Video Indexing and Retrieval summary

Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 2.6.1

Artificial Intelligence & Information Analysis Lab

Video Indexing and Retrieval

- Hierarchical video structure
- Shot cut/transition detection
- Video Summarization
- Audiovisual content description
- Video indexing and retrieval

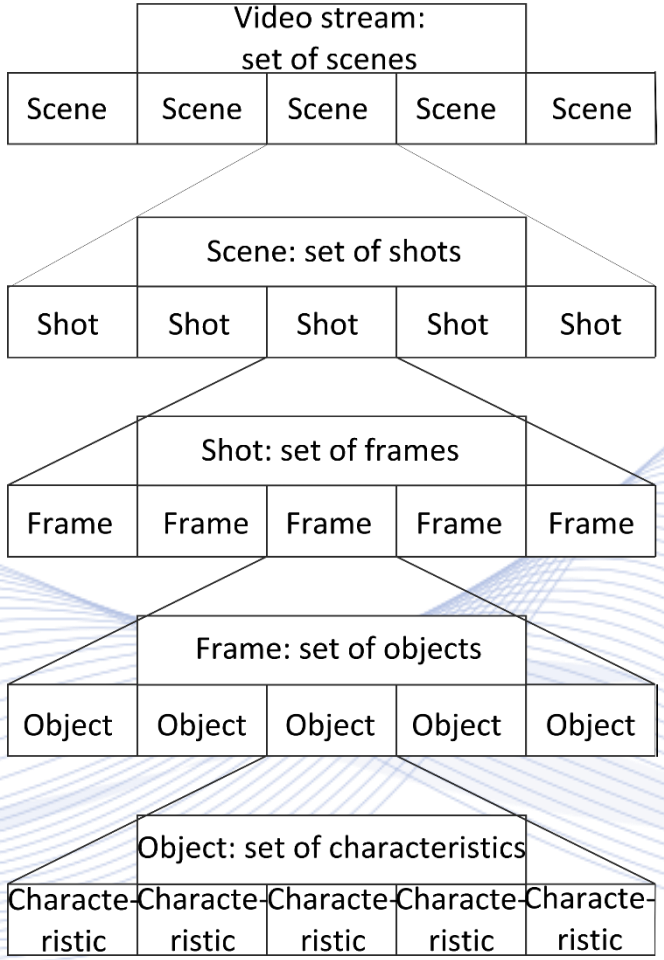
Introduction

- Video content search:
  - in a digital video file;
  - in broadcasting archives;
  - in social media sites (e.g., YouTube).
- Content-based video search and retrieval is difficult:
  - Too big content size;
  - Unstructured video content
  - Time consuming video browsing.

Introduction

- Many techniques have been proposed for content-based video indexing and retrieval:
  - Shot cut/transition detection
  - Video Summarization
  - Video key-frame selection
  - Audiovisual content description
  - Video indexing and retrieval.
- Low-level and semantic (content-based) video retrieval techniques.

Hierarchical video structure



Hierarchical video segmentation.

Hierarchical video structure

- A video (e.g., a movie) consists of a sequence of scenes.
- A **video scene** is a sequence of video shots focusing on an object or objects or story of interest.
- A **video shot** is a single sequence of frames which are captured by a stationary or continuously moving camera.
  - A movie which contains alternating views of two persons consists of multiple shots.

Shot cut and transition detection

There are various types of **shot transitions**:

Abrupt shot transitions:

- A **shot cut** is an abrupt shot change.
- Abrupt changes are easier to detect, compared to gradual ones.

Gradual shot transitions:

- A **fade-in/fade-out** is a slow change in shot luminance, which usually leads to, or starts with, a black frame.

