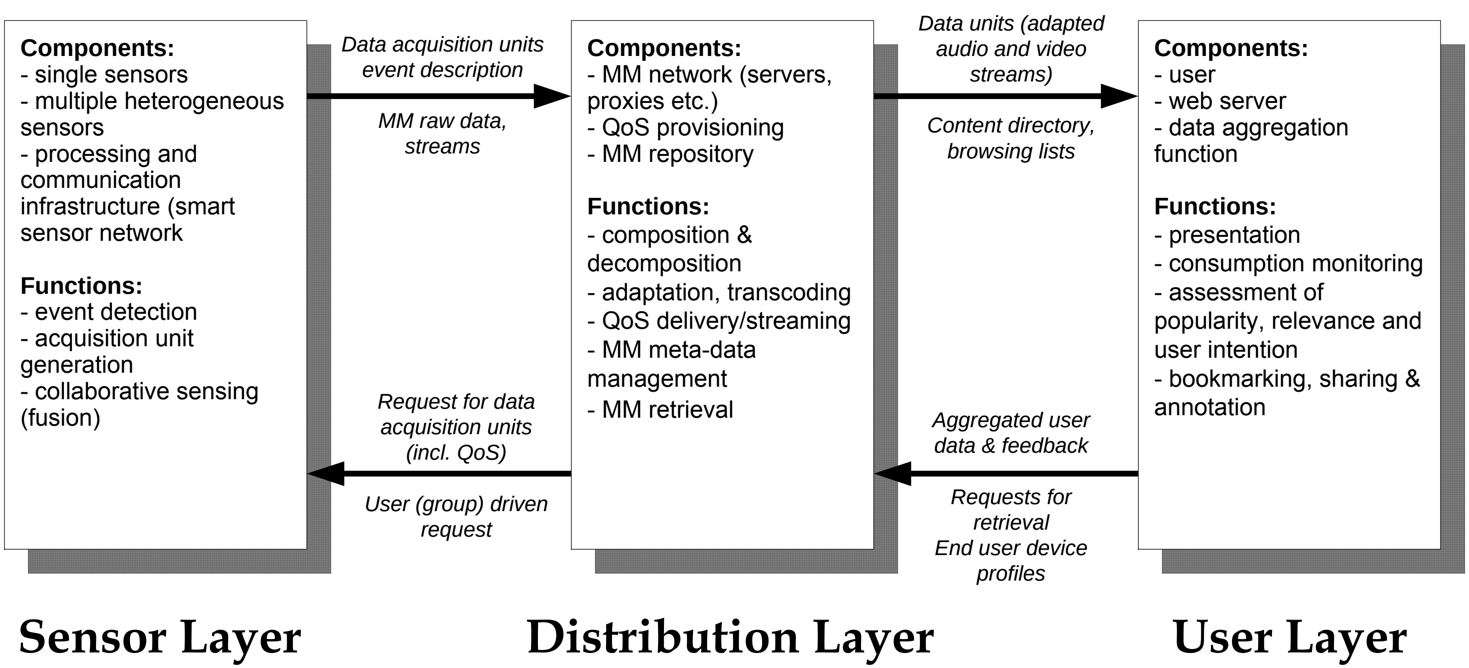
PB = 5u/sec

s_1 = download from P1 to Client, s_2= Playback on client

$$u_{1,s_1} \leftarrow_{BW,PB} (u_{2,s_1} || u_{1,s_2}) \leftarrow_{BW,PB} \cdots \leftarrow_{BW,PB} (u_{100,s_1} || u_{99,s_2}) \leftarrow_{PB} u_{100,s_2}$$

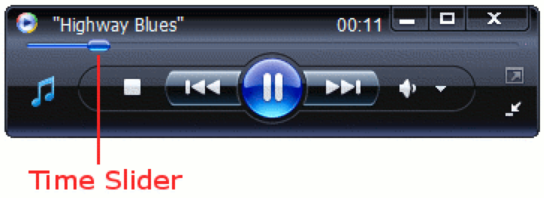
1.4. Self-Organizing Multimedia Arch.

- User intention → Composition → Deliver content
- Events (decomposition) → Offer content → ↑



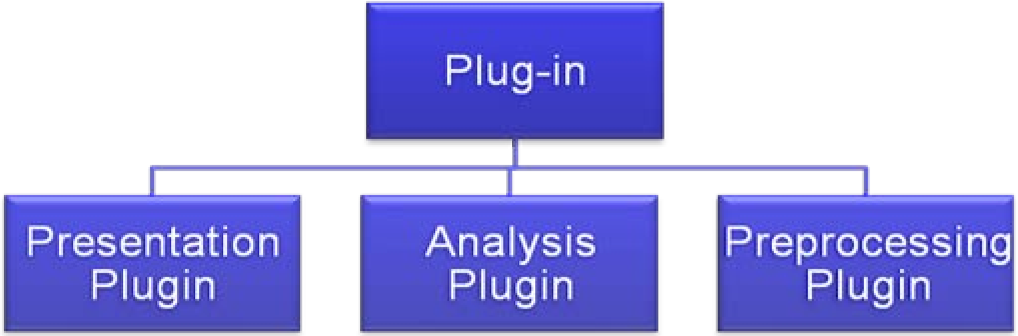
2.1. Video Exploration – Motivation

- Find certain video segments as fast as possible
- Identify scenes quickly we are *not* interested in
- Common video player are easy to use but restricted



