Video Indexing and Retrieval (VIR)

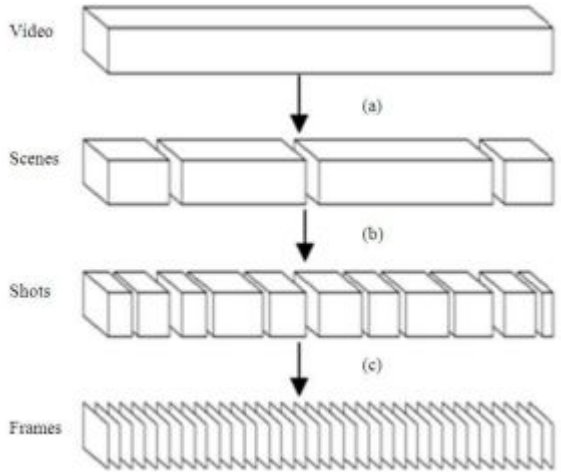
By Rupsi Kaushik

What is VIR?

VIR is the indexing and retrieval of video data from large databases

Video data is data in the form of text, audio, and images, incorporated with a time dimension

Video databases have a shot as a basic unit of data



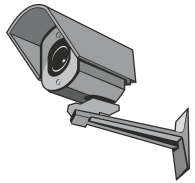
Keywords: video, scenes, shots, frames

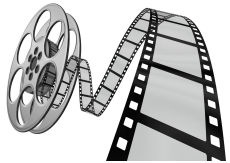
Motivation

- **Growth of multimedia data**
  Need for more **efficient tools** to classify and retrieve high volumes of video data
- **The complex nature of indexing and retrieving video data**
  Call for **better ways** to help machines evaluate challenging video data
- **The limitations in our current multimedia retrieval systems**
  Require more **practical solutions** than the ones we have now for managing and retrieving video data

