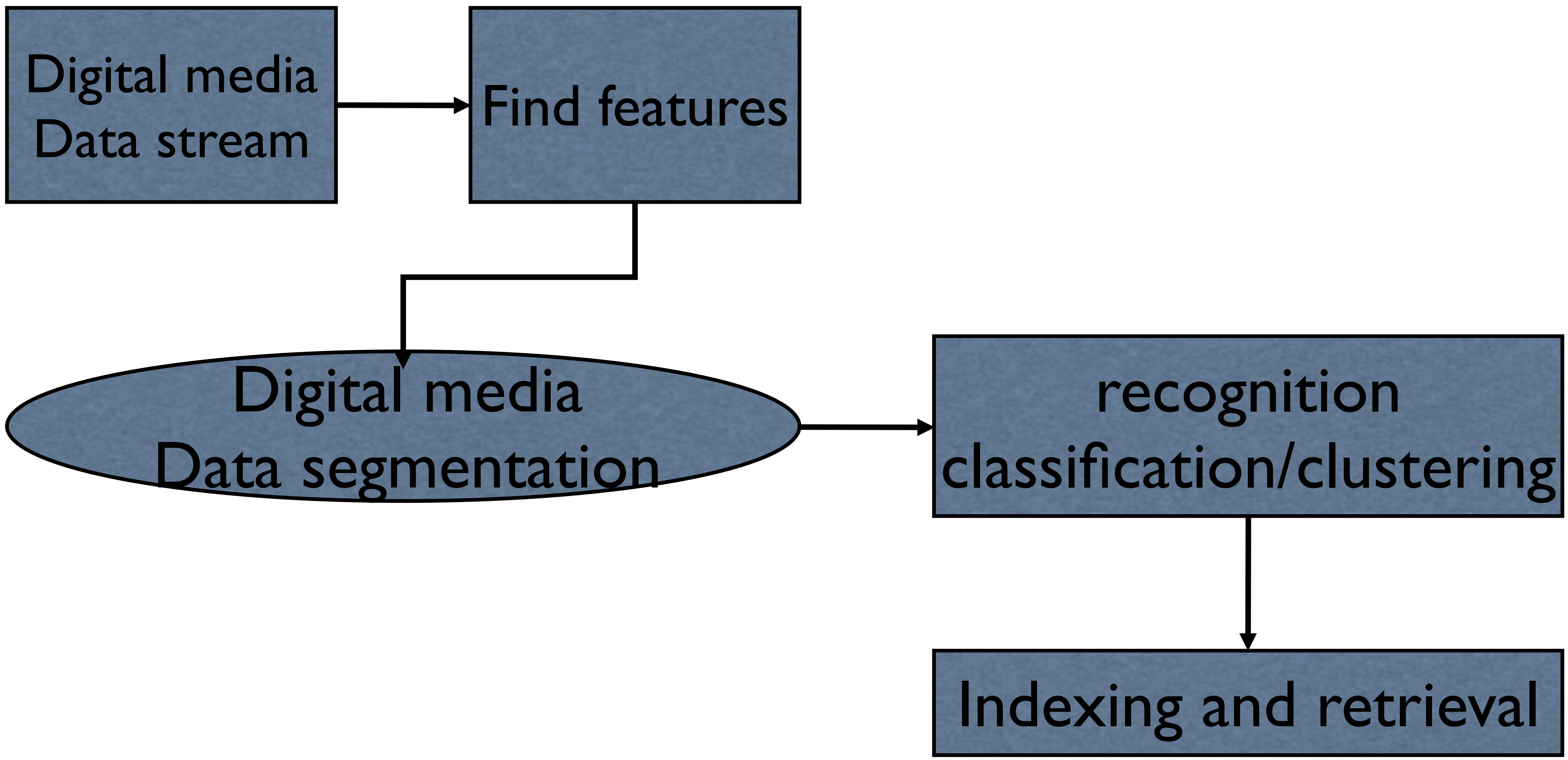
浙江大学计算机学院
数字媒体与网络技术

Digital Asset Management
数字媒体资源管理

6. Introduction to digital media retrieval

任课老师：张宏鑫
2008-10-08

The workflow of digital media analysis and retrieval



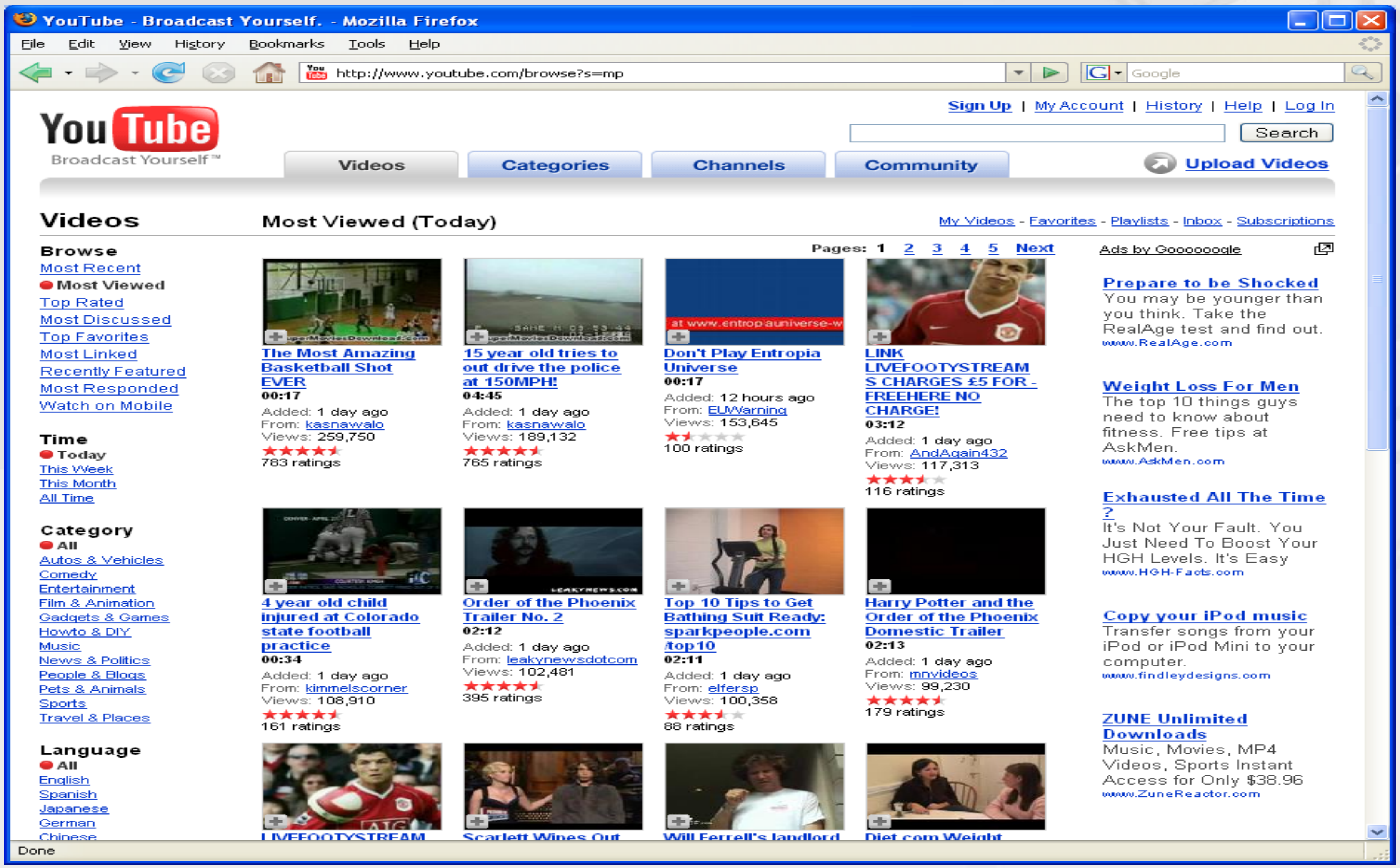
3. Video retrieval techniques

Differences and relations between image and video

- Images are **static**, but video are **dynamic**.
- Video stream can be viewed as sequence of image frames.

CBVR

- Sample YouTube Video page:



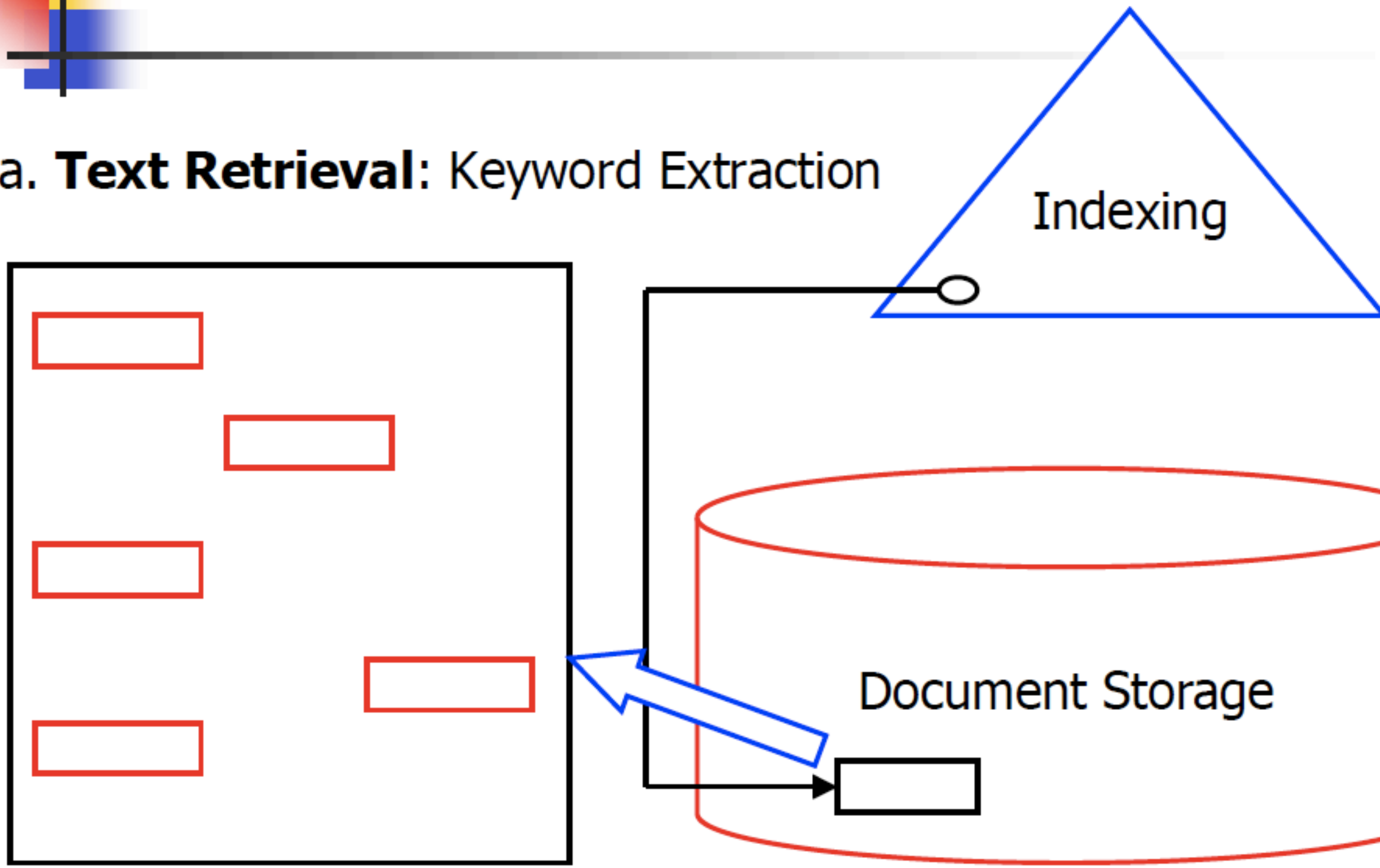
Main methods of digital media retrieval

- **Text-based** digital media retrieval

- **Content-based** digital media retrieval

Why we need video shots?

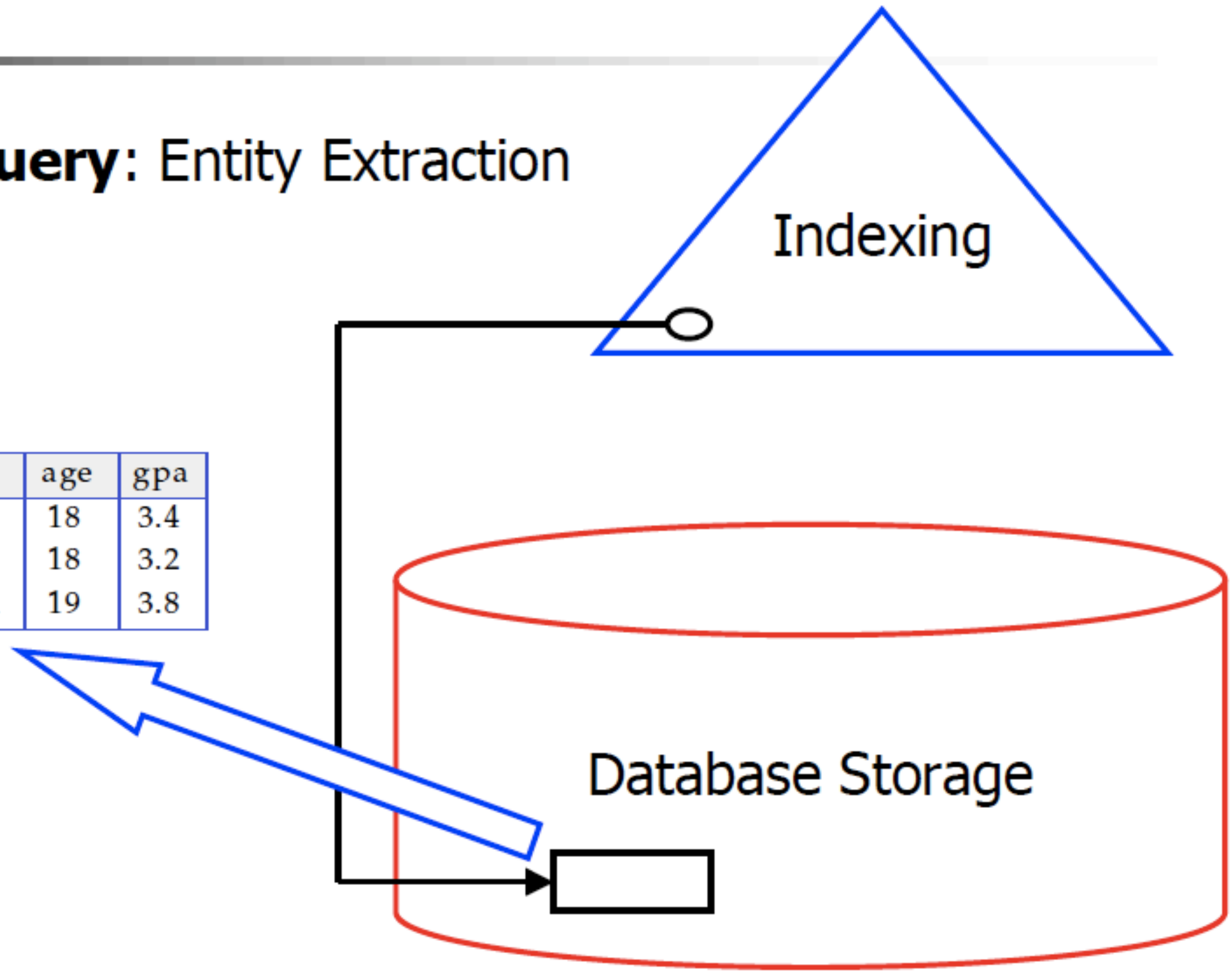
a. **Text Retrieval**: Keyword Extraction



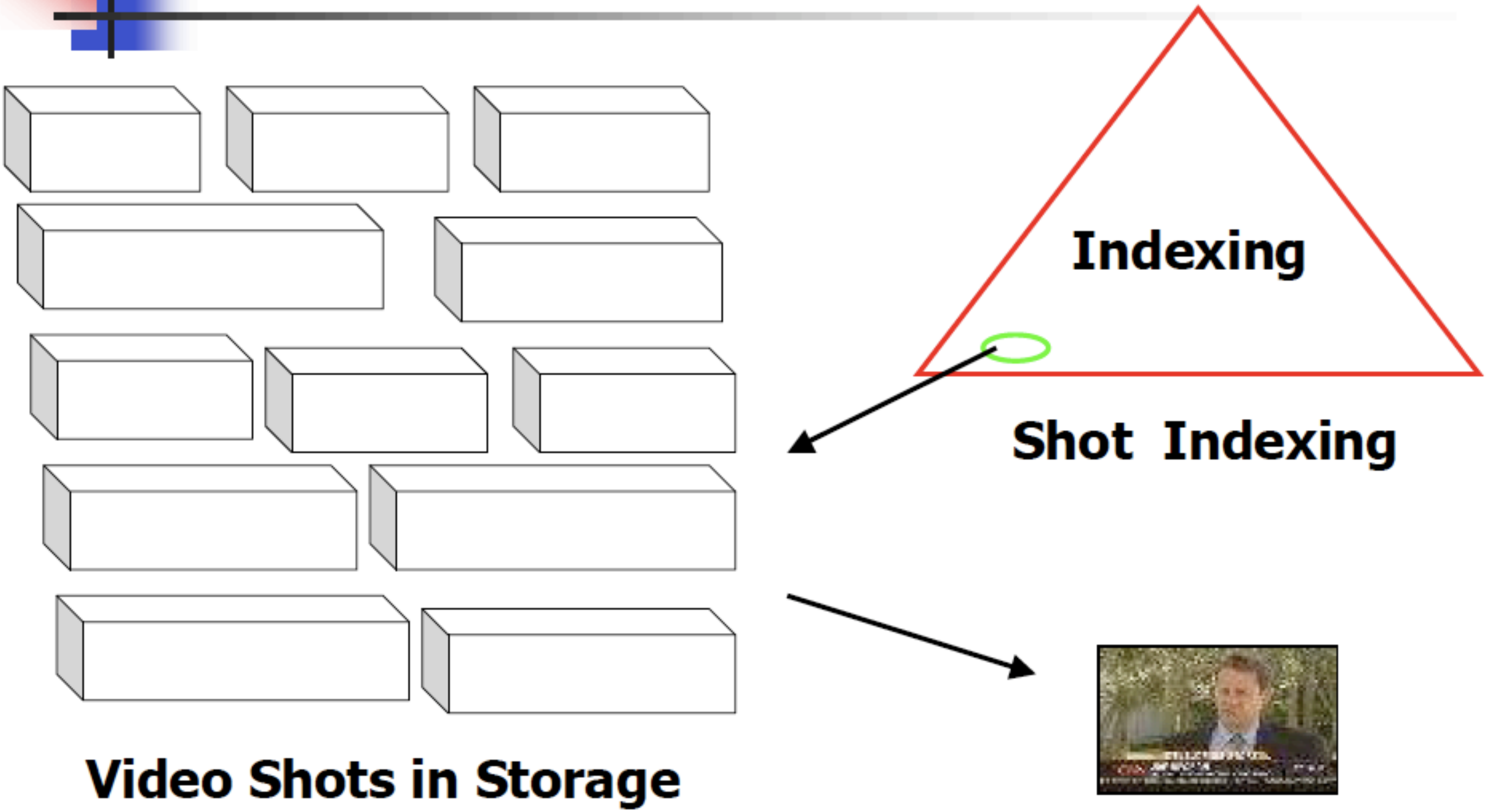
Why we need video shots?