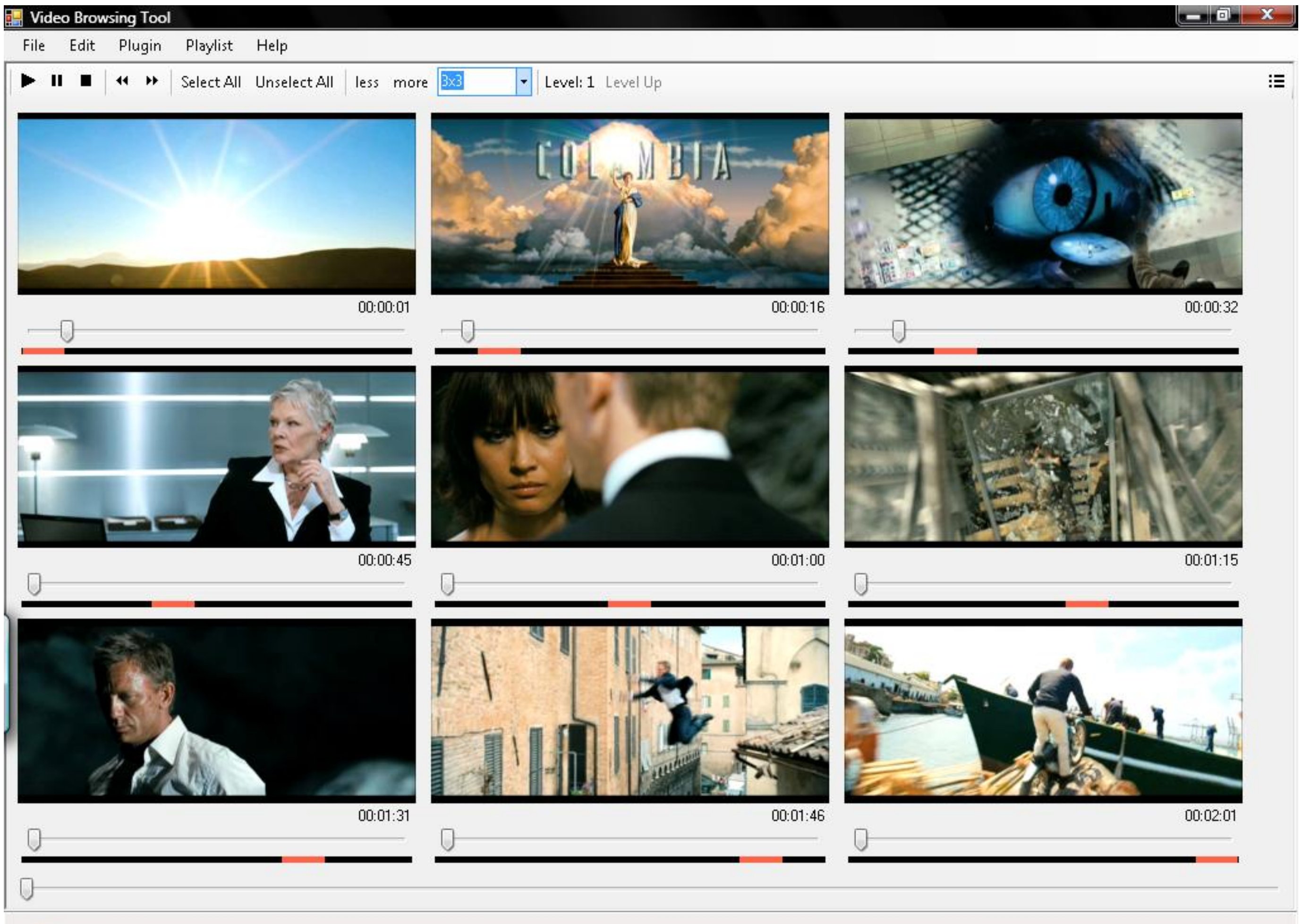
Navigation features from the VCRs of the 1960s!

2.2. Flexible Browsing – Architecture



- Plug-in architecture
  - Preprocessing
    - E.g. transcoding
  - Presentation
    - E.g. parallel, hierarchical
    - Tree-like, hierarchical
  - Analysis
    - Avoid lengthy analysis
    - Mainly compressed domain (H 2.64)

2.3. Flexible Browsing – Example

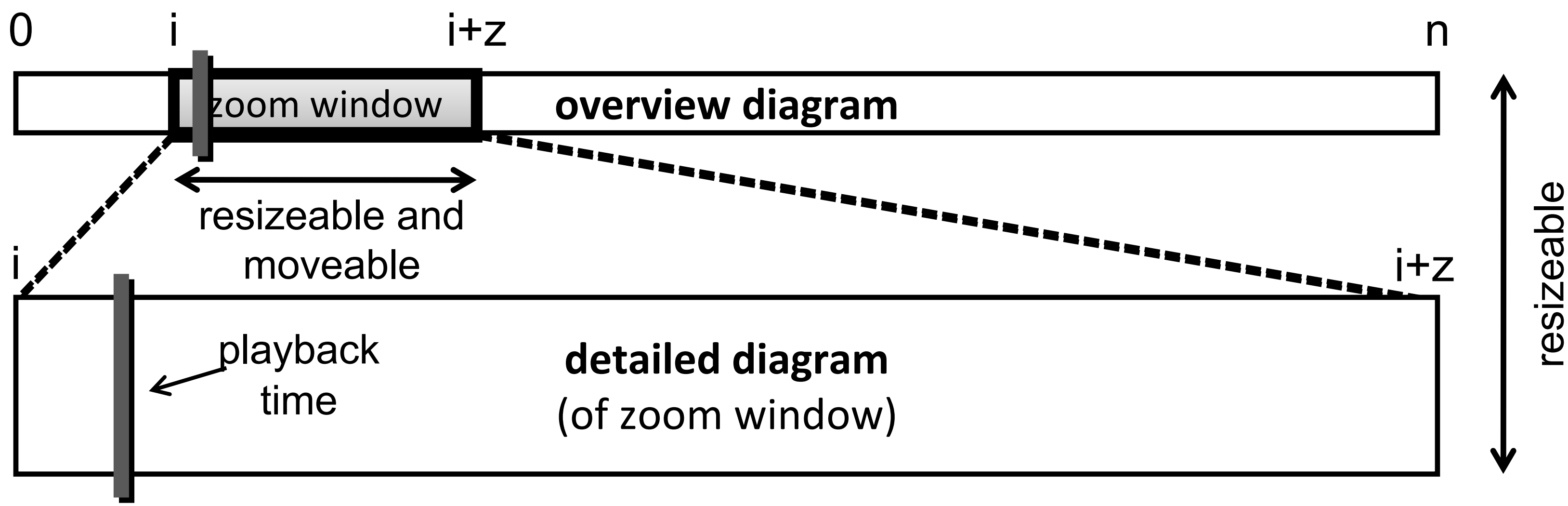


2.4. Video Exploration – Main goals

- Keep well-known user interactions, incl. time slide
  - Previous tests showed the users are "conservative"
- Flexible interaction via
  - Interactive Navigation Summaries
  - ROI and SOI search
  - Simple query formulation
- Learn semantics from users
- Prototype implementation
  - Evaluation, incl. user tests, incl. experts

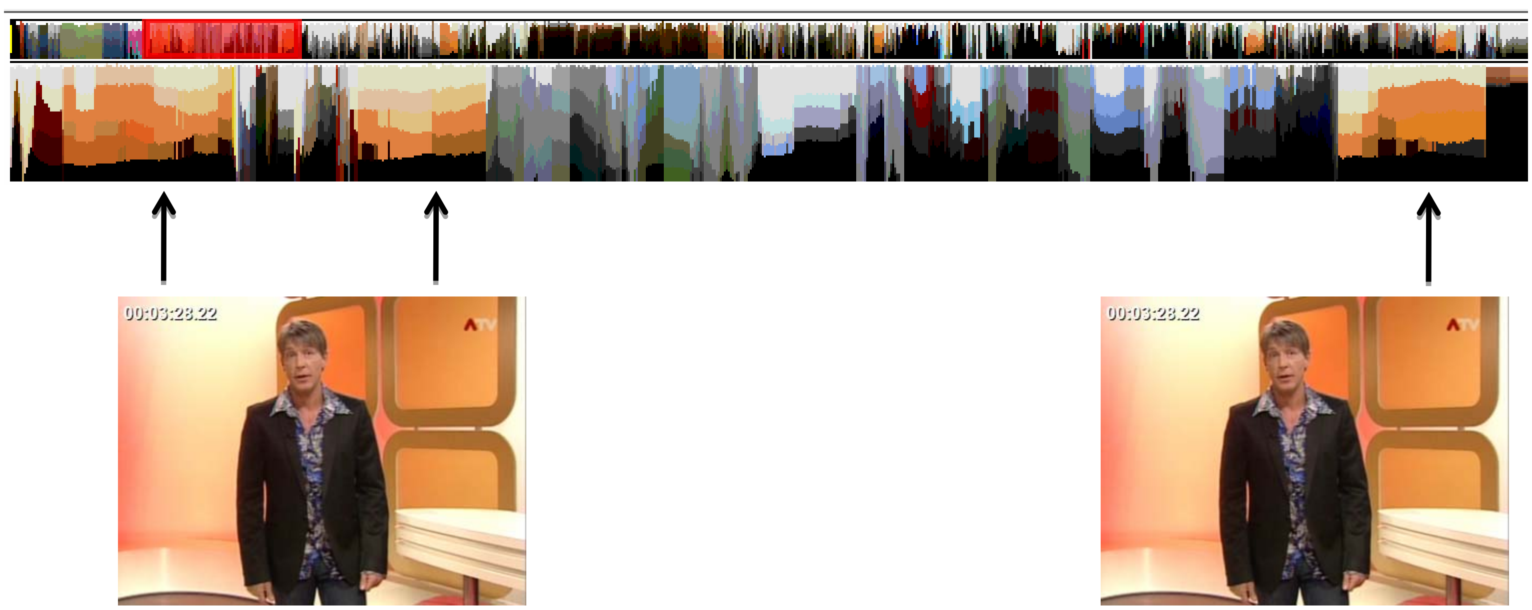
2.5. Interactive Navigation Summaries (1)