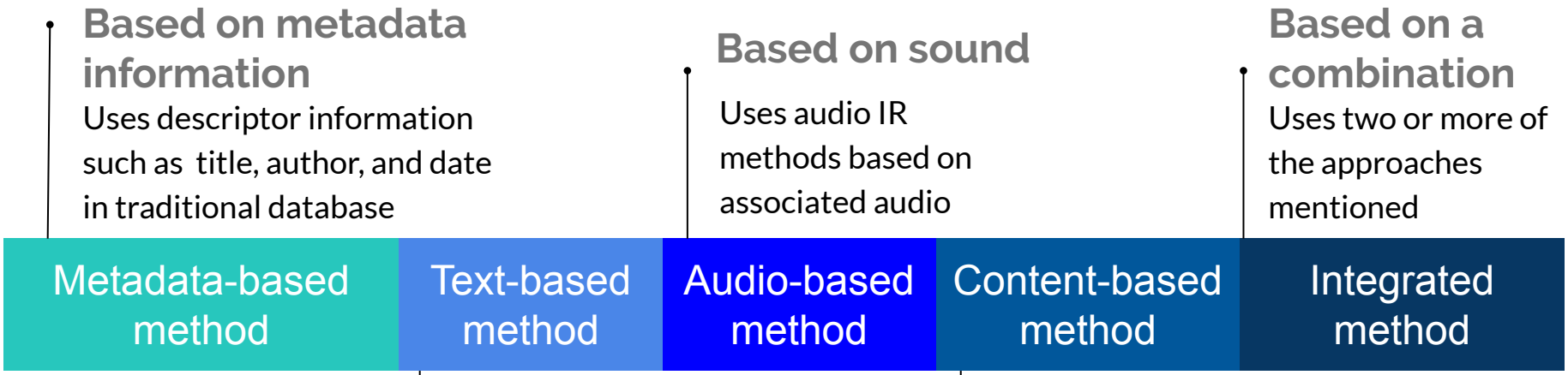
Applications of VIR

- ➔ News videos
- ➔ Surveillance videos
- ➔ Sports highlights
- ➔ Film archives

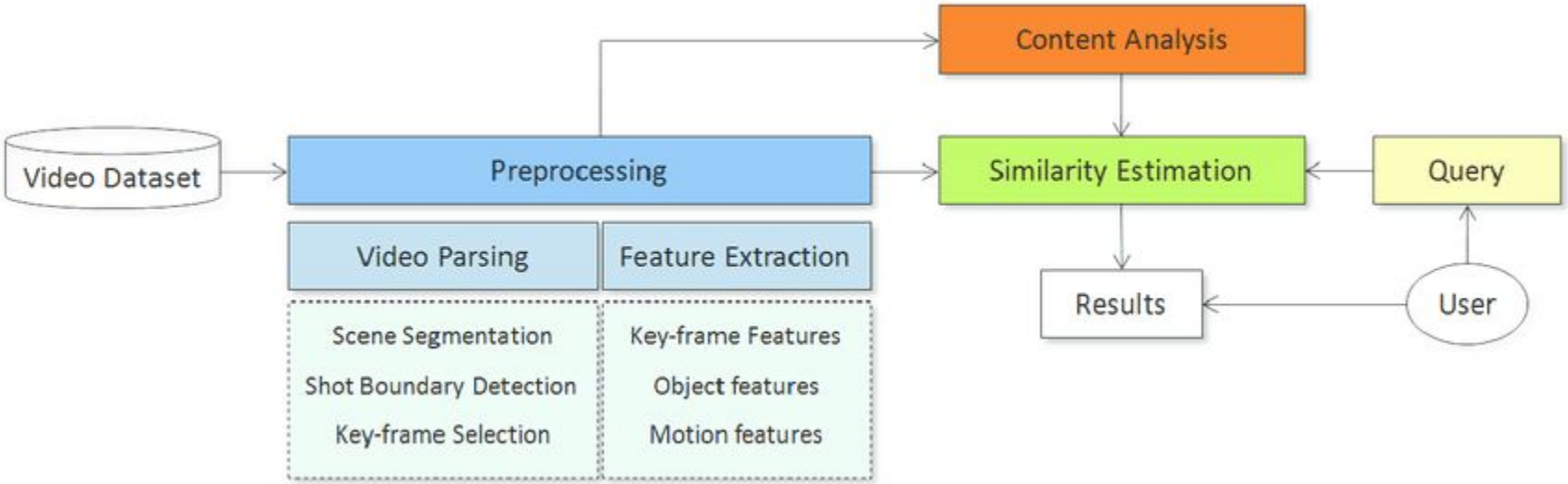


VIR Methods


Based on metadata information
Uses descriptor information such as title, author, and date in traditional database
Based on sound
Uses audio IR methods based on associated audio
Based on a combination
Uses two or more of the approaches mentioned
Metadata-based method
Text-based method
Audio-based method
Content-based method
Integrated method
Based on subtitles
Uses traditional IR techniques based on associated subtitles
Based on the media content
Uses the video content itself
Focuses on three main methods to VIR

Content-based Video Retrieval

Architecture



Shot-based video retrieval

Based on representative frames (r-frames)

Step 1: Video segmentation

➔ Decompose video into elementary shots and frames

Step 2: Index each shot

➔ Use an r-frame to represent each shot in order to create a shot index

➔ Image IR is used to extract features of key frames for information

Step 3: Apply a similarity measure

➔ Compute similarity between r-frames and user query and rank them

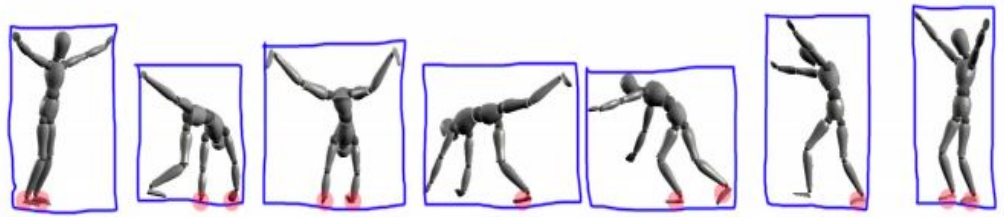
➔ The most relevant/ similar frames are displayed

Motion-based video retrieval

Based on motion parameters

Captures parameters from video such as:

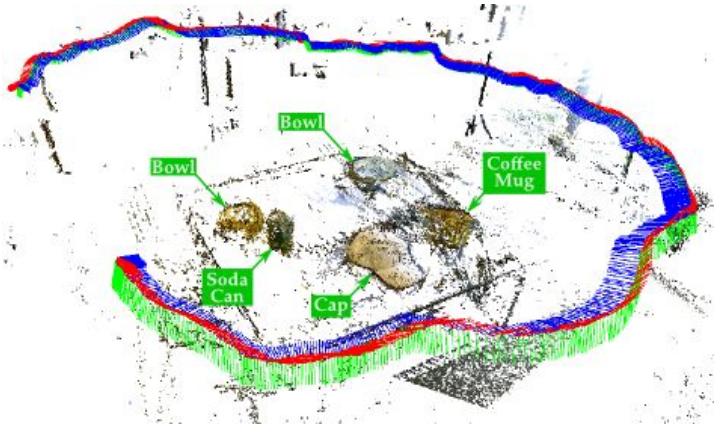
- ➔ Motion content
- ➔ Motion uniformity
- ➔ Motion panning
- ➔ Motion tilting



Object-based video retrieval

Based on object detection

Primarily concerned with detecting occurrences of objects, their interactions, and relationships among other objects in a video content



- ➔ Group pixels that move together as an object
- ➔ Use these segmented objects to index the video
- ➔ Continue with feature extraction

Comparative Analysis of VIR Methods

Shot-based video retrieval

- Simple computations
- Less computational cost to use keyframes to represent a shot
- Ignores the concept of motion
- Does not directly represent content

Motion-based video retrieval

- Extends the concept and benefits of r-frames
- Captures the motion information

Object-based video retrieval

- Aims to represent actual content
- Computationally complex to extract objects from videos
- Easier when video is compressed

Evaluating VIR systems

- ➔ Precision (P)
- ➔ Recall (R)
- ➔ Harmonic average values (F1 score)

$$P = \frac{Similar \cap Retrieved}{Retrieved} \qquad R = \frac{Similar \cap Retrieved}{Similar}$$

$$F1\ score = 2PR / (P + R)$$

Key Takeaways

1

Growing need of VIR in multimedia retrieval

VIR has an increasing importance in today's ever-growing multimedia repositories

2

Different approaches to video retrieval

We are able to apply metadata IR, text IR, annotative IR, image IR, and so on, to retrieve videos

3

3 main methods for content-based retrieval

Representative frames, motion information, and object-based retrieval are the most common methods for content based video retrieval

4

Precision & recall can be used for evaluation

Precision, recall, and their harmonic average values can be used to evaluate a VIR system