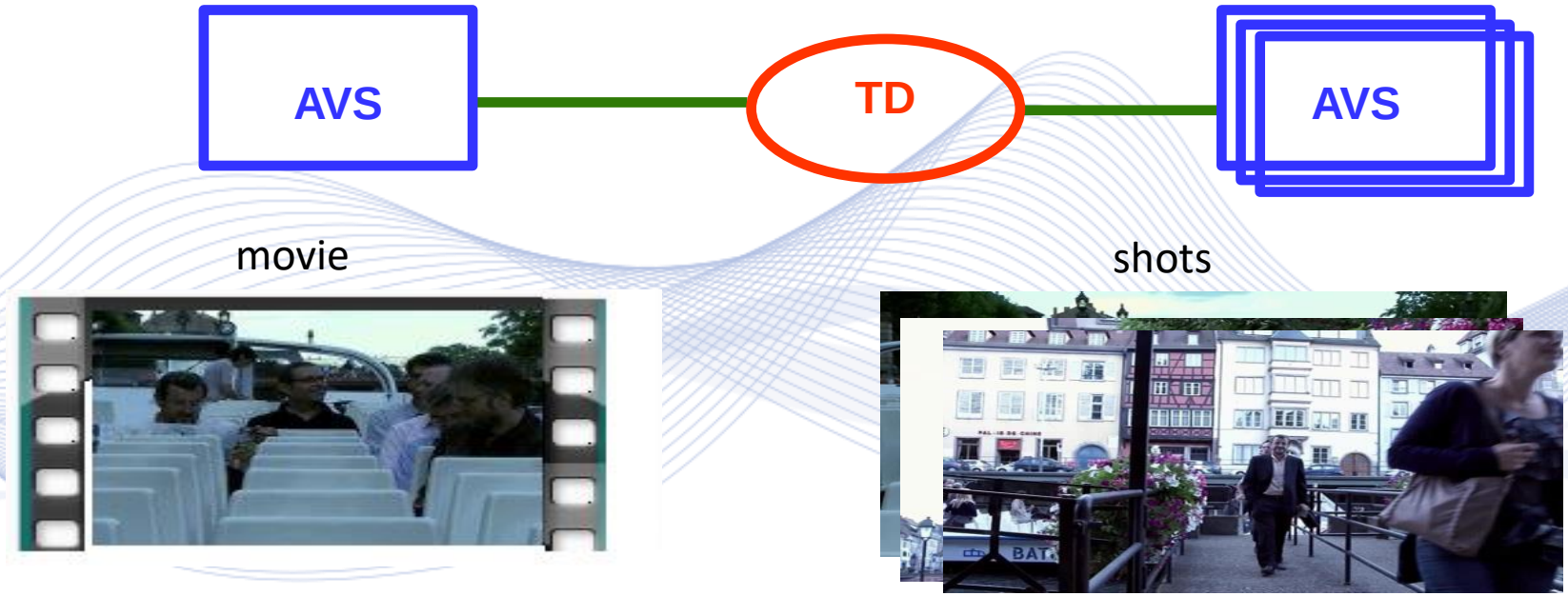
Shot cut and transition detection

- A **dissolve** takes place when there is a spatial overlay of the frames of the two shots for the duration of the transition. The luminance of the images of the first shot decreases and that of the second shot increases.
- A **wipe** occurs when the pixels of the second shot gradually replace those of the first shot with a local motion, e.g., from left to right.

- Many other gradual shot transition types are possible.

Shot cut and transition detection

Temporal Decomposition of a video into shots.



Shot cut detection

Let video frames f, f' have luminance vectors $\mathbf{Y}, \mathbf{Y}'$.

- The simplest **distance metric** between two consecutive shots is given by:

$$D(f, f') = \|\mathbf{Y} - \mathbf{Y}'\|.$$

- Various error norms, e.g., the L_1 or L_2 ones can be used.
- This shot cut detection method has a limited success rate and can detect only 73% of the actual shot changes.