- “Intelligent time-sliders” – well known concept



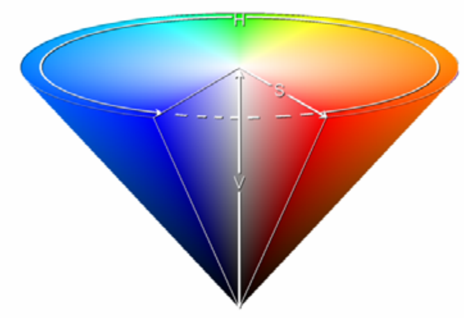
2.6. Interactive Navigation Summaries (2)

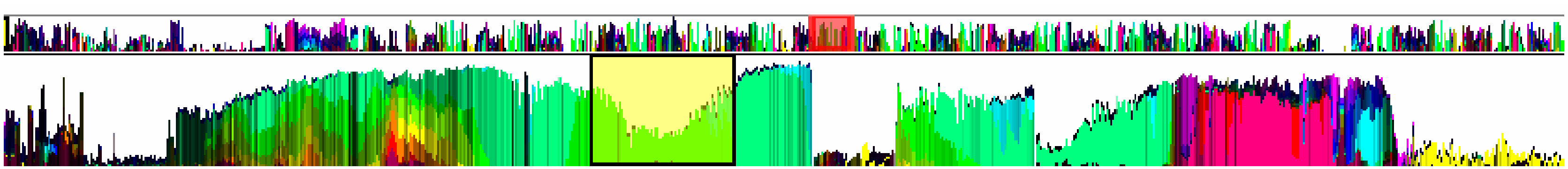
- k dominant colors per frame in a “river-like” layout
  - Vertical bars constructed from k (e.g. 5) values of 64 bins
  - Presentation order is constant (dark at the bottom etc.)

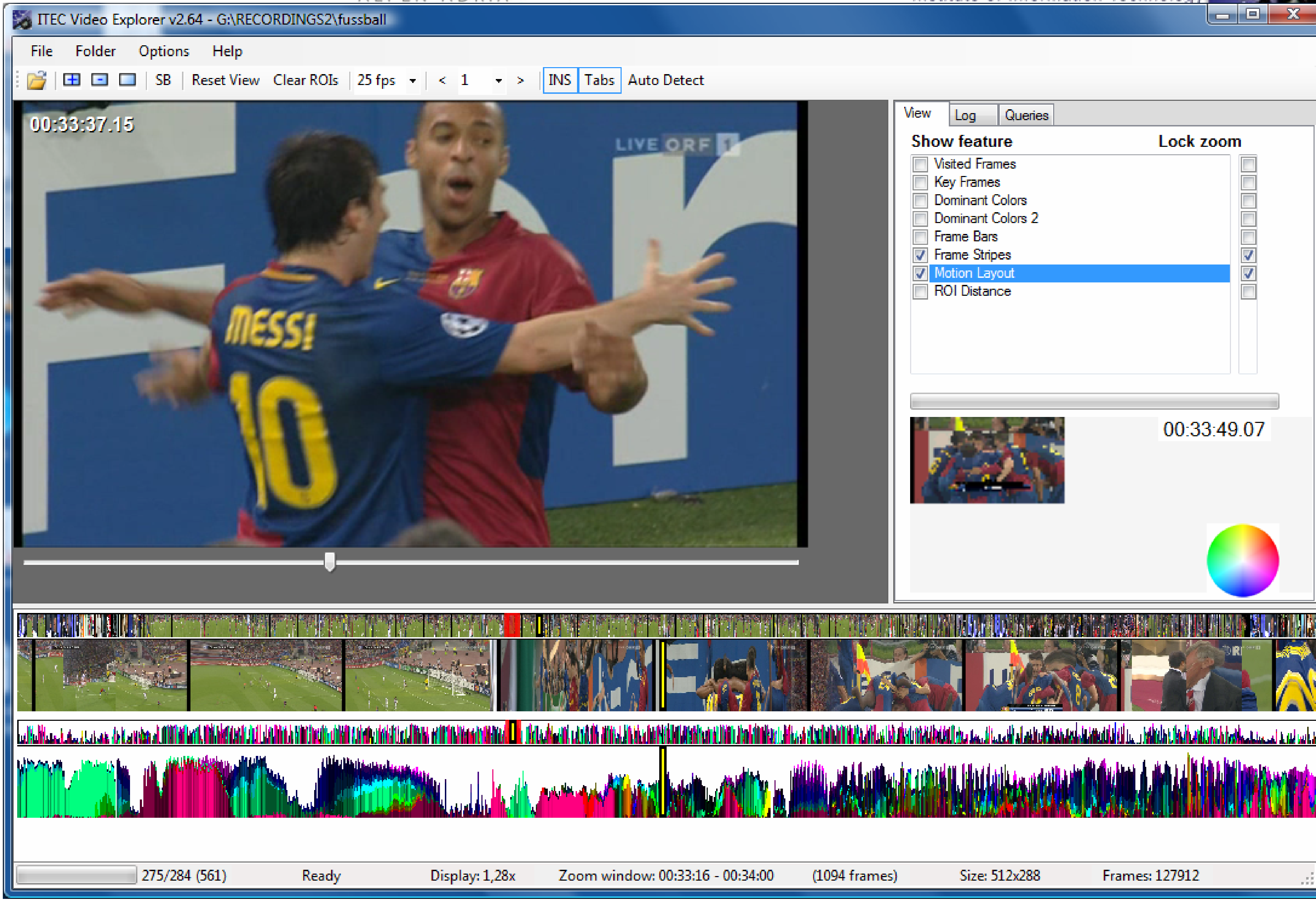


2.7. Interactive Navigation Summaries (3)

- Frame stripes: one pixel (under-sampled) column from each frame
- Motion layout, mapped to the HSV (hue/saturation/value) color space
  - E.g. light blue: fast movement to the left