b. **Database Query**: Entity Extraction

sid	name	login	age	gpa
53666	Jones	jones@cs	18	3.4
53688	Smith	smith@eecs	18	3.2
53650	Smith	smith@math	19	3.8

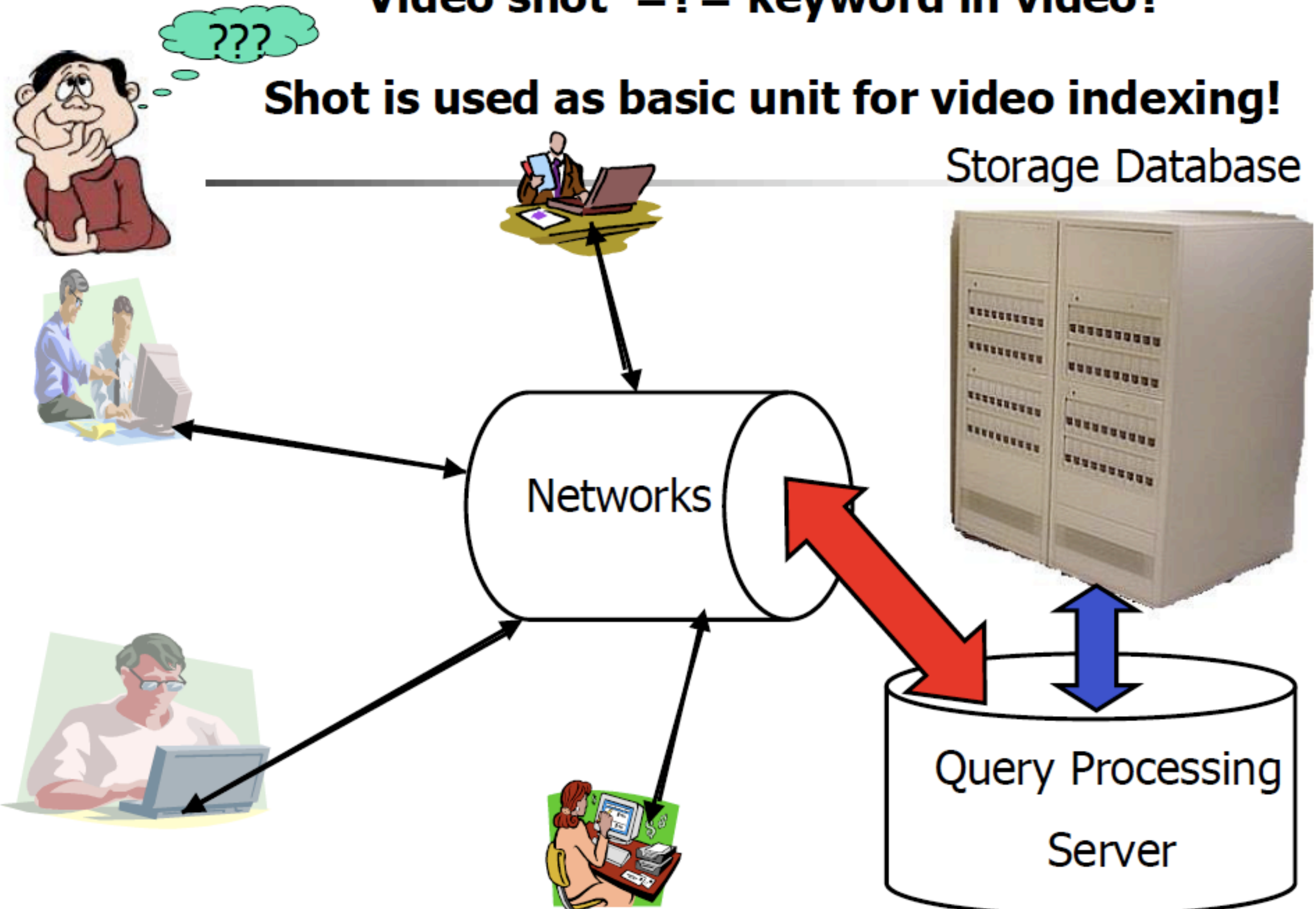


Why we need video shots?



Video shot =?= keyword in video?

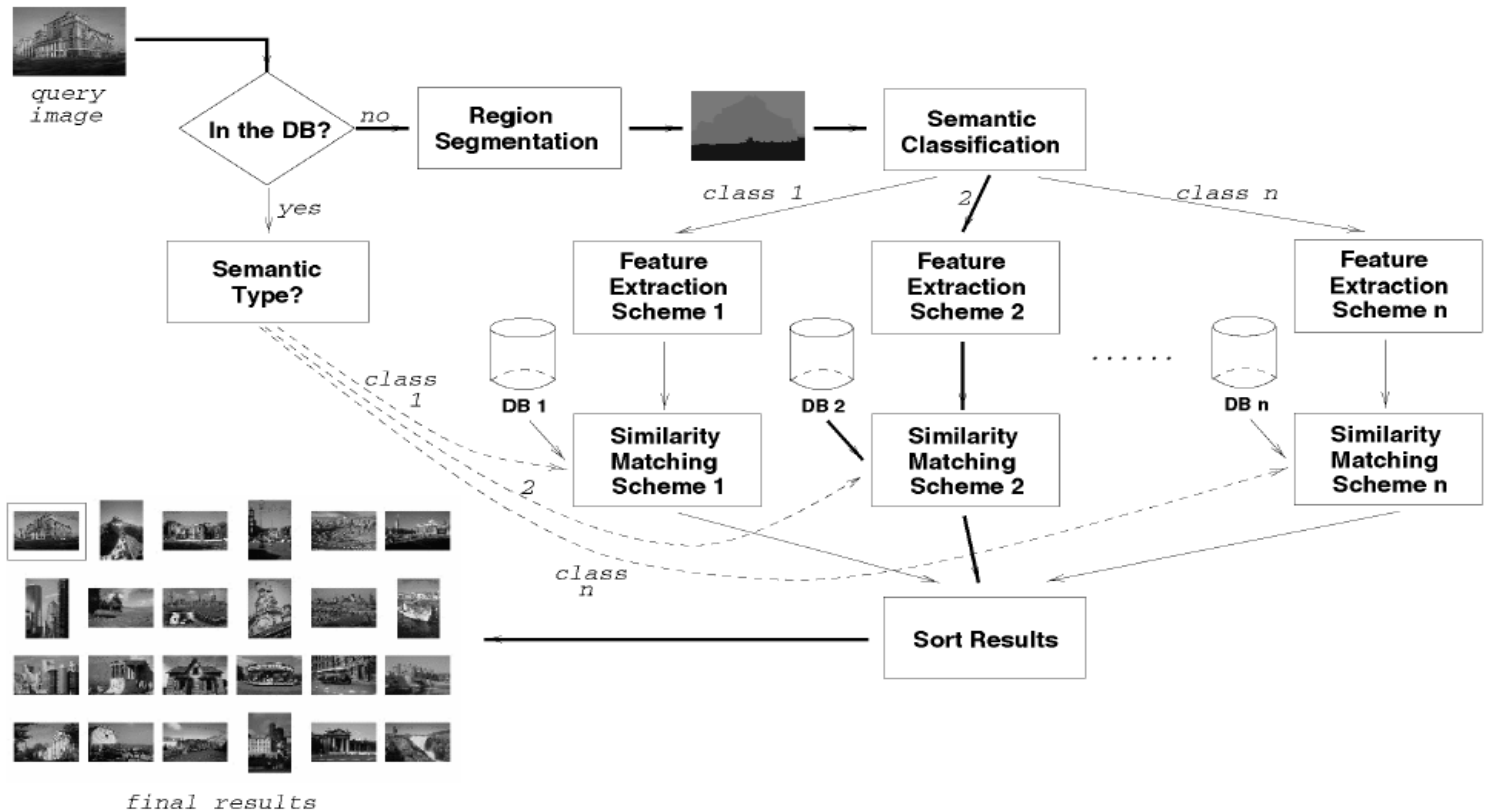
Shot is used as basic unit for video indexing!



Overview

- CBVR has two phases:
  - –Database Population phase
    - Video shot boundary detection
    - Key Frames selection
    - Feature extraction
  - –Video Retrieval phase
    - Similarity measure

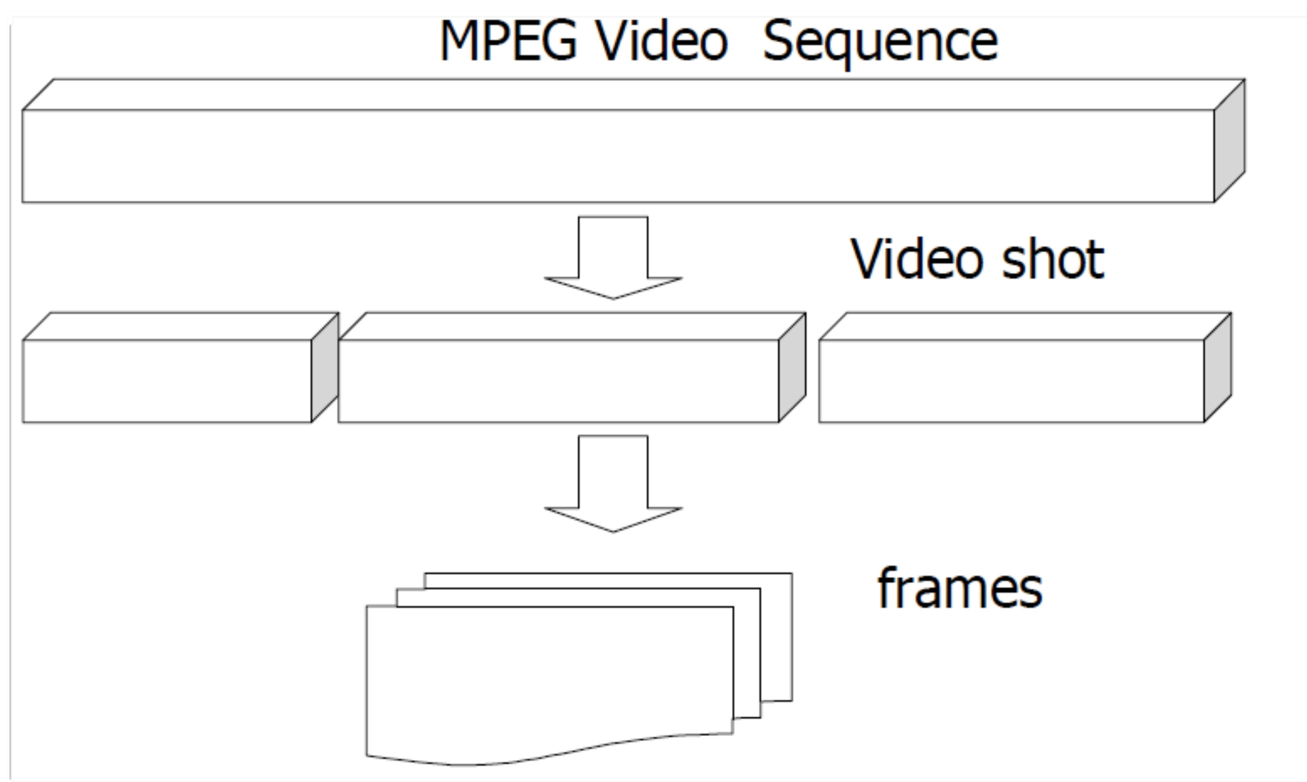
Overview (cont.)



[Wang, Li, Wiederhold, 2001]

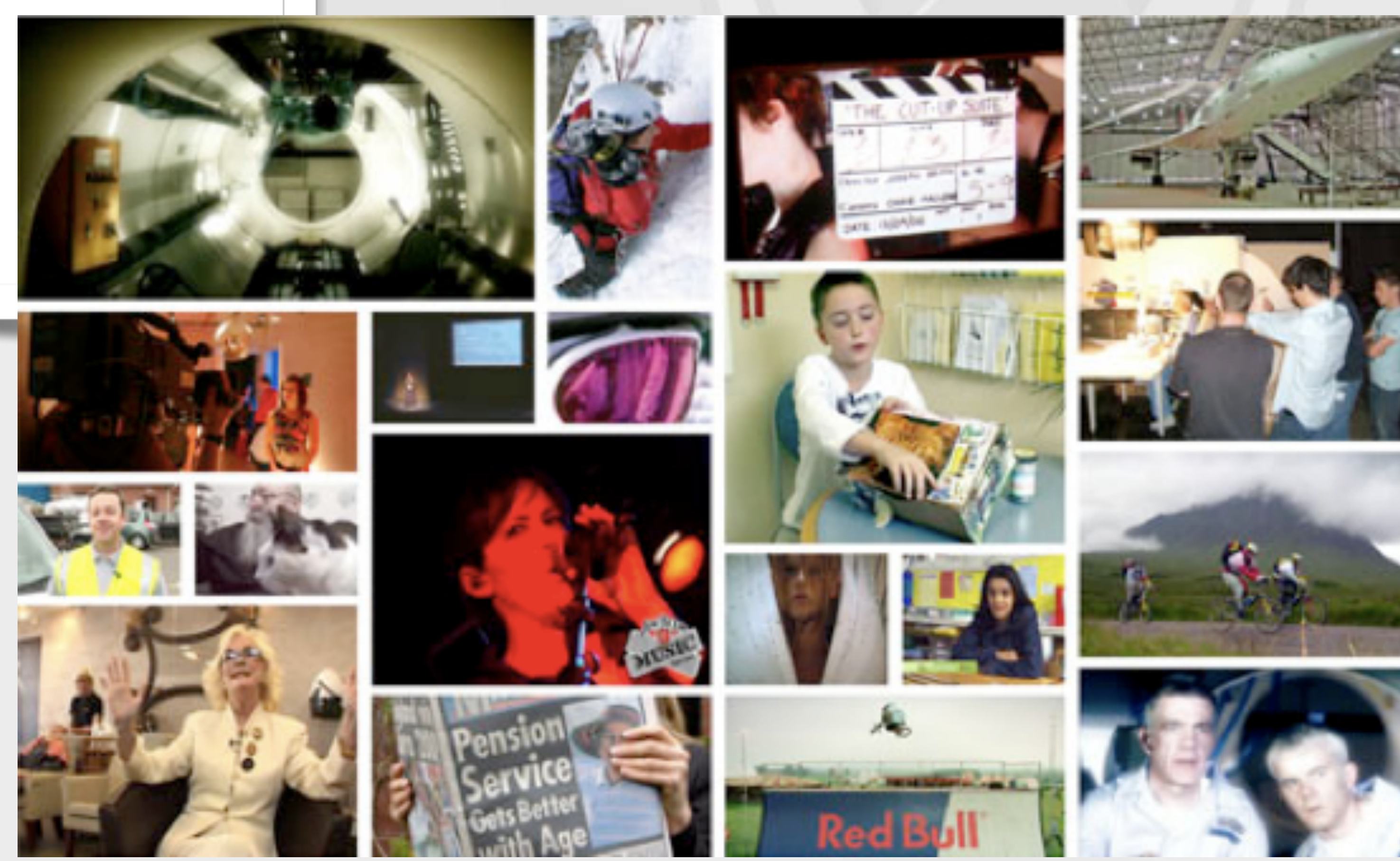
Structuralizing video data

- **semantic content layers**, e.g., scenes and shots in a video program.
  - These layers are erased when they are displayed for audience, which weakens the ability for user dealing with raw video data.