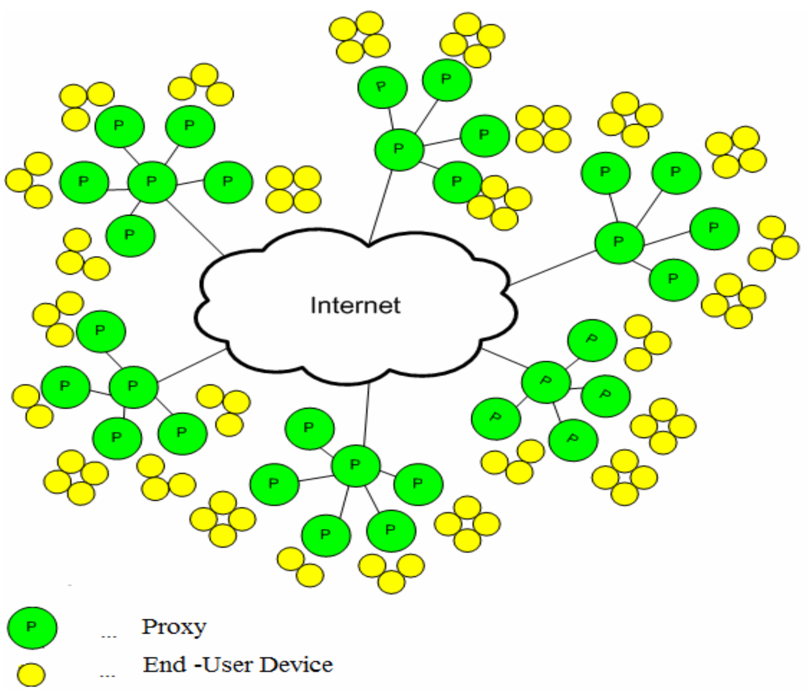

ITEC Video Explorer v2.64 - G:\RECORDINGS2\fussball
File
Folder
Options
Help
SB
Reset View
Clear ROIs
25 fps
< 1 >
INS
Tabs
Auto Detect
00:33:37.15
LIVE ORF 1
MESSI
10
View
Log
Queries
Show feature
Lock zoom
Visited Frames
Key Frames
Dominant Colors
Dominant Colors 2
Frame Bars
Frame Stripes
Motion Layout
ROI Distance
00:33:49.07
275/284 (561)
Ready
Display: 1,28x
Zoom window: 00:33:16 - 00:34:00
(1094 frames)
Size: 512x288
Frames: 127912


ITEC Video Explorer v2.64 - G:\RECORDINGS2\fussball
File
Folder
Options
Help
SB
Reset View
Clear ROIs
25 fps
< 1 >
INS
Tabs
Auto Detect
00:22:57.05
LIVE ORF 1
View
Log
Queries
Player zoom
00:22:54
00:22:56
00:23:07
00:23:09
00:23:33
00:26:45
00:29:13
00:29:15
00:29:46
00:31:02
00:32:48
00:32:50
00:32:52
00:33:16
00:36:53
00:36:55
00:41:19
00:41:21
00:41:31
00:42:16
528/166 (353)
Ready
Display: 0,97x
Zoom window: 00:22:37 - 00:23:12
(881 frames)

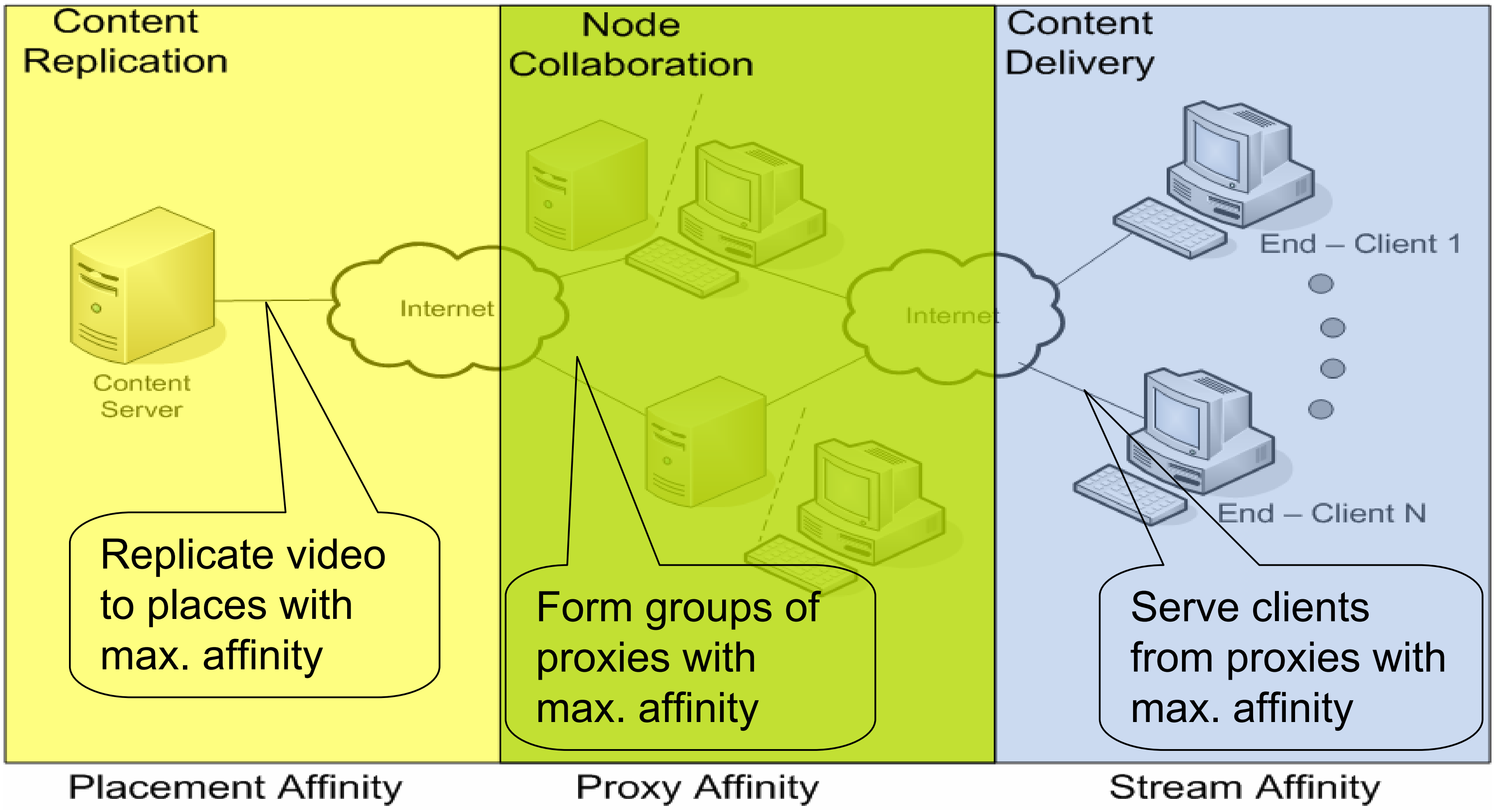
3.1. Proxy-to-Proxy (X2X) Video Delivery

- 3 main components
  1. Videos
  2. Proxies
  3. Client requests
- Components have own affinity function
  - Strive for maximal affinity
  - A global optimum should be reached
  - Self-organizing arch.

Sharing ⇔ Replication

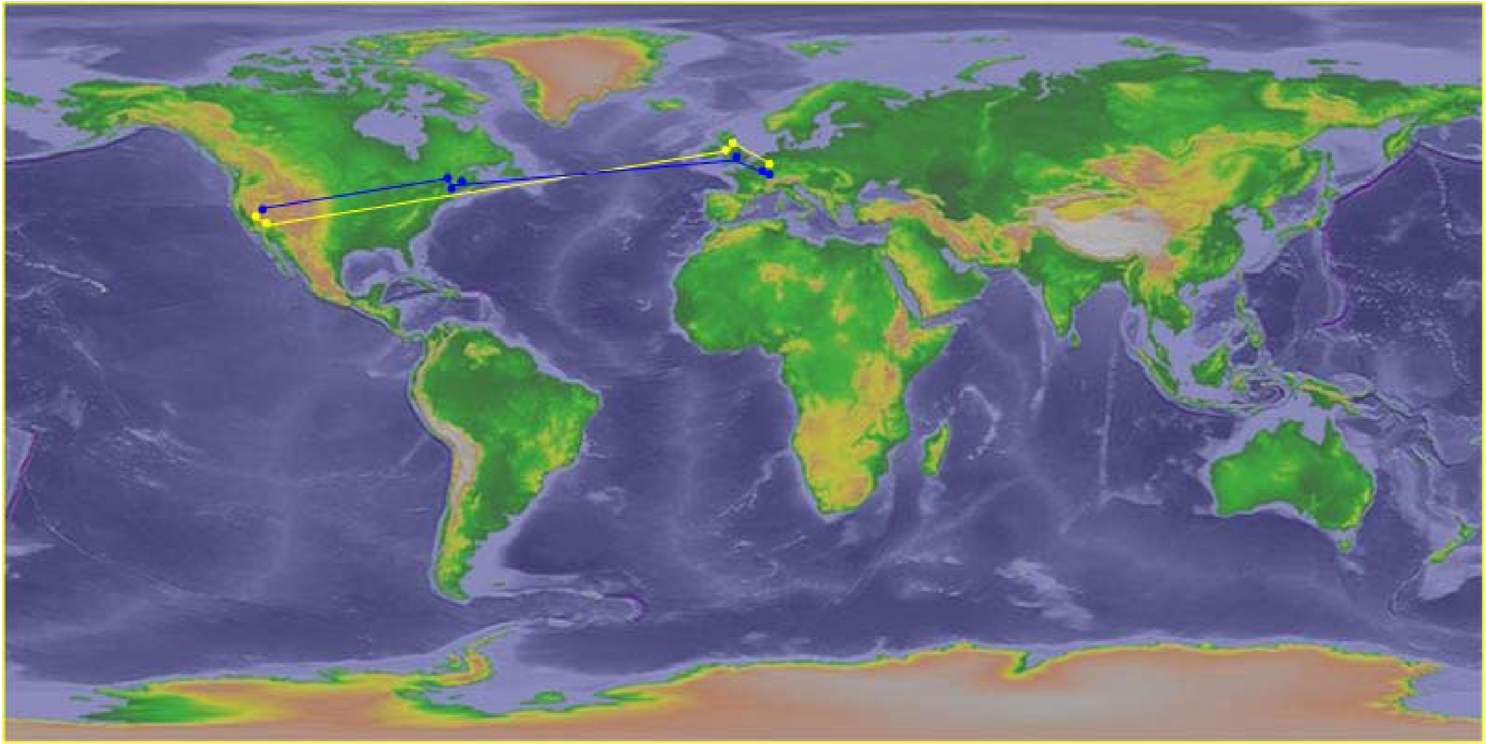


3.2. Overview of Affinity



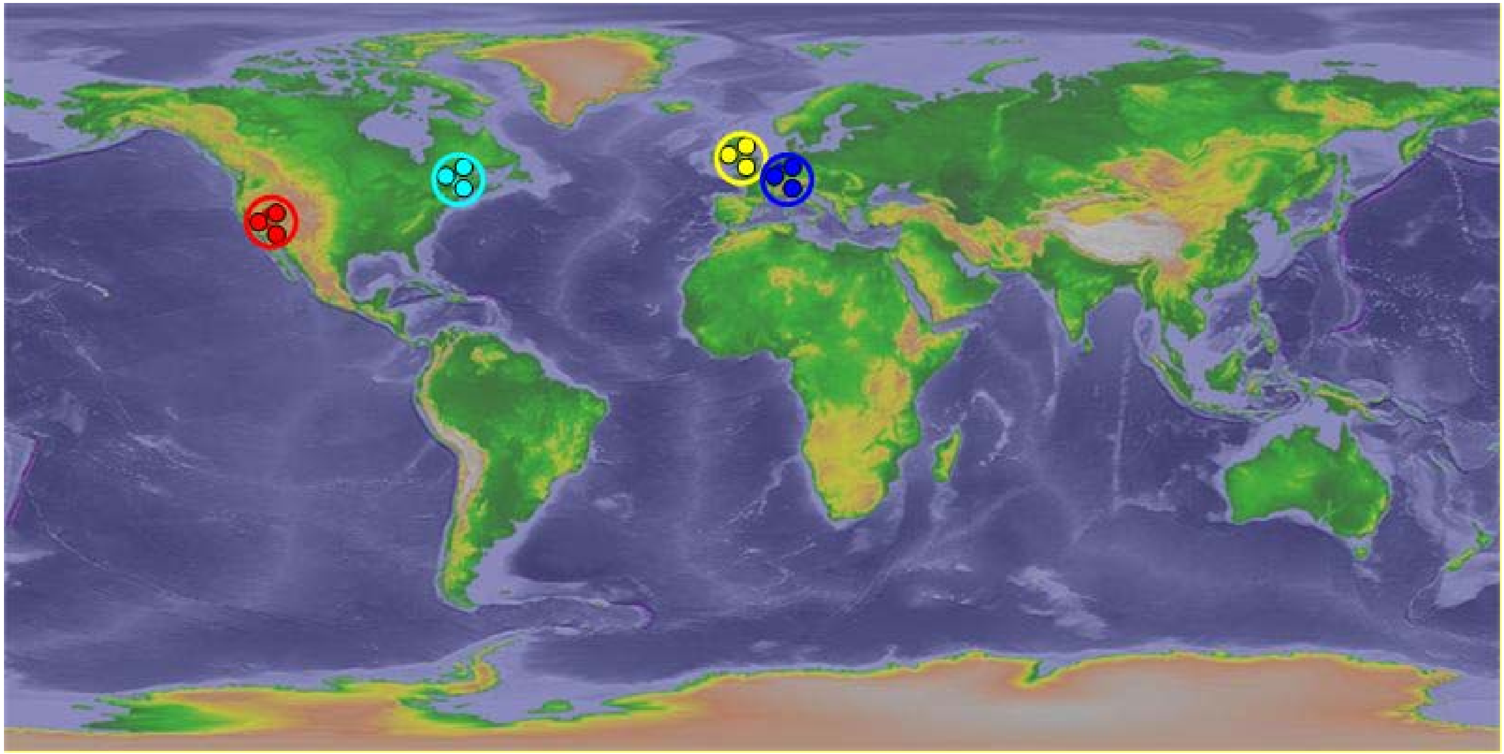
3.3. Group formation in PlanetLab (1)