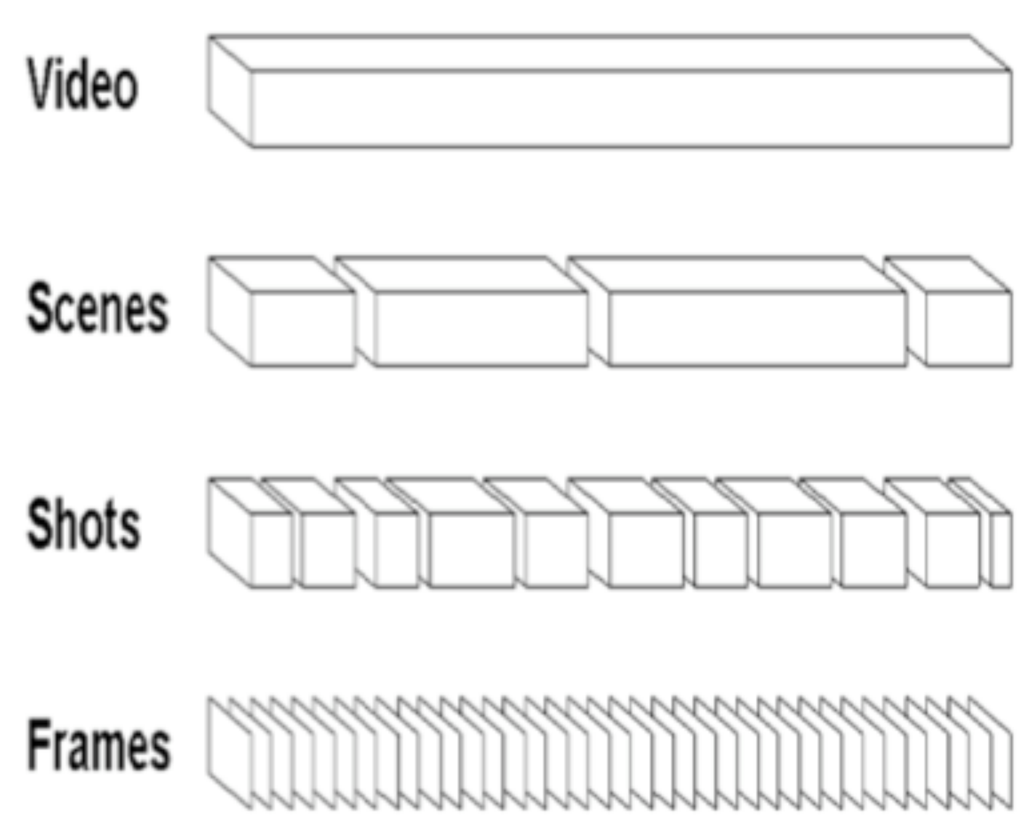
Fundamental definitions in video structurization

- Frame (帧)
- Shot (镜头)
- Key frame (关键帧)
- Scene (场景)
- Group (组)



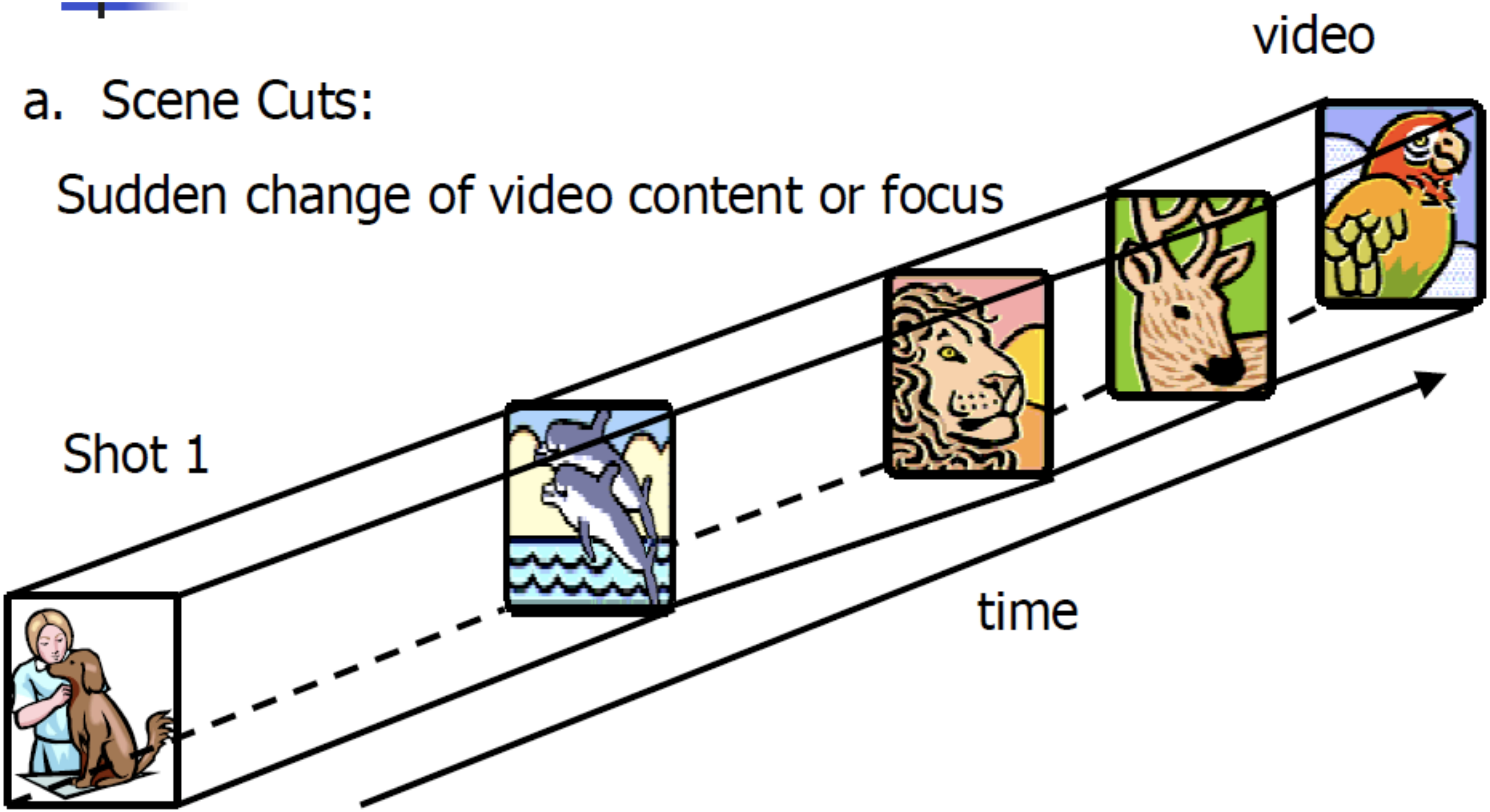
Fundamental definitions in video structurization

- Frame (帧)
- Shot (镜头)
- Key frame (关键帧)
- Scene (场景)
- Group (组)



a. Scene Cuts:

Sudden change of video content or focus



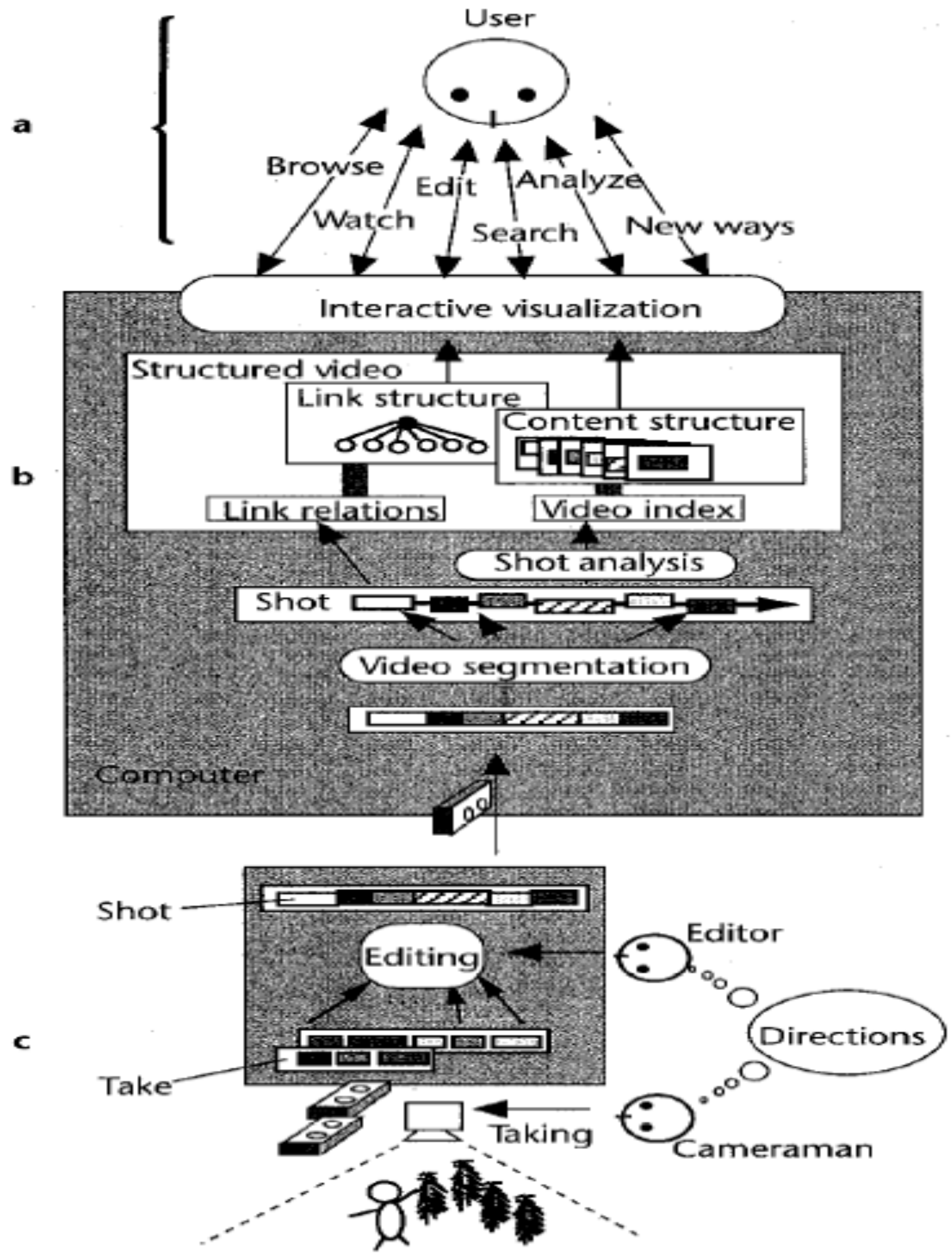
Proposal

- Analyze a video stream
- Segment the stream into shots
- Index shots using extracted features
  - Camera work characteristics
  - Color representations
- Browsing methods and user interfaces

Desired Video Interaction

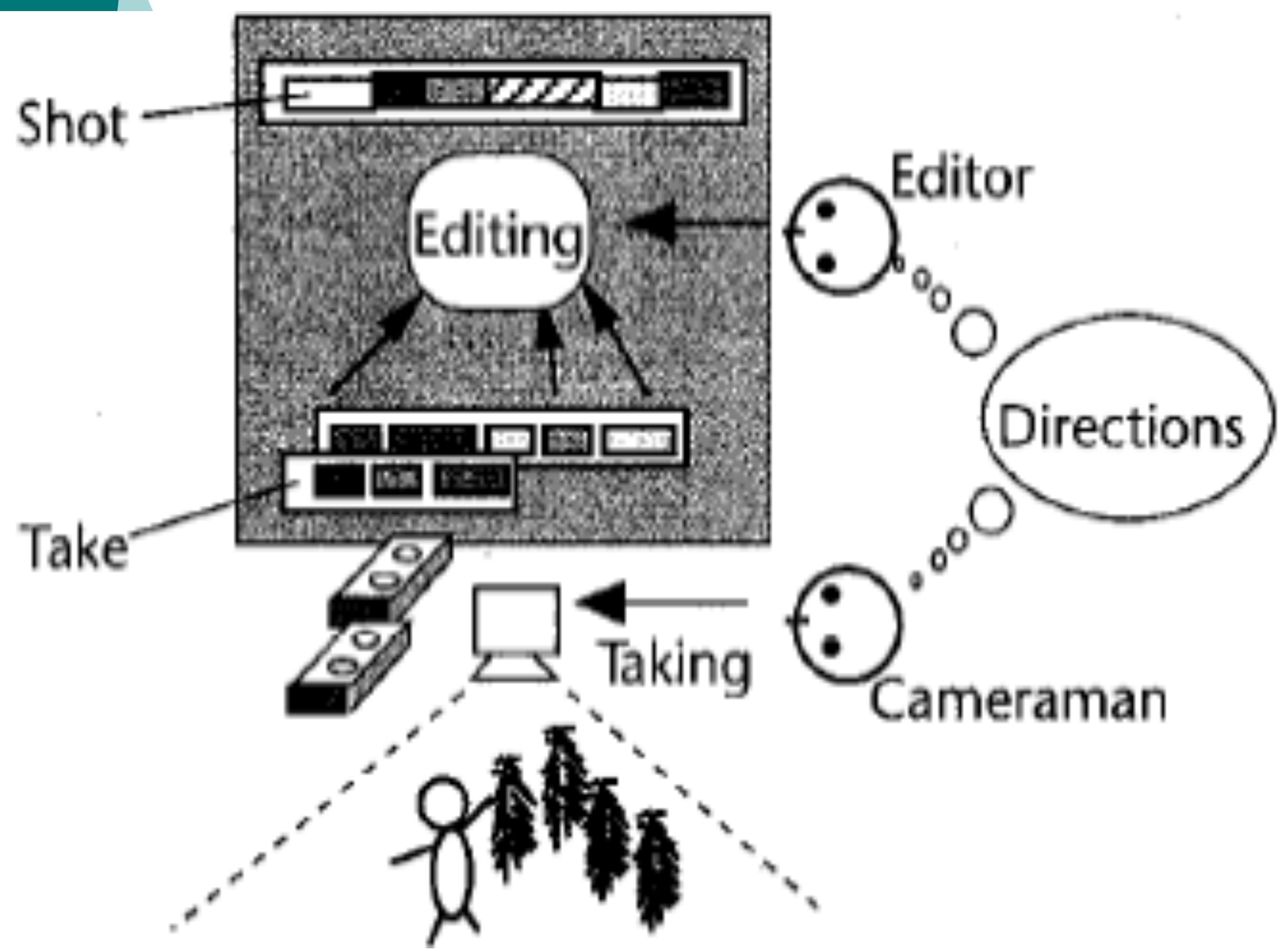
- Focus on fast visual browsing.
- Ability to grasp idea of lengthy video in short time.
- Not simply fast forward.
- Challenge: find and manage essential visual cues, then present them visually in an effective way

Viewer-Video Interaction: Conceptual Model



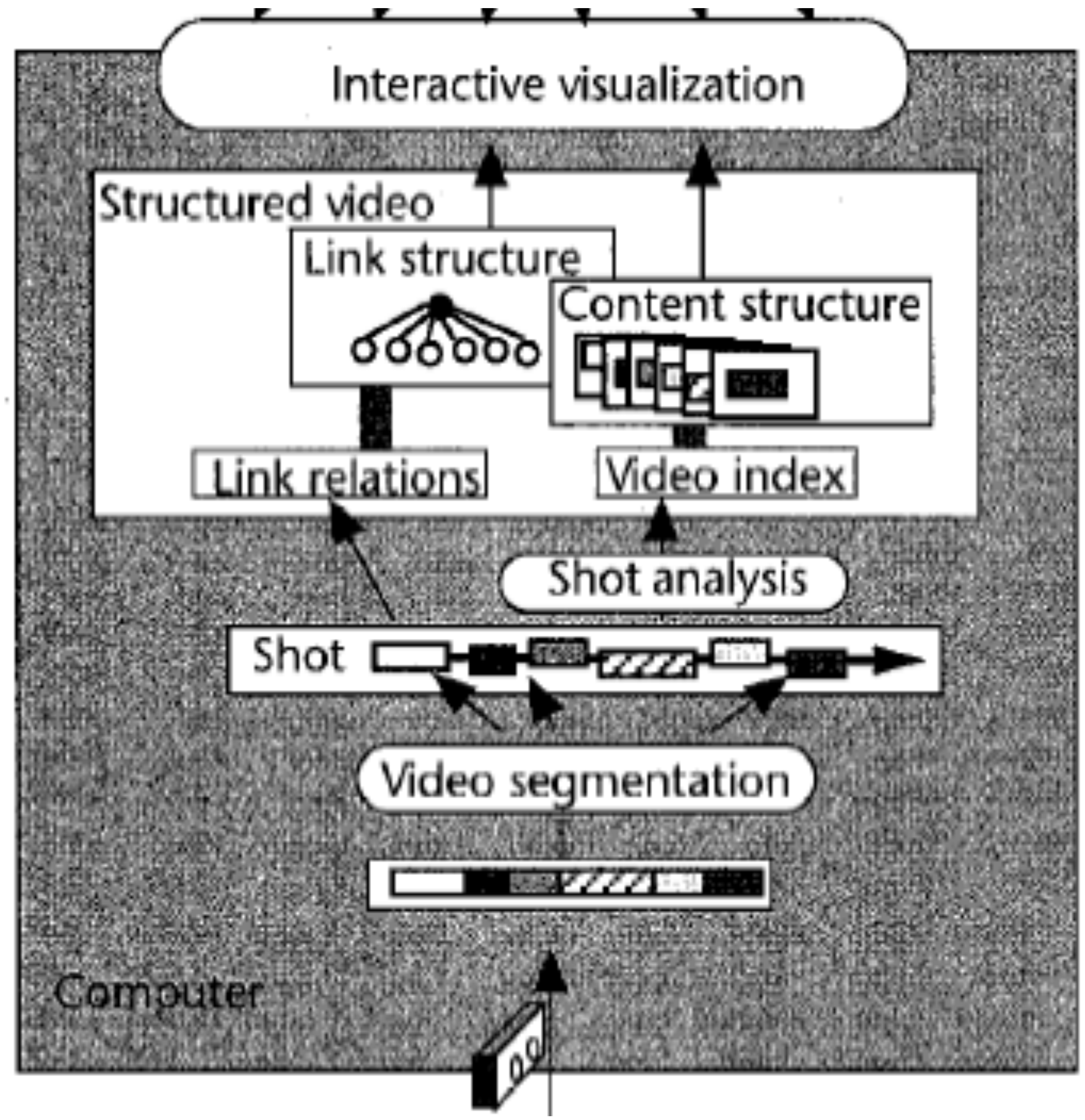
- a) Viewer Interaction
- b) Video Computing
- c) Video Production & Editing

Video Production



- Key Concepts:
  - Take: continuous video
  - Cut: separates takes
  - Camera characteristics
    - Pan, tilt, zoom, etc.
  - Shot: edited takes
- Resulting video contains embedded info: cut points, camera characteristics

Video Computing



- Main Function: Make the implied video structure explicit.

Video Segmentation: Problems

- Traditional Cut Detection – detect differences between frames using inter-frame comparisons (intensity, RGB, motion vectors).
- Misdetection due to rapid object motion, slow motion, animation, strobes, fading, wiping, dissolving, etc.
- Result: Low successful detection rate.

Basic video segmentation metrics

- Pair-wise comparison
  - Pixel-level
    - Sensitive to camera movement and motion
  - Block-level (Likelihood ratio)
    - Can tolerate small motion

$$DP_i(k,l) = \begin{cases} 1 & \text{if } |P_i(k,l) - P_{i+1}(k,l)| > t \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\sum_{k,l=1}^{M,N} DP_i(k,l)}{M * N} * 100 > T$$

$$\frac{\left[\frac{S_i + S_{i+1}}{2} + \left(\frac{m_i - m_{i+1}}{2}\right)^2\right]^2}{S_i * S_{i+1}} > t$$

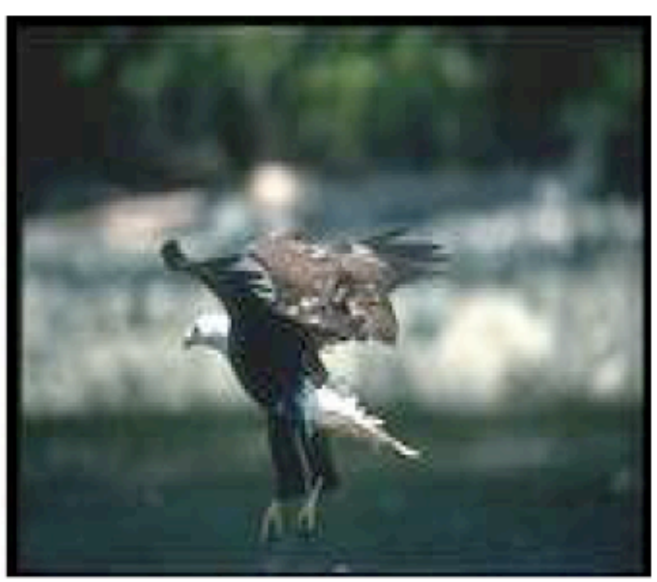
mi: mean intensity

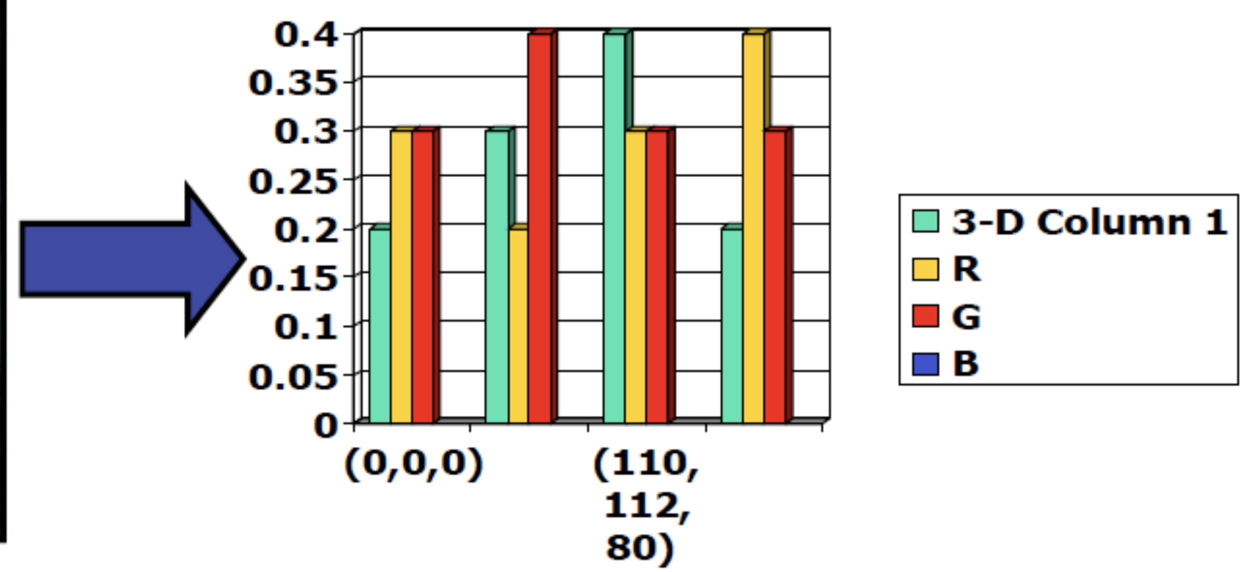
Si: corresponding variance

Basic video segmentation metrics

How to measure statistical property of video frames?

Color Histogram





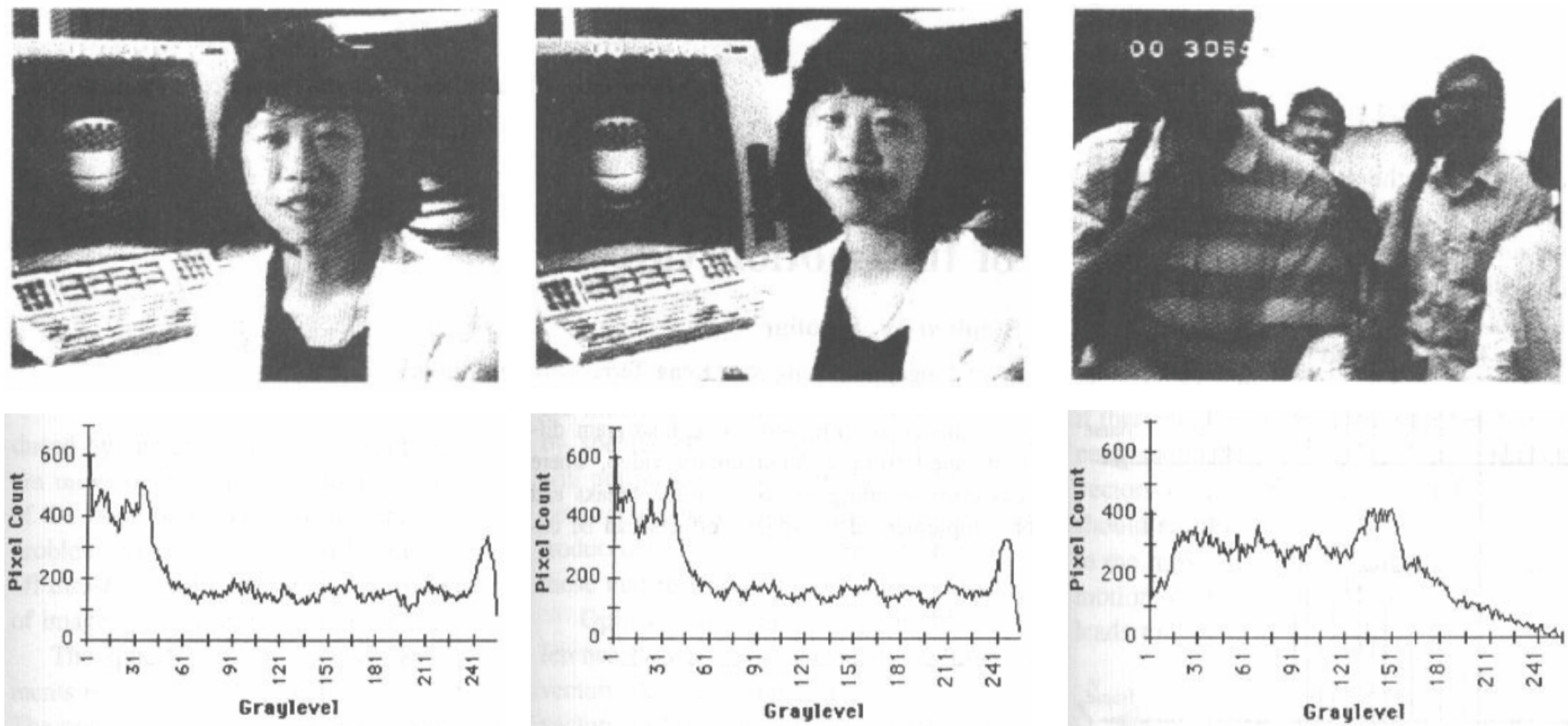
Basic video segmentation metrics

- Histogram comparison
  - Basic
    - Tolerate motion better
  - χ2-test
  - Color level can also be used but only the MSB to save the number of bins

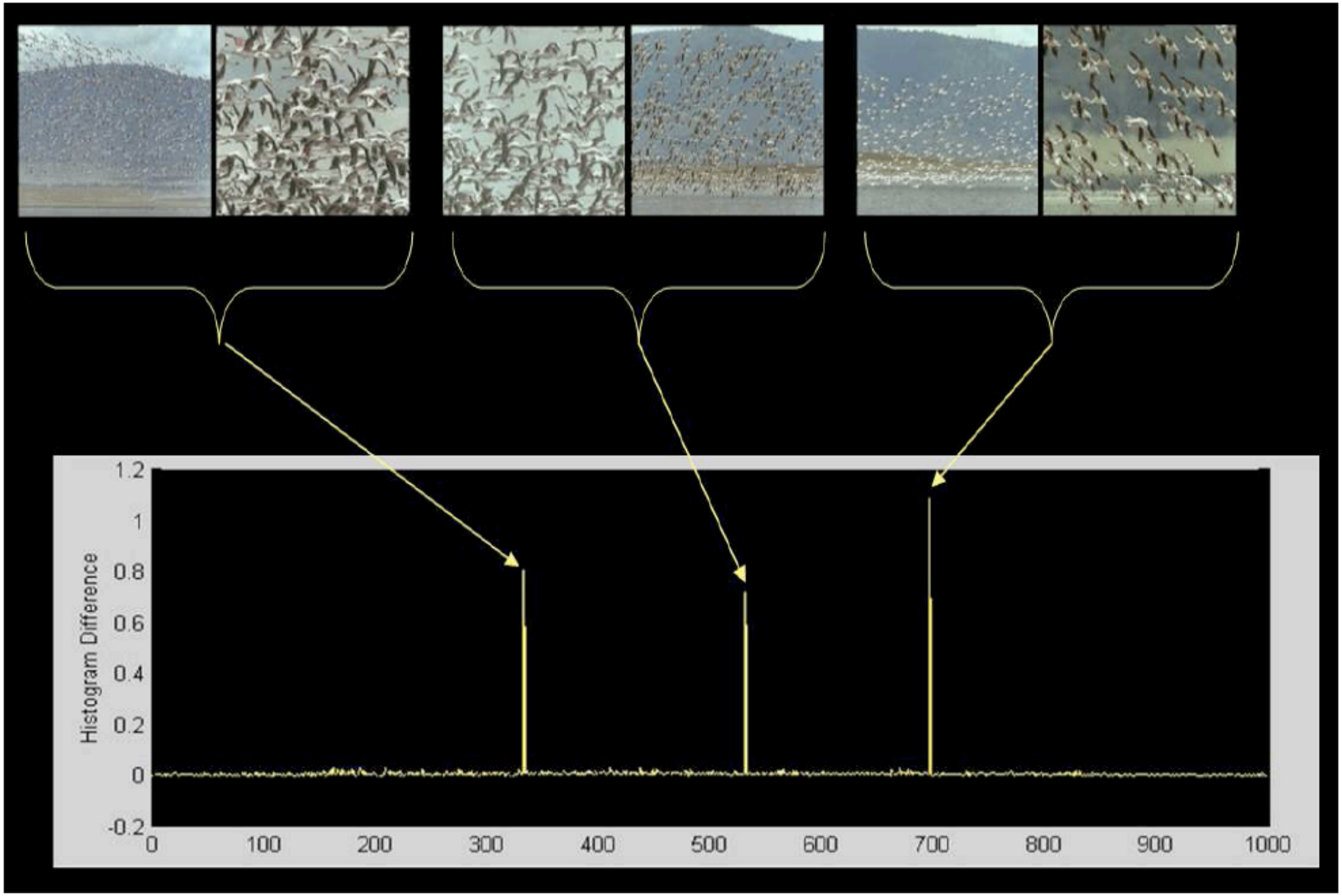
$$SD_i = \sum_{j=1}^{G} |H_i(j) - H_{i+1}(j)|$$

$$SD_i = \sum_{j=1}^{G} \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)}$$

Sample of using histogram

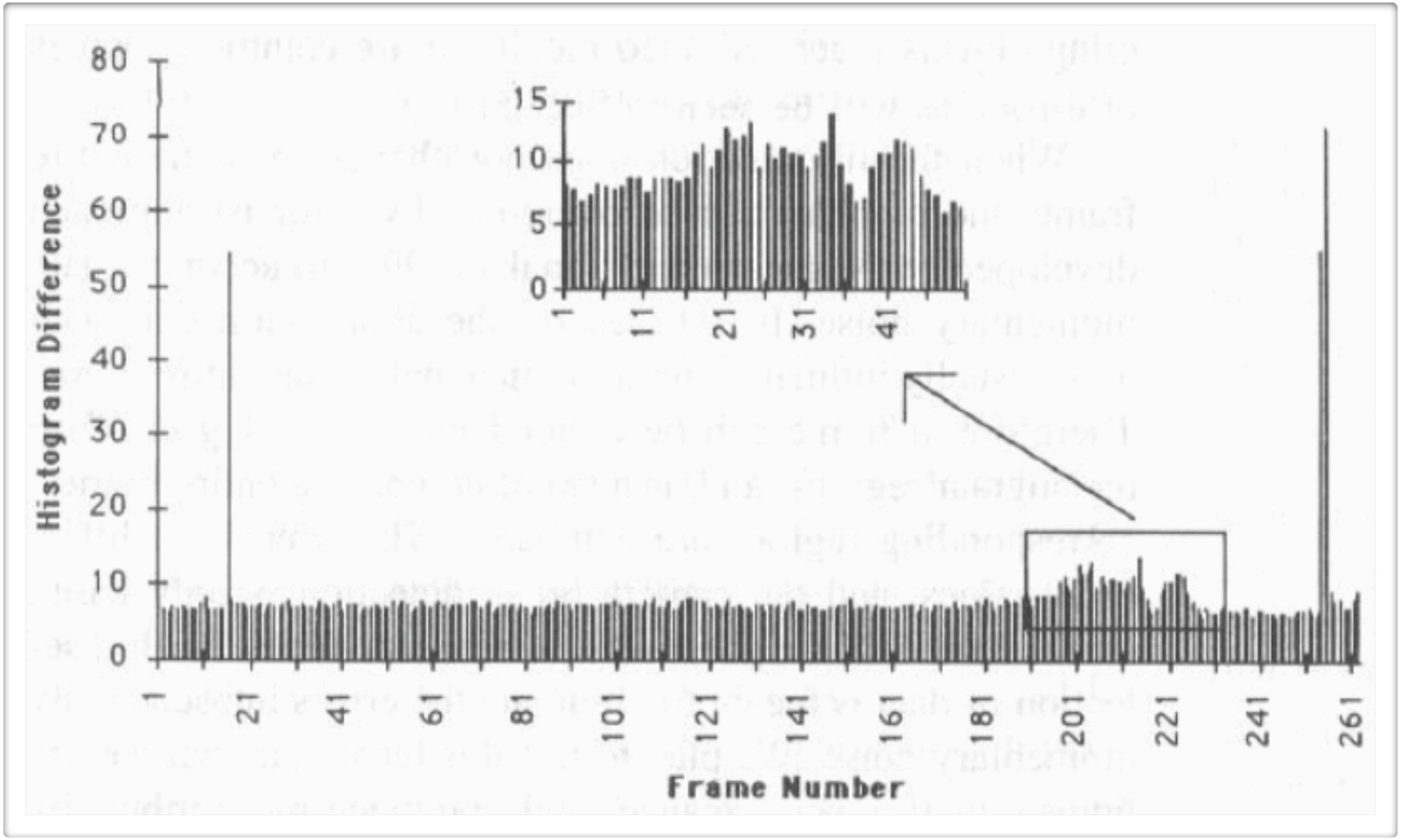


Scene Cut



Histogram Difference
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
100
200
300
400
500
600
700
800
900
1000