- Based on semantic closeness (α=0), 2 genres (blue and yellow)
  - ProxyAffinity = α * NetworkCloseness + (1- α) * SemanticCloseness
  - Long distances, common interest



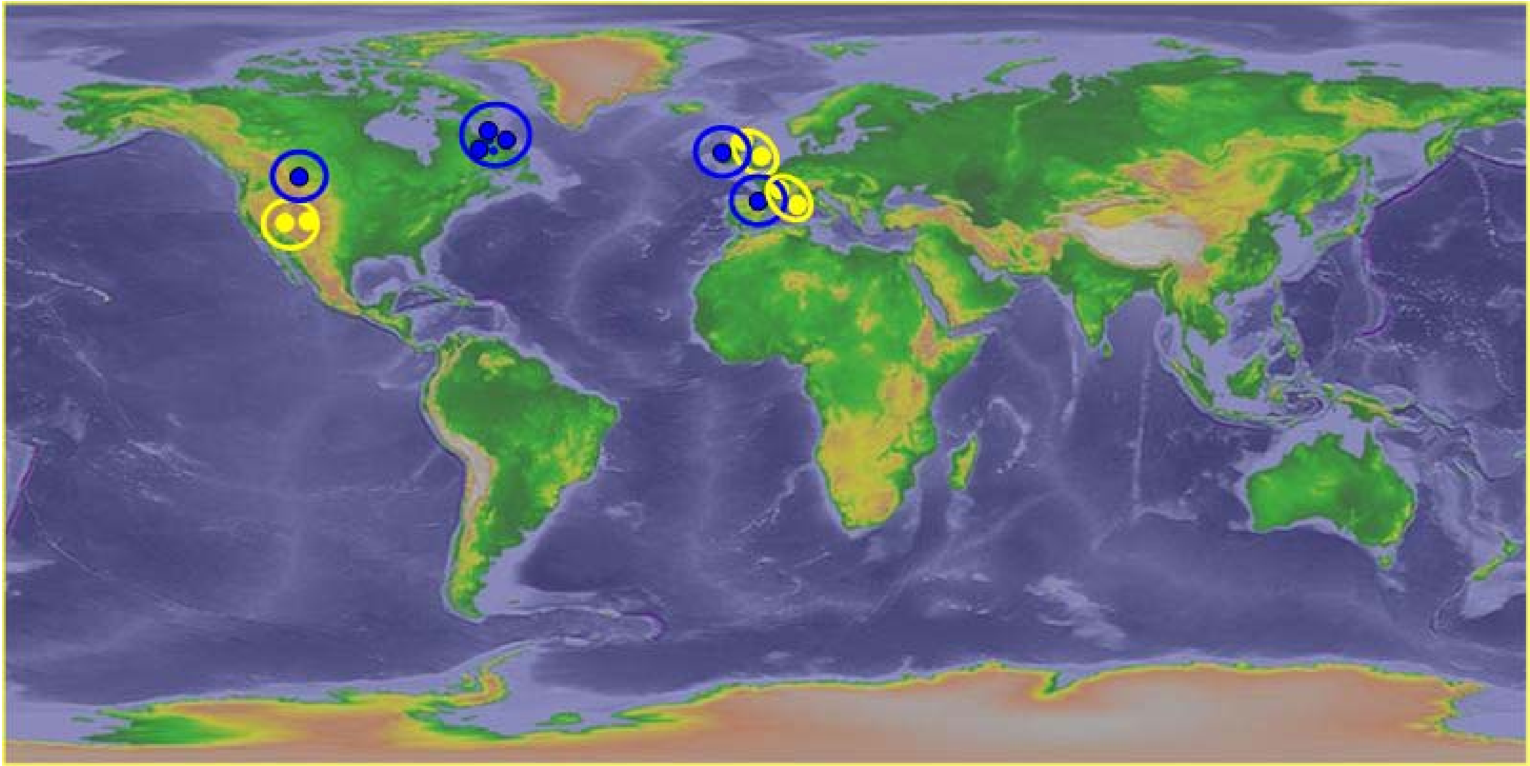
3.4. Group formation in PlanetLab (2)

- Based only on network closeness (α=1)
  - ProxyAffinity = α * NetworkCloseness + (1- α) * SemanticCloseness
  - Short distances; semantically meaningless (“stupid”) groups

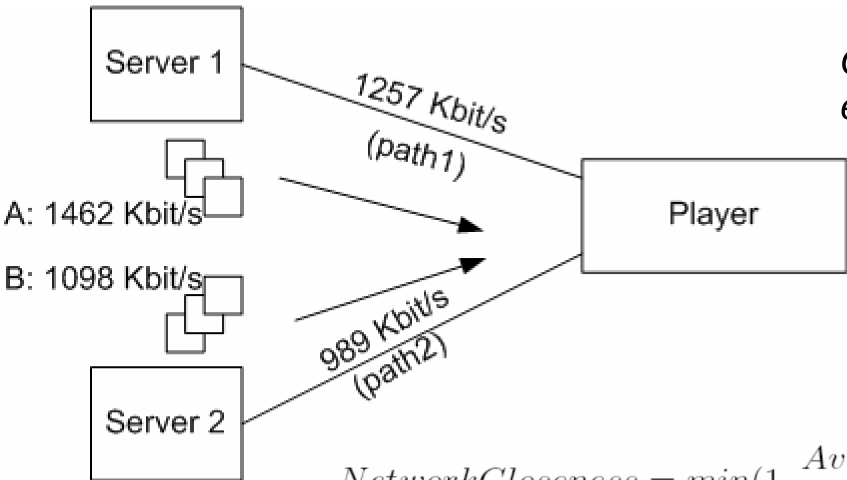


3.5. Group formation in PlanetLab (3)

- Based on combination of semantic and network closeness (α=0.5)
  - ProxyAffinity = α * NetworkCloseness + (1- α) * SemanticCloseness
  - Good compromise: common interest; fairly closed geopraphically



3.6. Stream affinity - example



$$StreamAffinity = \frac{1}{N} * QualityCloseness * \prod_{i=1}^{N} NetworkCloseness_i$$