Gradual transition detection



Gradual transition detection

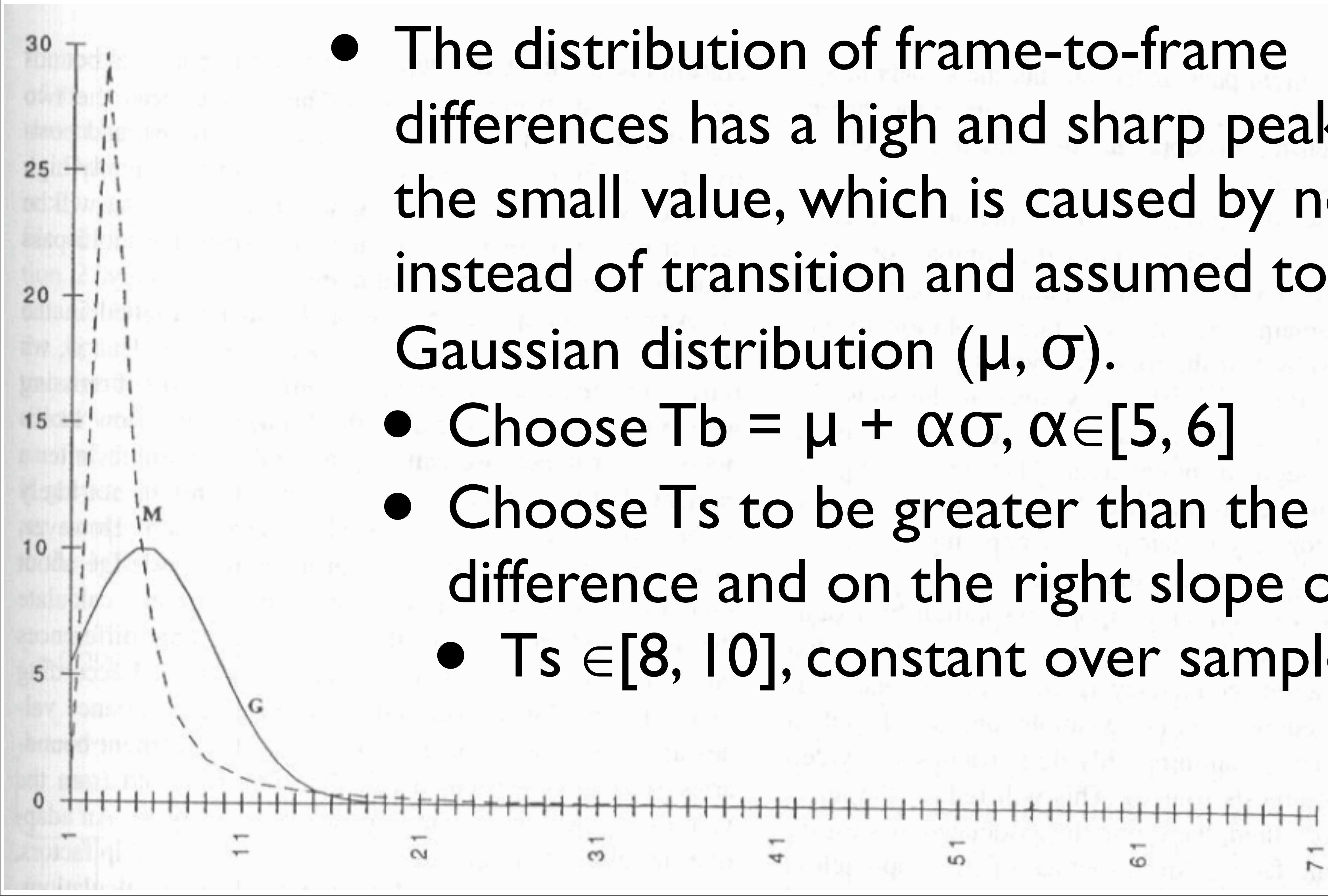
- Twin-comparison
  - Use two thresholds Tb and Ts to accommodate both **short-term** and **long-term** transitions
  - Differences of (F1, F2), (F2, F3), (F3, F4) are small, but difference of (F1, F4) is still big

1 2 3 4

- Twin-comparison
  - F_s — the potential beginning frame of the transition
  - F_e — the ending frame of the transition

scan frame
if $(Diff(F_i) \geq T_b)$
    detect as camera break
else if $(T_b > Diff(F_i) \geq T_s)$
    $F_s \leftarrow F_i$
    $i \leftarrow i + 1$
    while $(Diff(F_i) \geq T_s)$
        $i \leftarrow i + 1$
    if $(Diff(F_s, F_i) \geq T_b)$
        $F_e \leftarrow F_i$

Threshold selection (Tb, Ts)



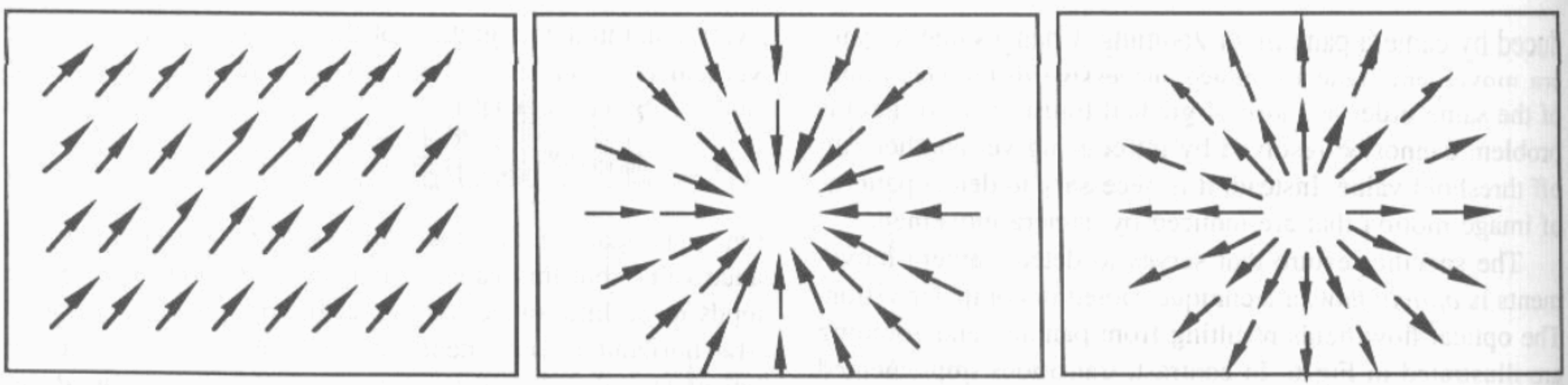
- The distribution of frame-to-frame differences has a high and sharp peak near the small value, which is caused by noise instead of transition and assumed to follow Gaussian distribution (μ, σ).
  - Choose $Tb = \mu + \alpha\sigma$, $\alpha \in [5, 6]$
  - Choose Ts to be greater than the mean difference and on the right slope of M
    - $Ts \in [8, 10]$, constant over samples

Multi-pass approach

- Scanning all frames could be computationally hard
- Temporal skipping is more useful
  - e.g. one out of every 10 frames
  - Better for detecting gradual transition
  - May miss camera break
  - May get false detection (distance increased)
- Multi-pass approach
  - First pass, use either pair-wise or histogram with large skip factor and smaller Tb to collect the potential regions.
  - Second pass, two methods may be applied together (hybrid) to search the candidate regions while increasing the confidence.

Distinguish camera movement

- To distinguish gradual transitions from changes made by camera movements
- Basic approach— observing optical flow via motion vectors



a b c

Fig. 6a–c. Motion vector patterns resulting from camera panning and zooming. a ⁄ Camera panning direction. b Camera zoom-out. c Camera zoom-in

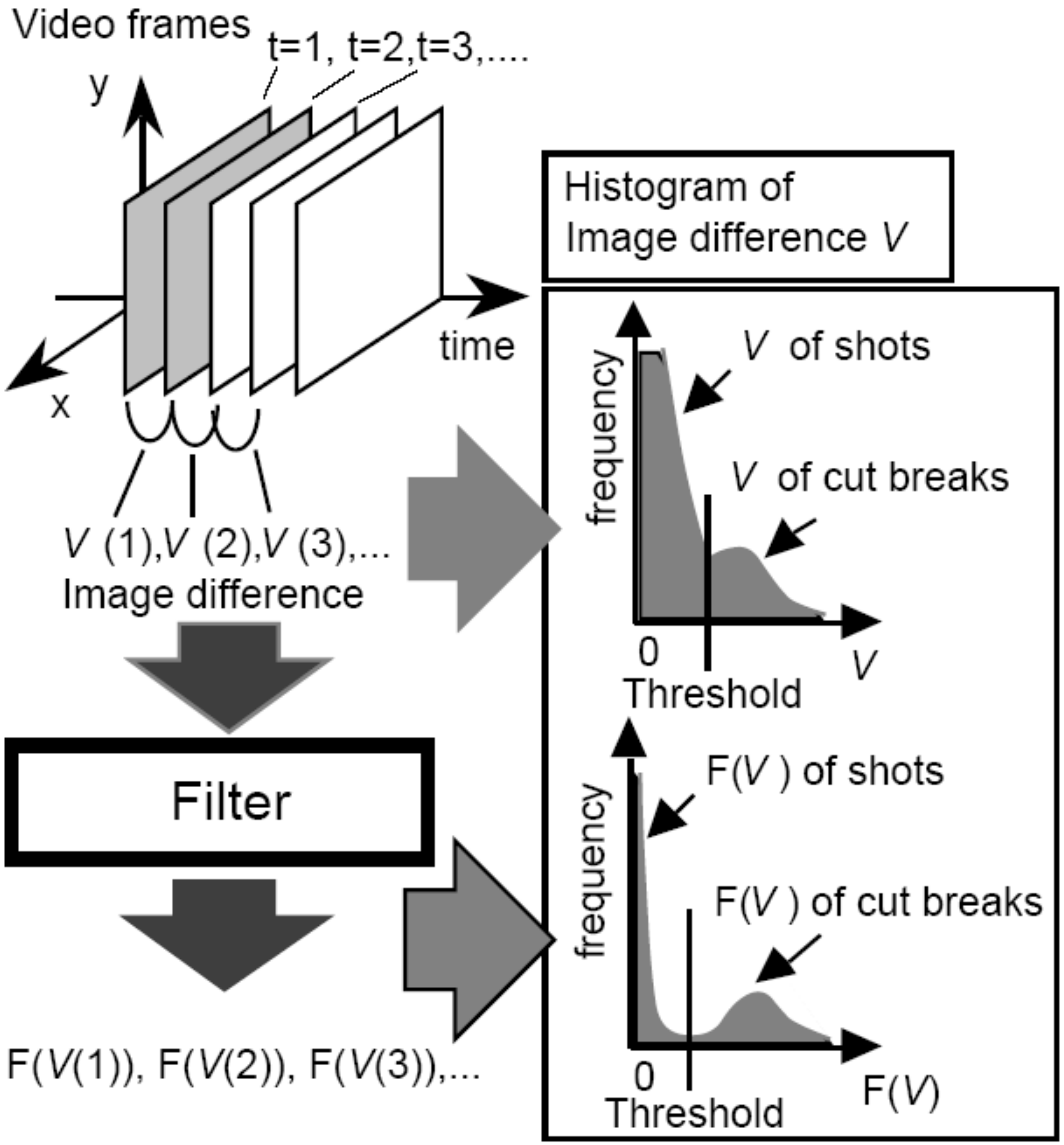
Distinguish camera movement

- **Panning**
  - Distribution of motion vectors has a single modal value (θm) that corresponds to the panning direction.

$$\sum_{k}^{N} |\theta_k - \theta_m| \leq \Theta_p$$

- **Zooming**
  - The vertical components of top and bottom motion vectors have different signs.
  - Similarly for horizontal components of left and right motion vectors.

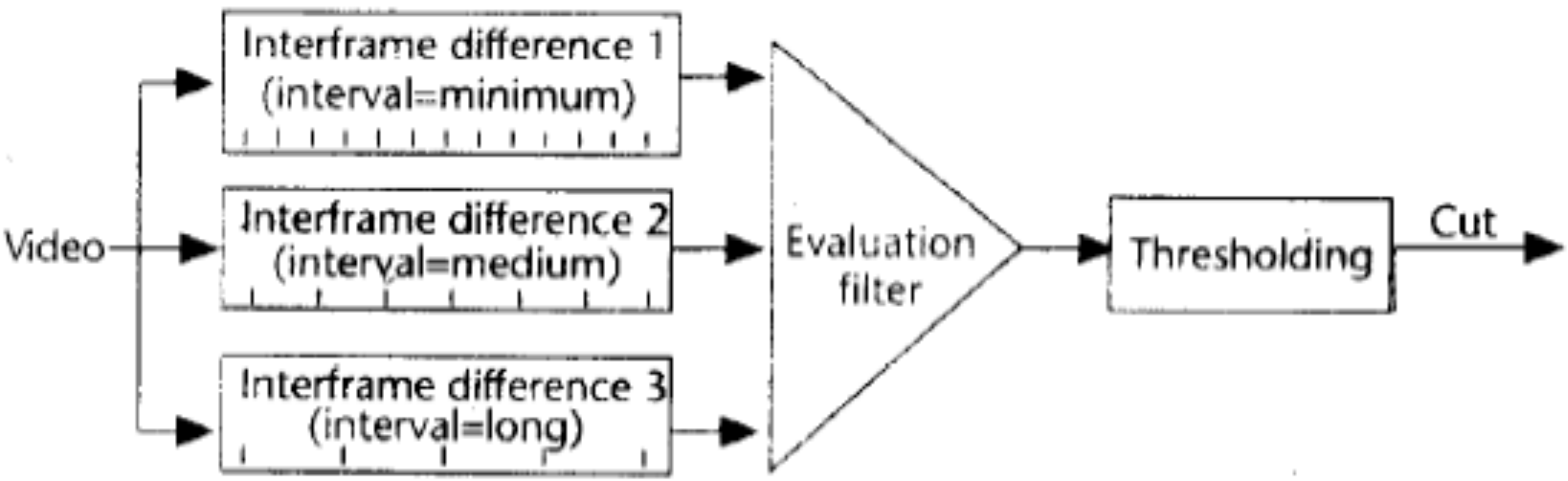
Yet Another Video Segmentation



V = image difference

Video Segmentation: Solution

- 92-98% success rate over 4.5 hours of video (news, movies, documentaries)
- 90% success when 1/3 of all cuts were via special affects

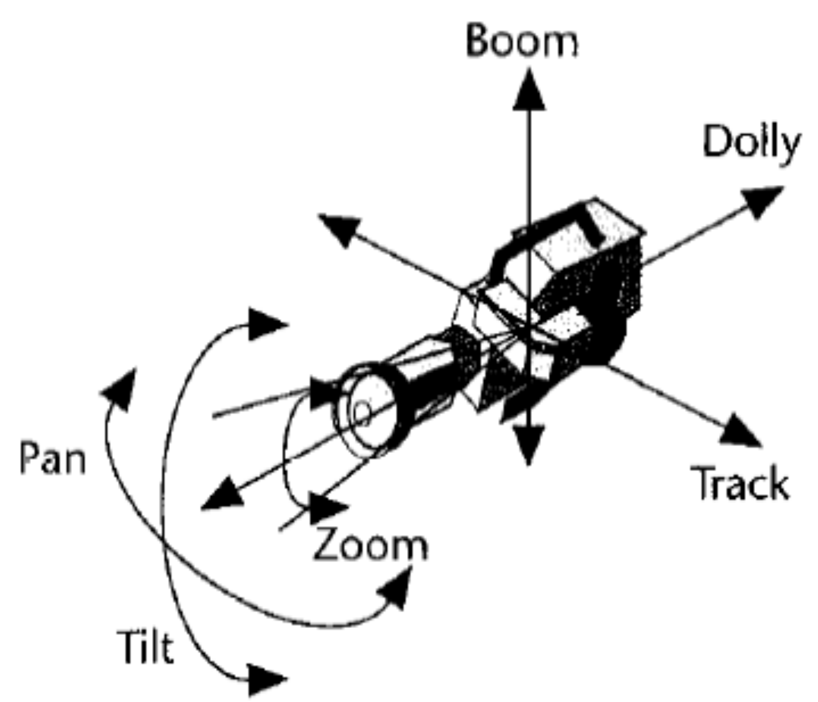


Shot Analysis

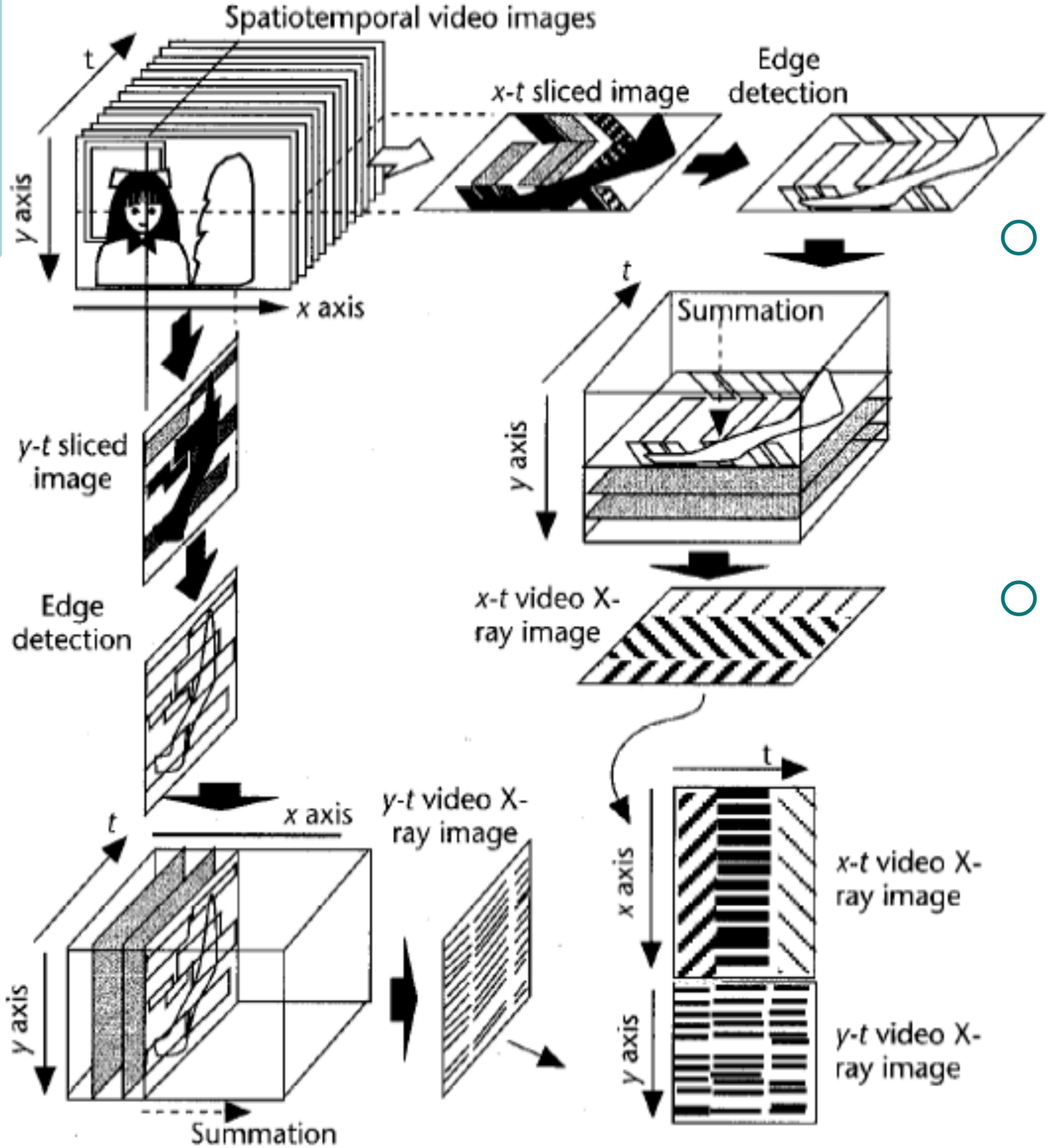
- Shot is simply sequence of frames capturing a scene’s spatial and temporal context.
- Extract this information:
  - Camera work yields spatial situation
  - Color info yields object information

Camera Work Information Extraction

- Camera movement causes global change in objects.
- Resulting point traces = motion vectors
- Motion vectors yield camera work parameters
- Computationally complex, not robust



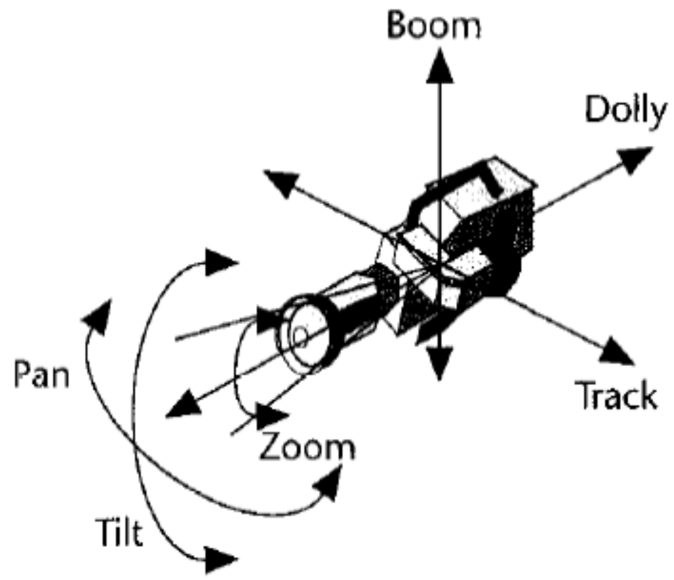
Camera Work Information Extraction

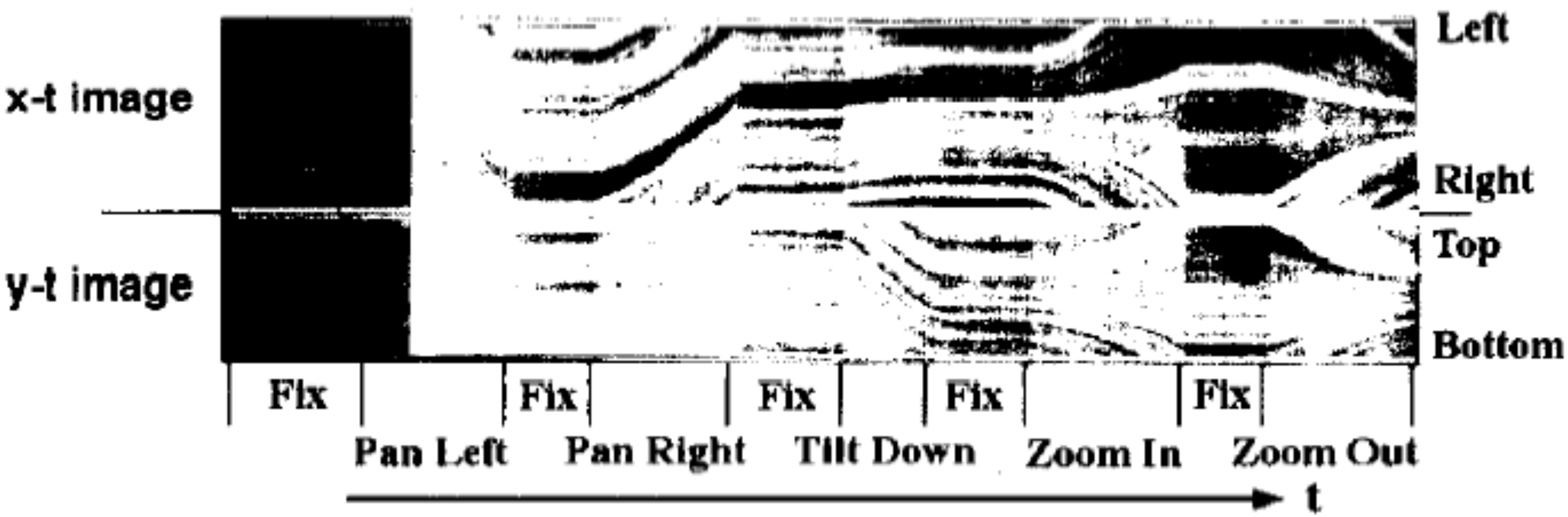


- Proposal based on video x-ray imaging.
- Easy calculations, robust

Camera Work Information Extraction

- Parallel to time = fixed camera
- Slant = camera pan
- Degree of slant = speed of pan
- Line spread = zoom
- No information present for track and dolly

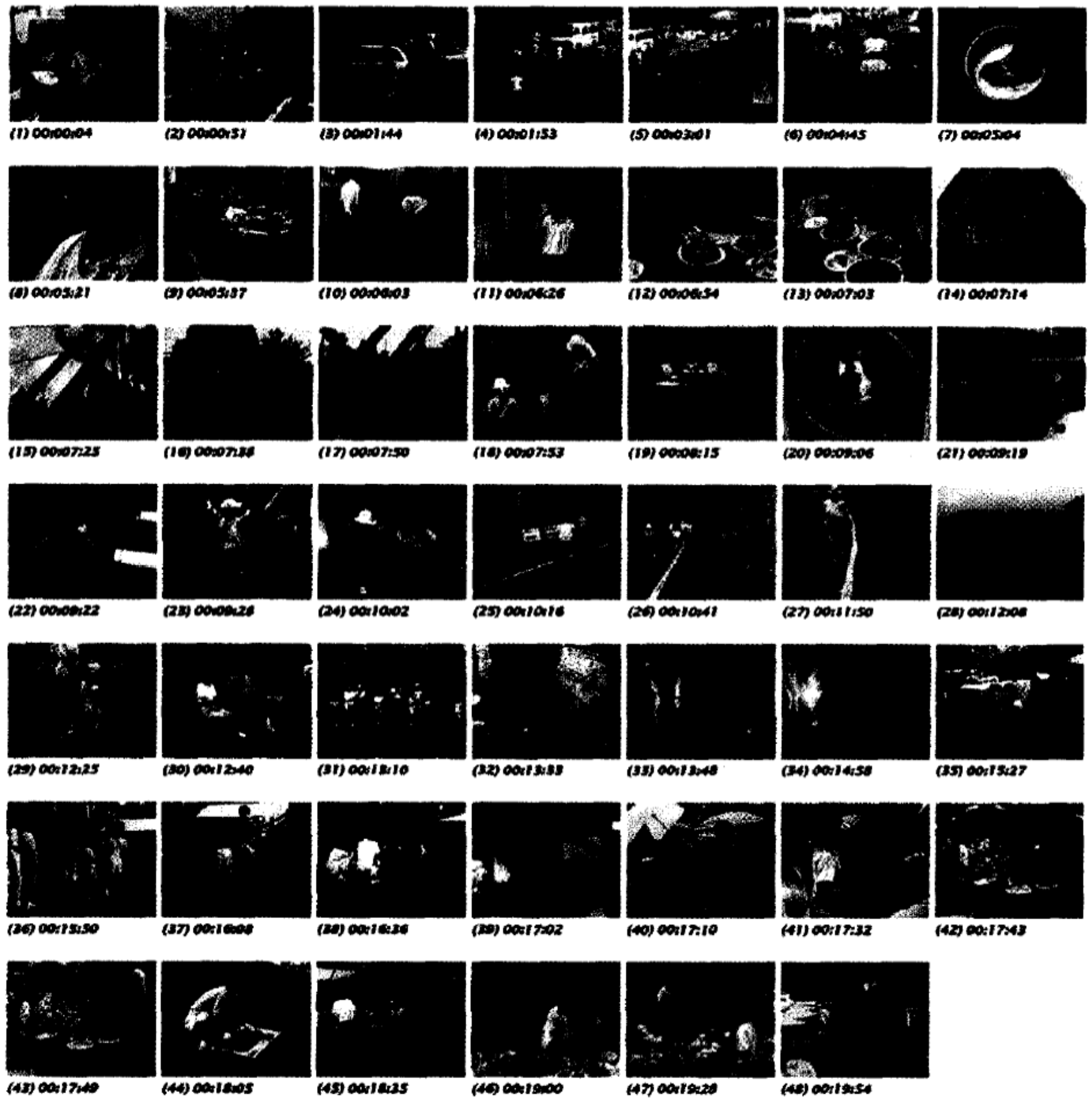




New Video Interfaces

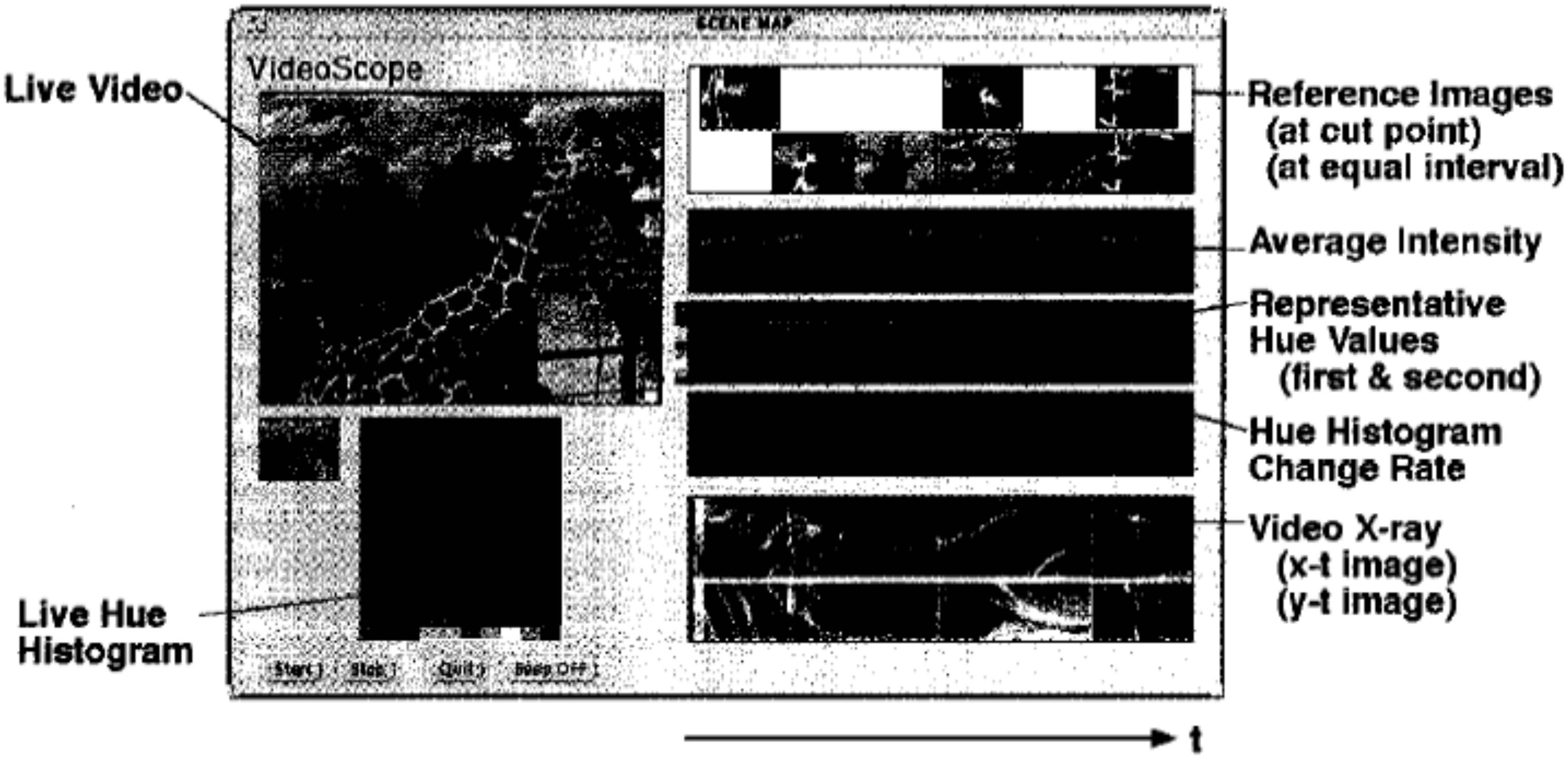
- VideoScope
- VideoSpaceIcon
- ViewSpaceMonitor
- PaperVideo

PaperVideo



- Photo albums and video indexing.
- Shows potential simplicity of structured video apps.

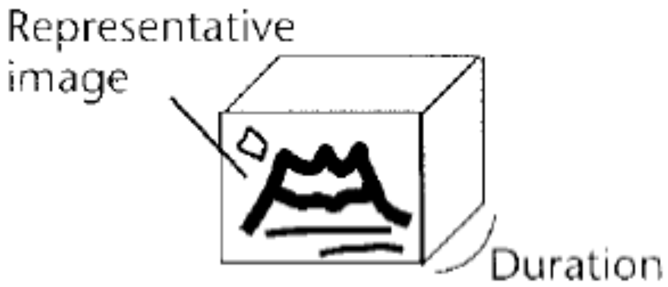
VideoScope

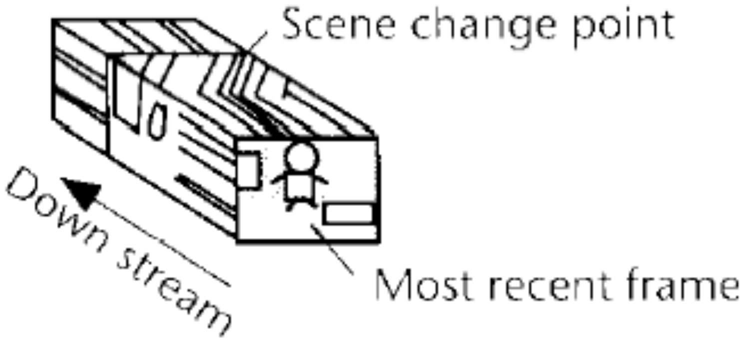


- Possible use as video engineering tool.
- Shows potential complexity of structured video apps.

Related Work

- Importance of visual interface
  - Must activate user's visual sense
  - Must stimulate user's ability to manipulate video





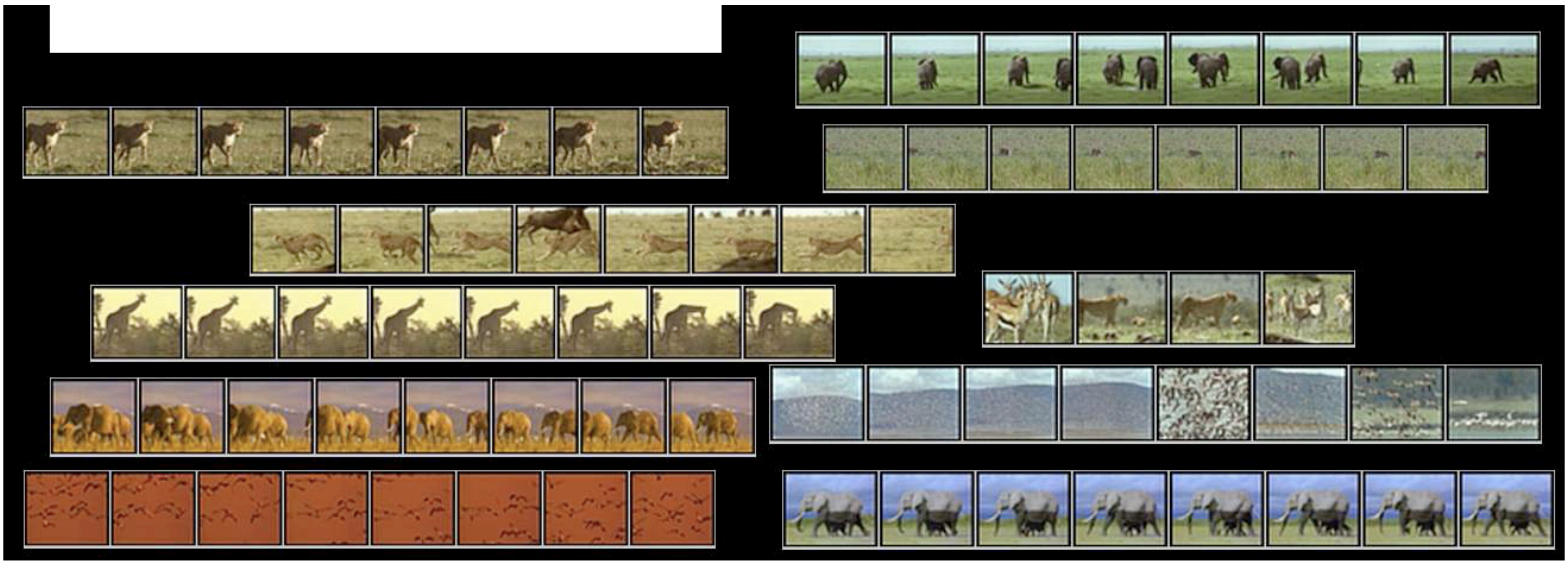
- What can be done in video production stage?