Qual.Cl. considers dependencies; e.g. error propagation

Stream	Stream Affinity
A	0.66
B	0.5

$$NetworkCloseness = min(1, \frac{AvailableBandwidth}{Bitrate})$$

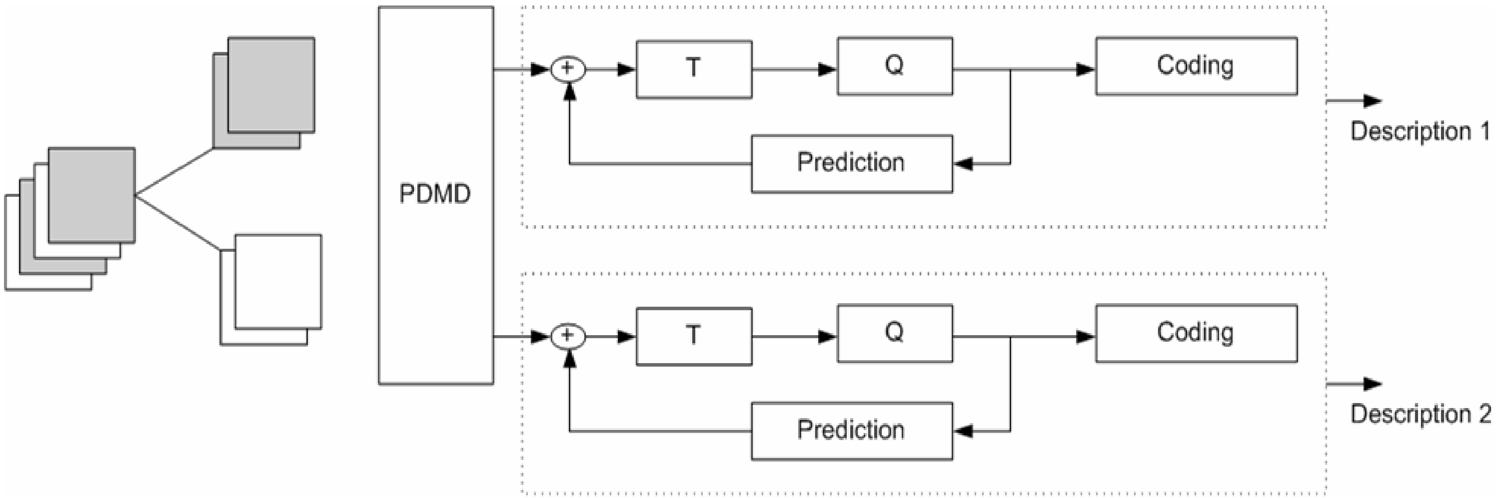
Stream	Bitrate	Avail. BW	NW - Closeness
A	1462	1257	0.86
B	1098	989	0.90

Evaluation

Decision	Mean Opinion Score
Server A	4,02
Server B	3,50

3.7. Multiple-source Streaming and MDC

- Multiple Description Coding (MDC)
  - In the temporal or in the spatial domain
  - Partial streams are independent (as opposed to SVC)



3.8. Demo – scalable streaming with MDC

Forman.mp4
Req: 1 Mbit/s
Avail: 1 Mbit/s

Forman.mp4
Req: 1 Mbit/s
Avail: 512 kbit/s
Using MDC (half)

Forman.mp4
Req: 1 Mbit/s
Avail: 512 kbit/s

Forman.mp4
Req: 1 Mbit/s
Avail: 2 x512 kbit/s
Using MDC (2*half)

4.1. Adaptive Video-Player in MMC#

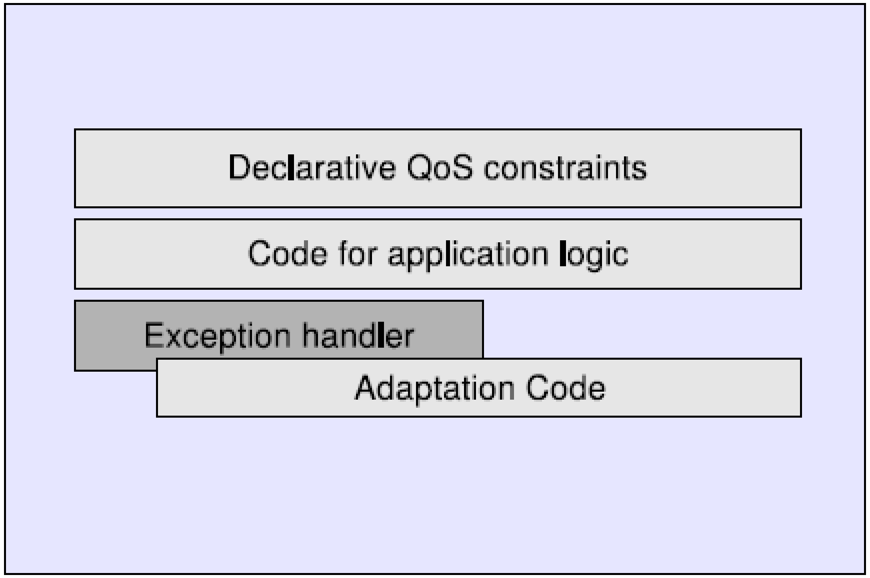
- Extending a general purpose language
  1. Enhance expressing power
  2. Enhance safety
  3. Enable optimizations

1. Automatic parallelism
   - Processing of independent units (e.g. slices) in parallel
   - Simple dependencies can be handled also easily
2. Automatic control of quality of service
   - Time as the n+1st dimension
   - Declarative QoS constraints with timed logic (e.g. fps)
   - Streaming operator controlled by QoS constraints
     - A kind of “timed assignment”

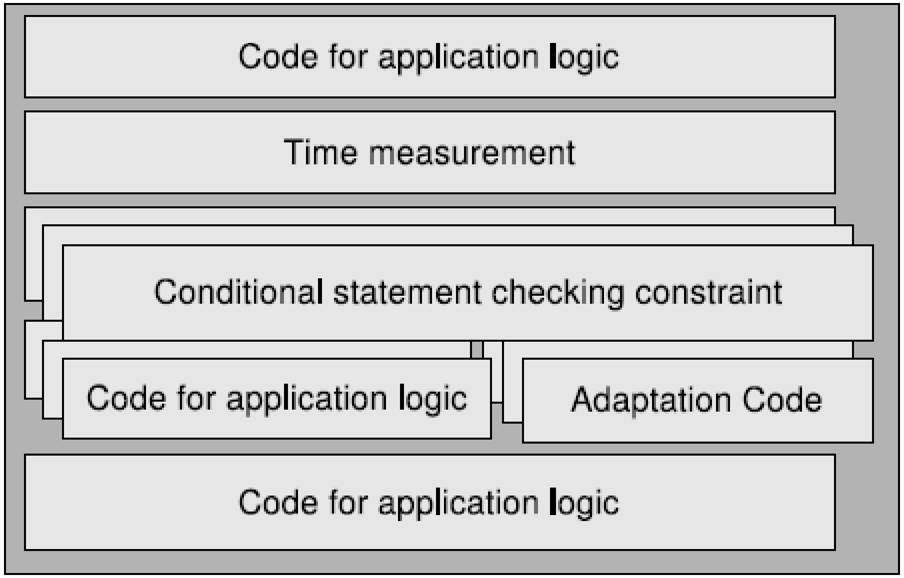
4.2. Adaptive Programming Model

1. Declare QoS constraints
2. Add code for „normal“ applications logic
3. Handle QoS violations as expressions

Adaptive Quality-Aware Programming Model



Common Model for Adaptive Programming



4.3. Timed Variables and Streaming

- The time dimension can be added to any type, as [~]
  - In type-compatibility checks time dimension is omitted

```
int [~] value1;       // denotes a time-sequence of integer values
int [][][~] value2;   // denotes a time-sequence of integer matrices
int [,][~] value3;    // denotes a time-sequence of packed matrices
Frame [~] f;          // denotes a time-sequence of Frames
```