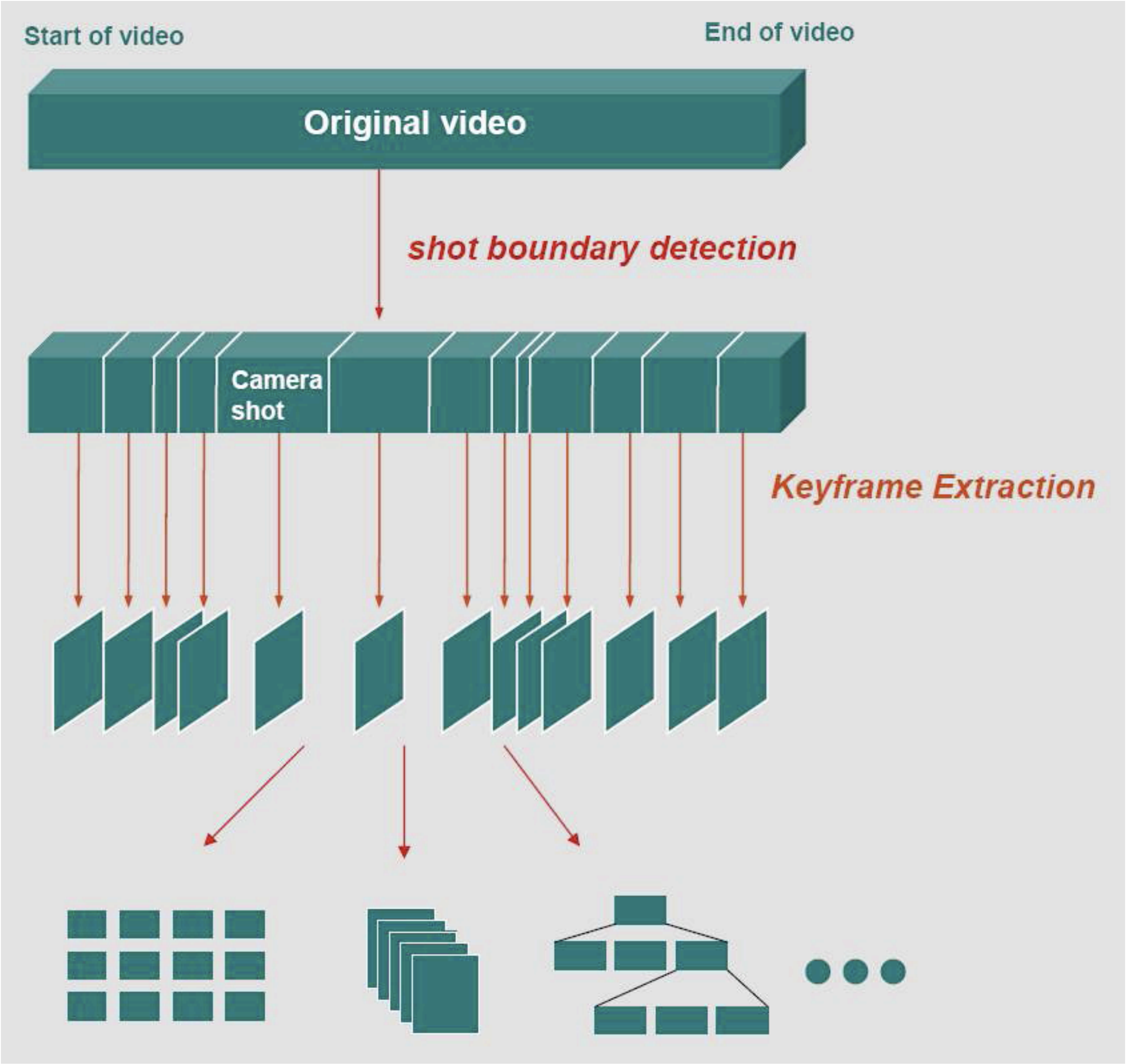
Notable Reference

Cut Detection

- K. Otsuji, Y. Tonomura, “Projection Detecting Filter for Video Cut Detection,” *Proc. ACM Multimedia 93*, ACM Press, New York, 1993.

Keyframe extraction




Start of video
End of video
Original video
shot boundary detection
Camera
shot
Keyframe Extraction

Reference

- Key Frame Extraction

http://www.cs.ust.hk/~rossiter/mm_projects/video_key_frame/key_frame_index.html

Reference

- Key Frame Extraction

http://www.cs.ust.hk/~rossiter/mm_projects/video_key_frame/key_frame_index.html

关键帧提取技术

- 镜头边界法
  - 选取镜头中的首帧和末帧
- 颜色特征法
  - 首帧为关键帧，其后比较与前面关键帧的颜色差异
- 运动分析法
  - 分析相机的运动
- 聚类分析法

聚类分析法

- 设一个镜头 $S = \{f_1, f_2, ..., f_m\}$
  - 找关键帧 $[F_1, F_2, ..., F_n]$
  - 定义帧间距离 $d(f_i, f_j)$

Step 0. 设定阈值，选定初始n个关键帧位置

Step 1. 按照到关键帧的最小距离重新划分

Step 2. 指定每一聚类的中心帧为新的关键帧。

如果与上次划分区别不大则停止，否则重复 Step 1和Step 2.

Brain storm

Brain storm

专辑:强殖装甲
视频:26 时长:10:10:51 播放:42,021

专辑:强殖装甲
视频:26 时长:10:00:23 播放:2,400

专辑:强殖装甲
视频:27 时长:10:11:26 播放:1,982

更多相关专辑>>

强殖装甲 1:31:28
强殖装甲强殖装甲强殖装甲
强殖装甲
malinkof 3个月前
播放: 17,361 | 评论: 21 | 收藏: 21
2条相似结果

Brain storm

专辑:强殖装甲
视频:26 时长:10:10:51 播放:42,021

专辑:强殖装甲
视频:26 时长:10:00:23 播放:2,400

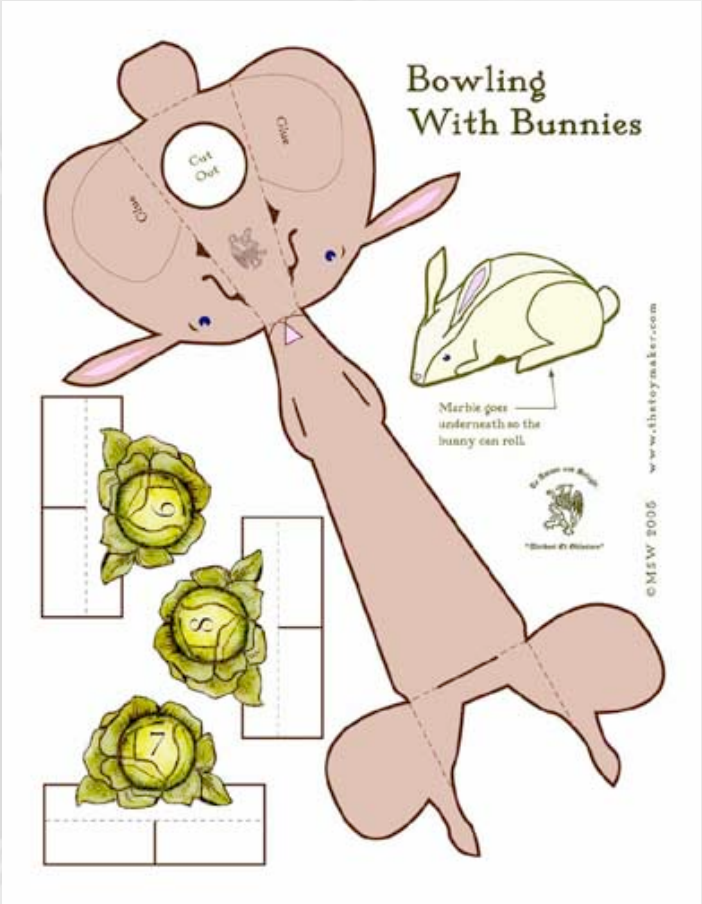
专辑:强殖装甲
视频:27 时长:10:11:26 播放:1,982

更多相关专辑>>

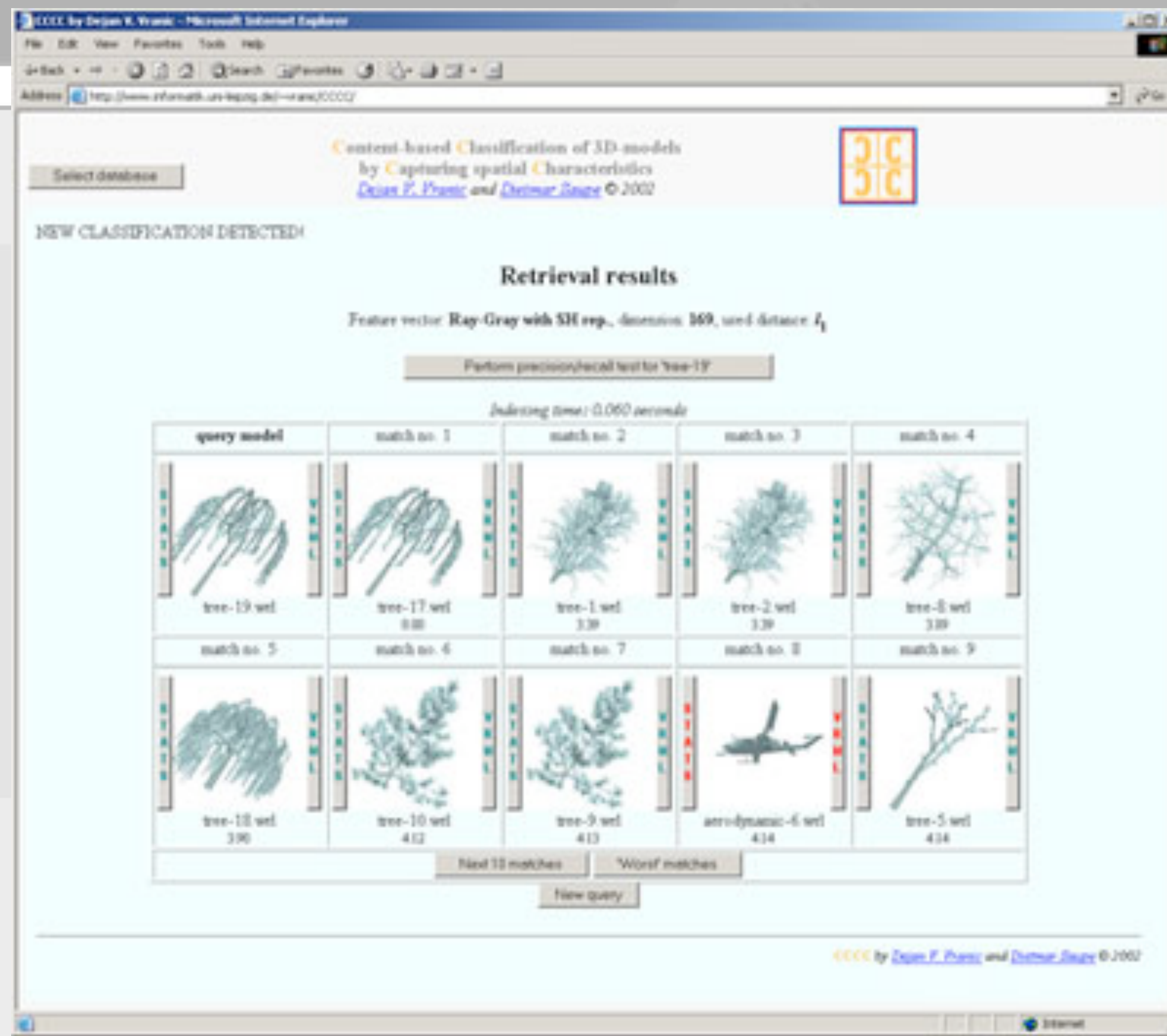
强殖装甲 1:31:28
强殖装甲强殖装甲强殖装甲
强殖装甲
malinkof 3个月前
播放: 17,361 | 评论: 21 | 收藏: 21
2条相似结果

?

4. Graphics retrieval techniques

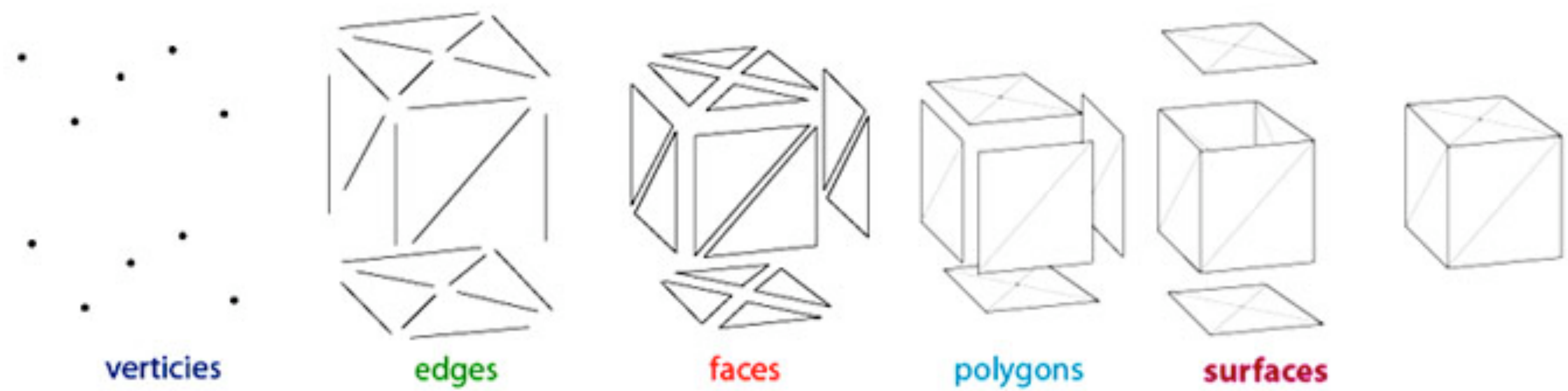


3D Model Similarity Search

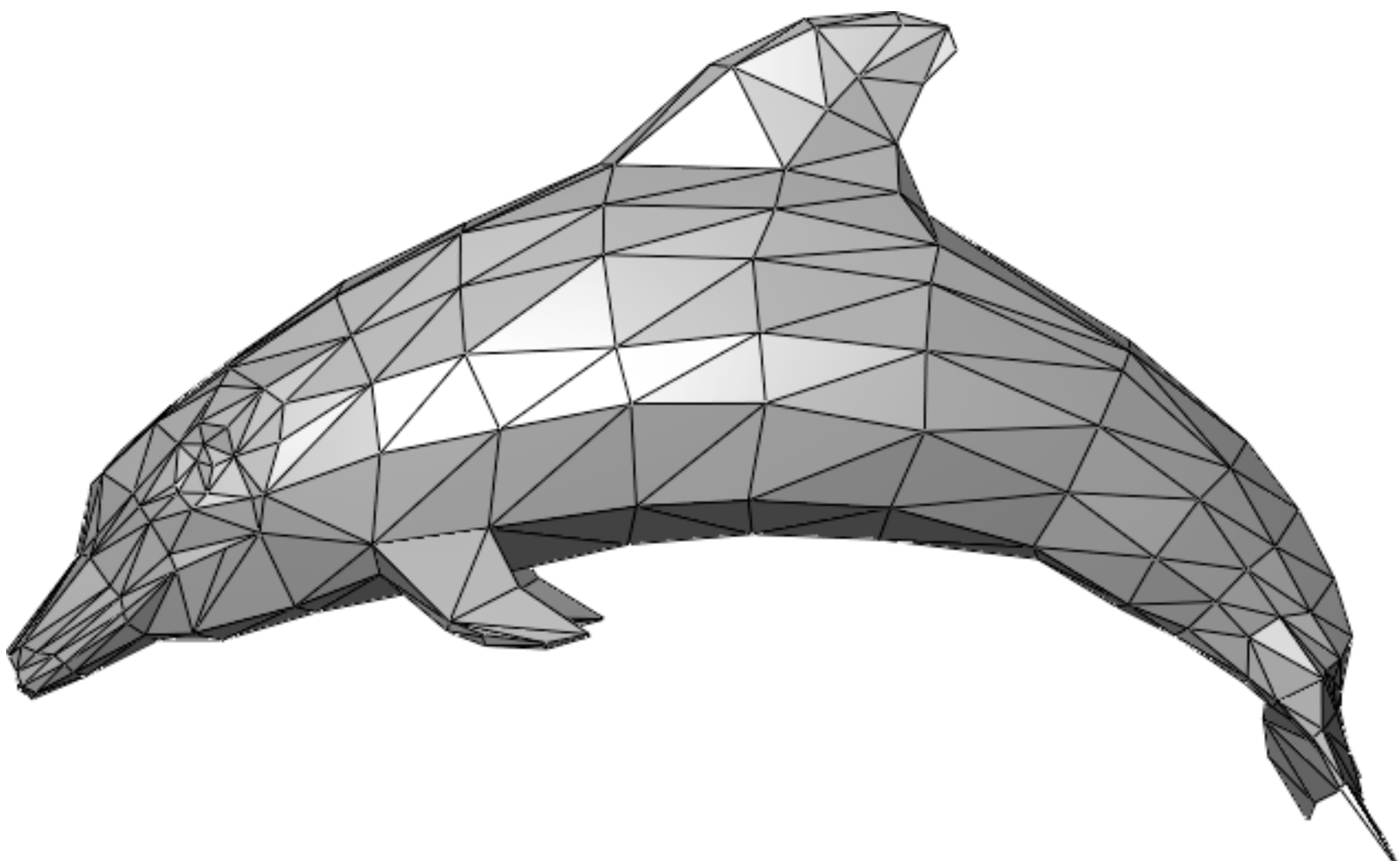


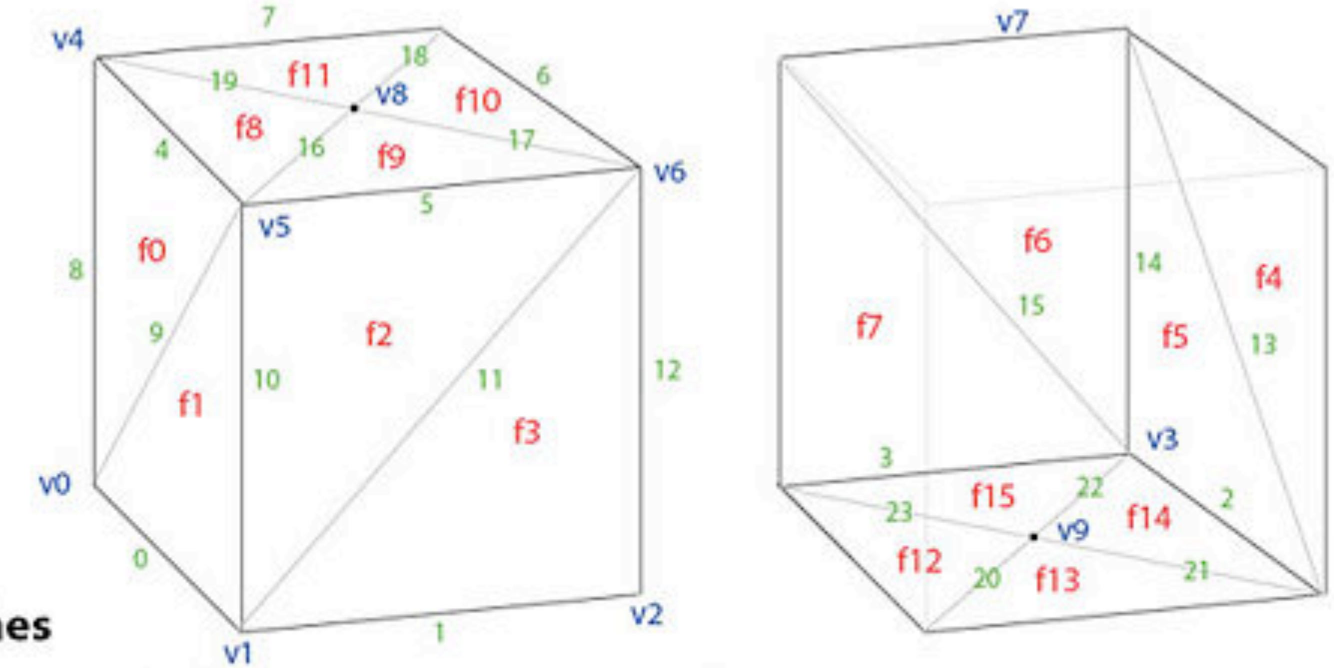
http://infovis.uni-konstanz.de/research/projects/SimSearch3D/