- A value can be "streamed" – instead of assigned
  - Both in assignment statements and at passing of value parameters
  - Streaming is executed under QoS constraint (see next slides)

```
int[] arr = {1,2,3,4,5,6,7,9}; int [~] val;
   for (int i = 0; i < arr.Length; i++)  {…val ~: arr[i];…}
```

Streaming to timed target

  - The timed-variable may stay either as target or as source

```
int[] [~] arrr = {1,2,3,4,5,6,7,9}; int val;
   for (int i = 0; i < arr.Length; i++)  {…val :~ arr [i];...}  //stream from
   arr[i]
```

Streaming from timed source

4.4. Declarative QoS constraints

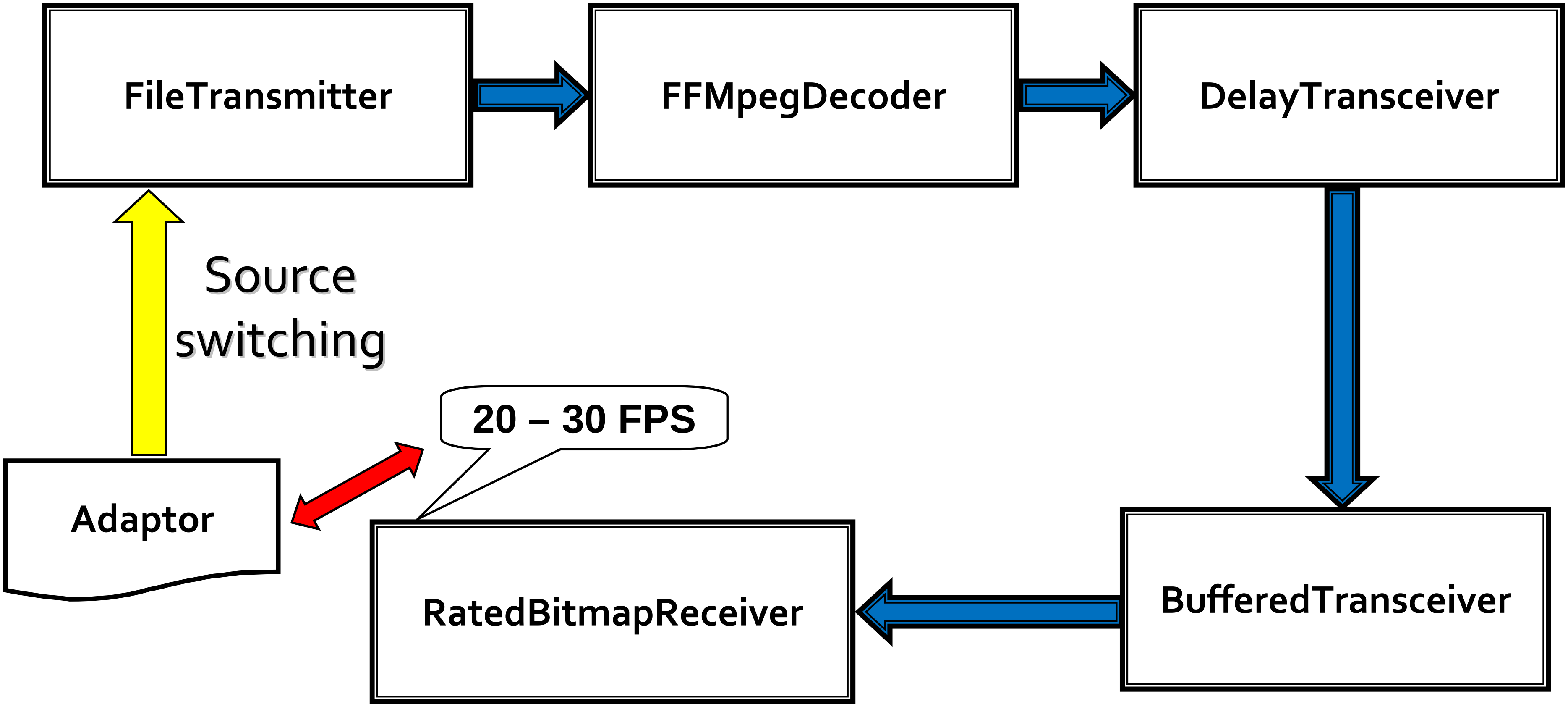
- Relies on the timed logic before

QoS Constraint	QL Syntax
Throughput	$\forall n, \lvert\tau(\epsilon_r, n+k) - \tau(\epsilon_r, n)\rvert \leqslant \delta$
Latency	$\forall n, \lvert\tau(\epsilon_r, n+k) - \tau(\epsilon_s, n)\rvert \leqslant \delta$
Delay Jitter	$\forall n, \delta_1 \leqslant \lvert\tau(\epsilon_r, n+k) - \tau(\epsilon_s, n)\rvert \leqslant \delta_2$
Bounded Execution Time	$\forall n, \lvert\tau(\epsilon_i, n) - \tau(\epsilon_o, n)\rvert \leqslant \delta$

- @ stays for ∀, [n] for the nth element in the history
- Examples (tVar, input and output are timed variables)

```
@n {output[n] - input[n] <= Units.MSec(50)}      // bounded time
// rate + jitter:
@n {tVar[n] - tVar[n-24] <= Units.MSec(1000)} &&
          Units.MSec(10) <= tVar[n] - tVar[n-1] <= Units.MSec(50)}
```

4.5. Component-chain of the adaptive player



4.6. QoS-aware rendering in adaptive player

```
private object [~] o;                                          // timed object
public void InitConstraint() {
    o = new object[~Constraint(o, parameters)]; // init constraint
}
// Example: (e[n]-e[n-1] >= 30) && (e[n]-e[n-1] <= 50) ms must hold

public void Display (stream object Data) { ...}     // input is constrained

public void ReceiveData(object inData) { // called with a timed object
    try {
        Display (inData);                          // may trigger QoS violation
    } catch (QoSException ex) {
        DoAdaption(ex);   // on QoS violation: call adaptation code
}   }
```

Conclusions

- Multimedia reached a critical point
- If we do further so, it remains “exotic”
- Interdisciplinary research is required
  - Streaming + self-organization
  - Searching + interaction
  - Compression + social aspects (popularity, intention)
  - Academia + industry
  - Science + art? →
  - …

Behold, a virgin shall conceive

- Isenheim Altarpiece
- Grünewald (1512 – 1516)
- Unterlinden Mus., Colmar
- Ecce virgo concipit et pariet

(Isiah, 7, 14)