Elements of polygonal mesh modeling



Triangle mesh





Winged-Edge Meshes

Face List

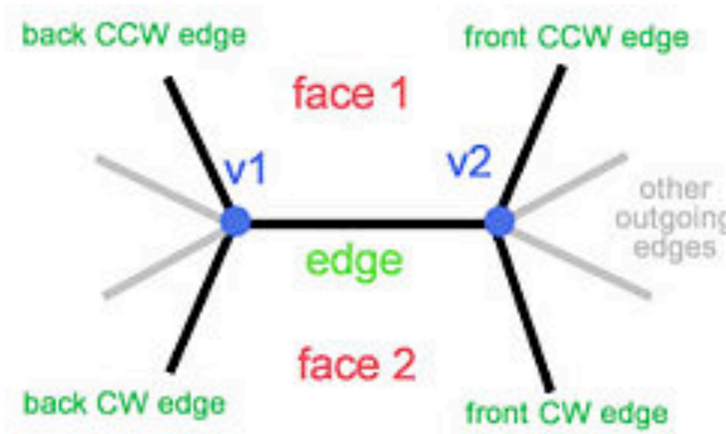
f0	4 8 9
f1	0 10 9
f2	5 10 11
f3	1 12 11
f4	6 12 13
f5	2 14 13
f6	7 14 15
f7	3 8 15
f8	4 16 19
f9	5 17 16
f10	6 18 17
f11	7 19 18
f12	0 23 20
f13	1 20 21
f14	2 21 22
f15	3 22 23

Edge List

e0	v0 v1	f1 f12	9 23 10 20
e1	v1 v2	f3 f13	11 20 12 21
e2	v2 v3	f5 f14	13 21 14 22
e3	v3 v0	f7 f15	15 22 8 23
e4	v4 v5	f0 f8	19 8 16 9
e5	v5 v6	f2 f9	16 10 17 11
e6	v6 v7	f4 f10	17 12 18 13
e7	v7 v4	f6 f11	18 14 19 15
e8	v0 v4	f7 f0	3 9 7 4
e9	v0 v5	f0 f1	8 0 4 10
e10	v1 v5	f1 f2	0 11 9 5
e11	v1 v6	f2 f3	10 1 5 12
e12	v2 v6	f3 f4	1 13 11 6
e13	v2 v7	f4 f5	12 2 6 14
e14	v3 v7	f5 f6	2 15 13 7
e15	v3 v4	f6 f7	14 3 7 15
e16	v5 v8	f8 f9	4 5 19 17
e17	v6 v8	f9 f10	5 6 16 18
e18	v7 v8	f10 f11	6 7 17 19
e19	v4 v8	f11 f8	7 4 18 16
e20	v1 v9	f12 f13	0 1 23 21
e21	v2 v9	f13 f14	1 2 20 22
e22	v3 v9	f14 f15	2 3 21 23
e23	v0 v9	f15 f12	3 0 22 20

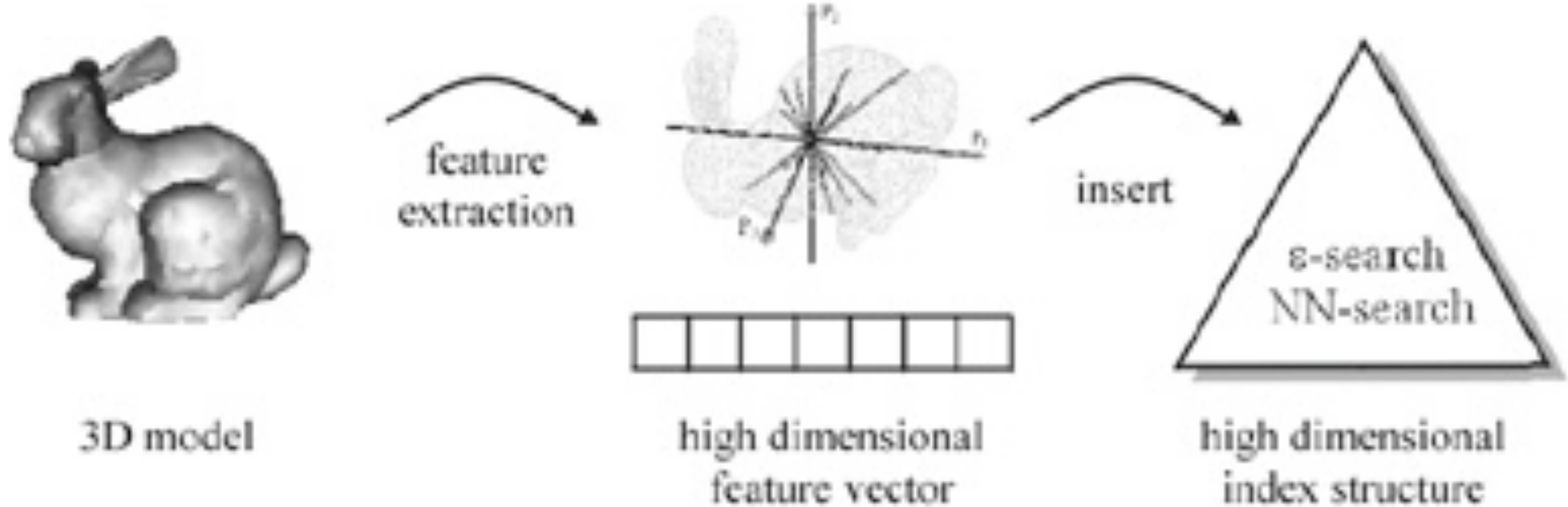
Vertex List

v0	0,0,0	8 9 0 23 3
v1	1,0,0	10 11 1 20 0
v2	1,1,0	12 13 2 21 1
v3	0,1,0	14 15 3 22 2
v4	0,0,1	8 15 7 19 4
v5	1,0,1	10 9 4 16 5
v6	1,1,1	12 11 5 17 6
v7	0,1,1	14 13 6 18 7
v8	.5,.5,0	16 17 18 19
v9	.5,.5,1	20 21 22 23



Winged Edge Structure

Main idea

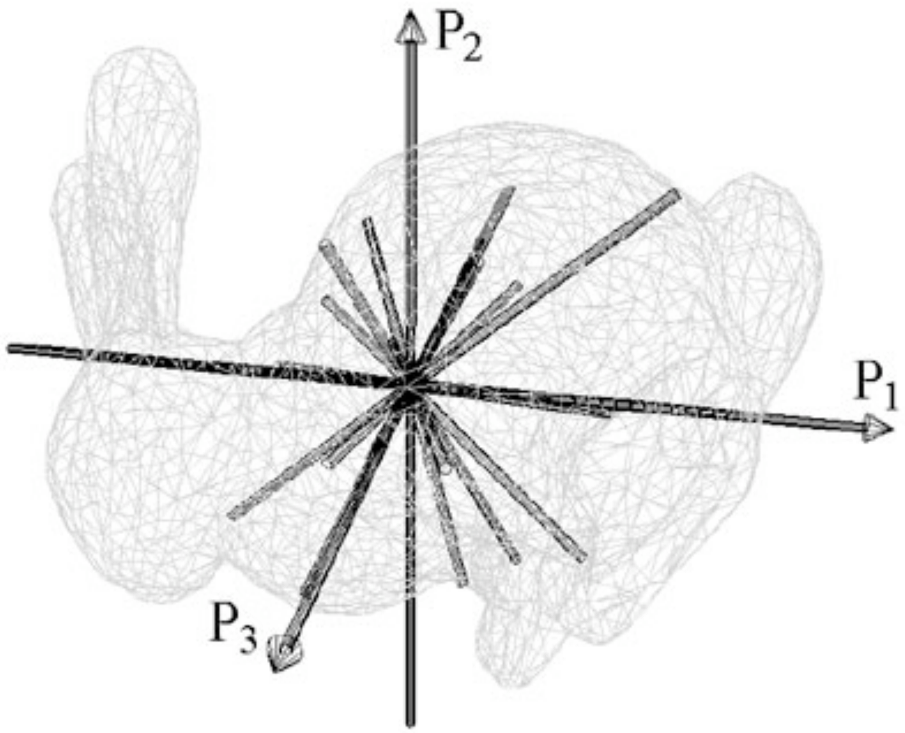

feature
extraction
insert
ε-search
NN-search
3D model
high dimensional
feature vector
high dimensional
index structure

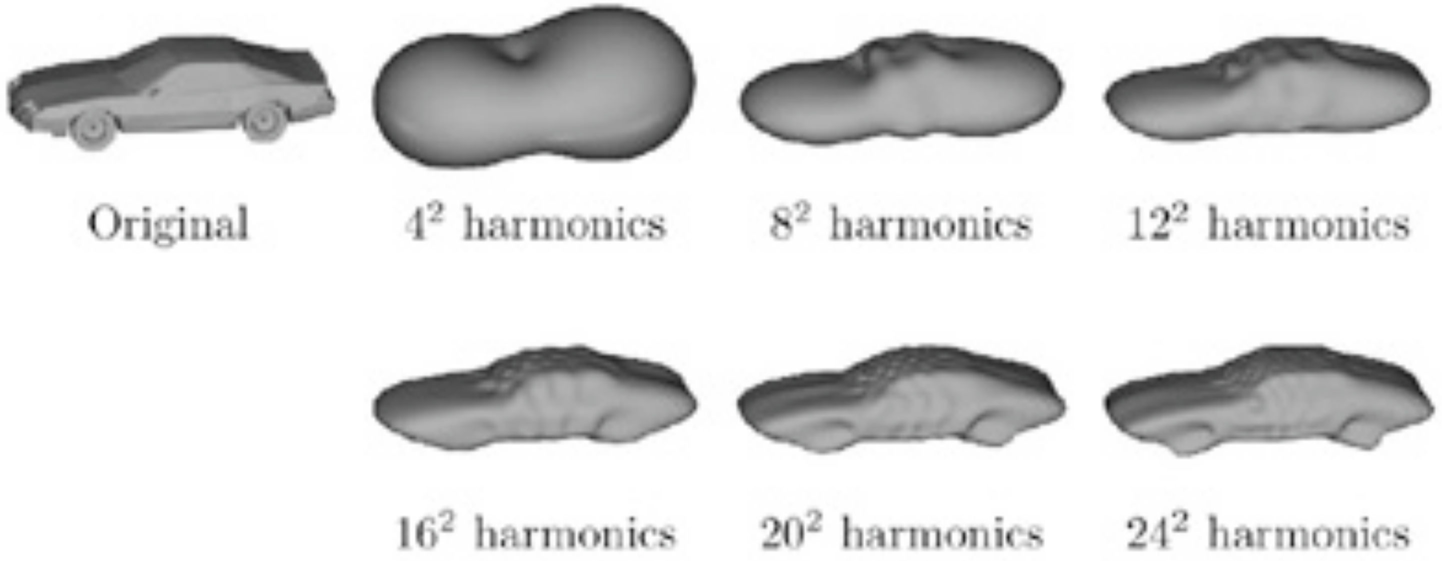
Feature vectors

- geometry based
- image based

Feature vectors

- Geometry based





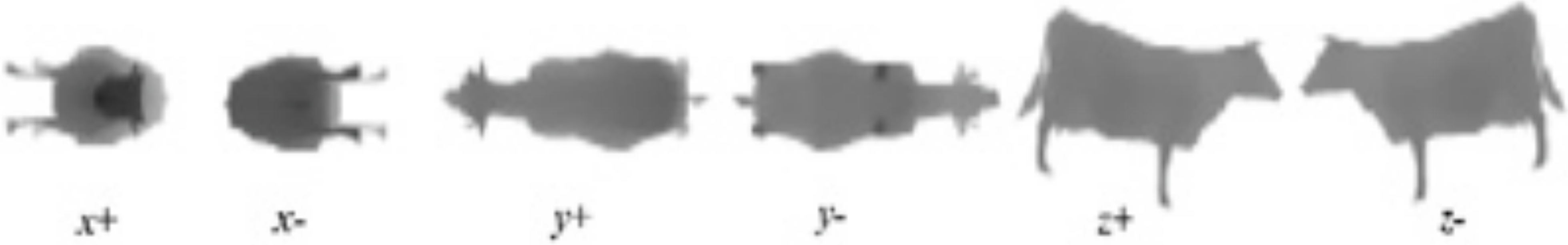
Ray-based scanning after principal axes transformation

Multi-resolution spherical harmonics representation

Feature vectors

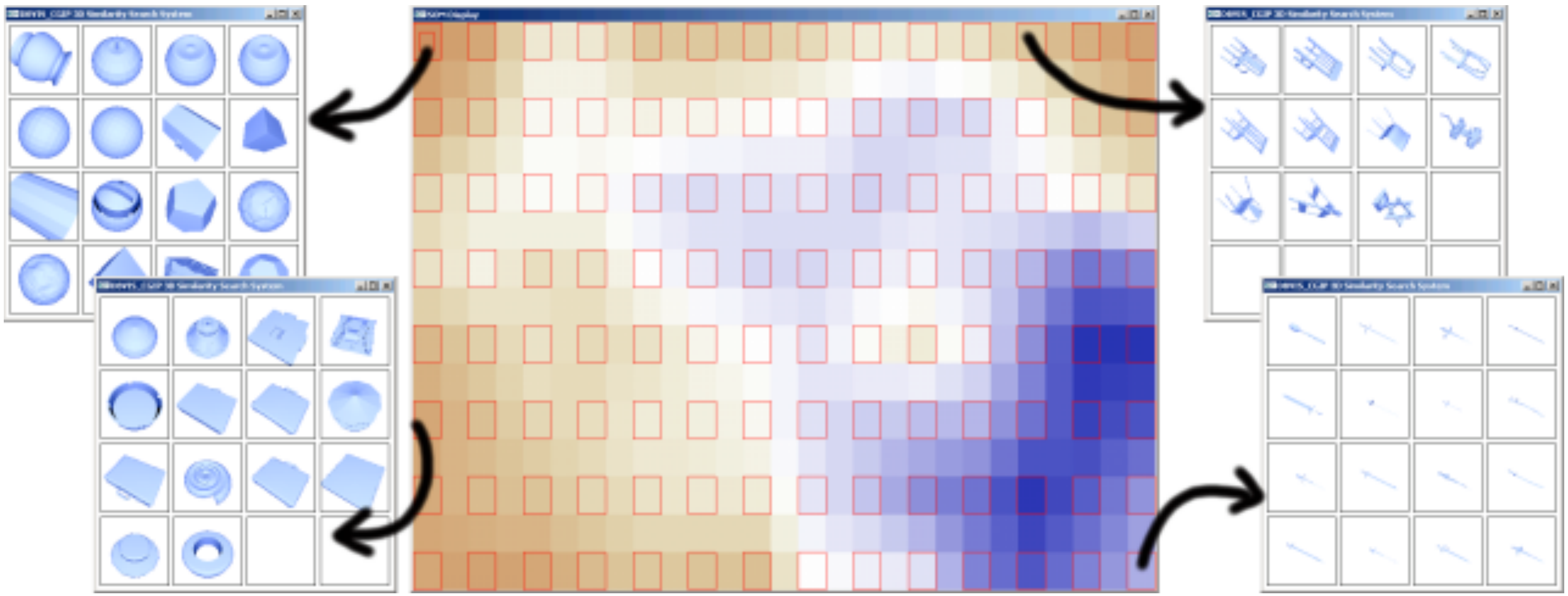
- Image based

Flat 2D silhouettes with Fourier coefficients



Depth buffer maps from 6 directions

What's good?



Self-organizing map of a 3D database