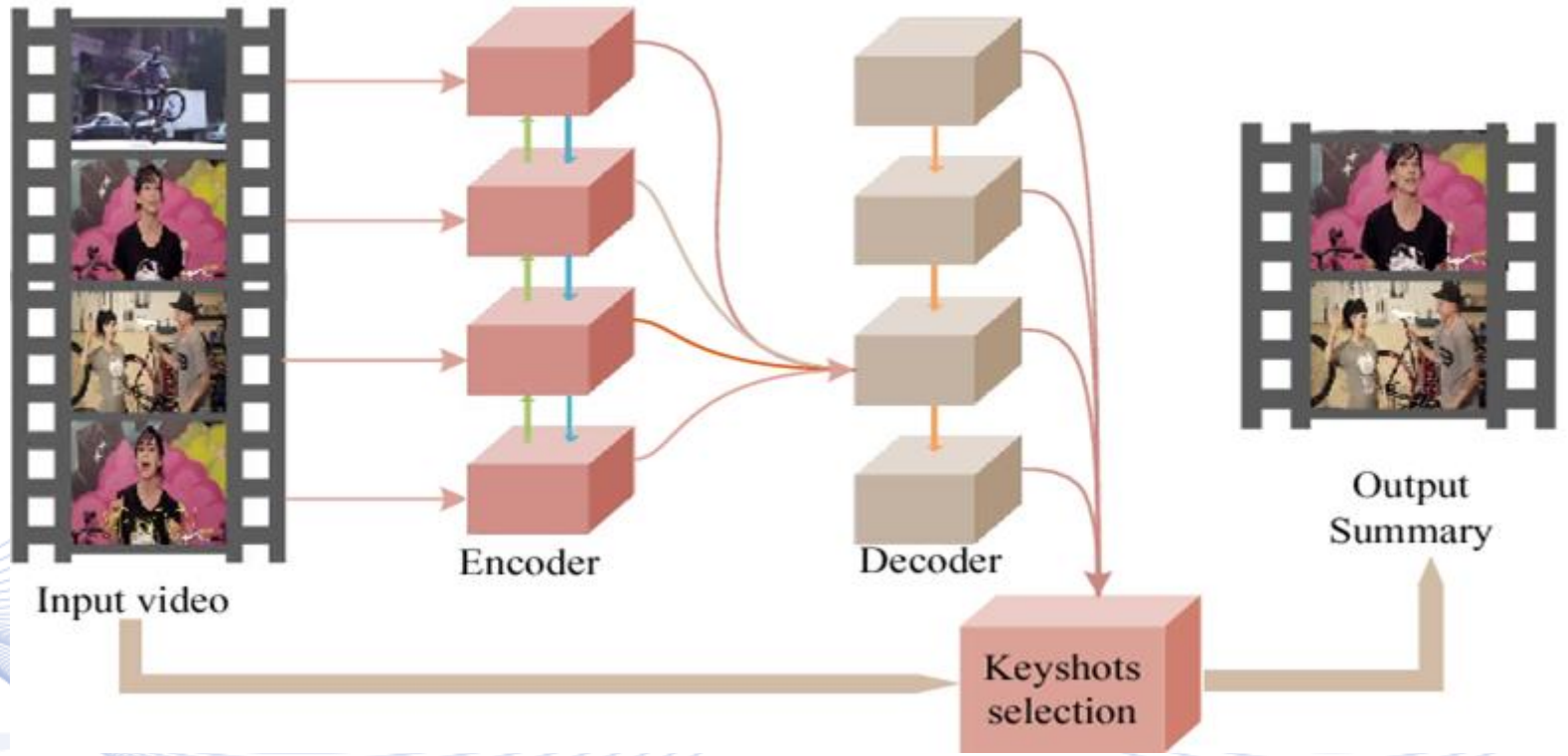
- False detections occur in case of object or camera motion.

Key frame selection and video summarization

- Consider we want to extract a number of *key frames*, able to summarize well the video content, for video description and fast browsing [Shan98], in a long video sequence. Their number may vary from 5% to 10% of the total frame number in the original video.

- There is no mathematical model, which defines the exact requirements for key frame selection. Many techniques are based on shot cut detection, while other approaches employ the visual content and motion analysis.

Key frame selection and video summarization



(Picture from CatalyzeX)

Object based shot description

- Many techniques have been proposed in the literature for video summarization and retrieval based on object detection.
- The objects can be detected using spatial features, such as their color, texture or shape.
- In a video, there are two sources of information which can be used for object detection and tracing : visual characteristics and motion information.

Object based shot description

- A typical strategy is to initially perform region segmentation based on color, texture and shape information.

- After the initial segmentation, regions with similar motion vectors can be merged based on certain limitations, such as region adjacency [Aslandogan99].

Object based shot description

The use of the adaptive K-means (C -means) algorithm is used in [Kompatsiaris01] for connected region (object) detection.

- A specific number of consecutive video frames are processed in the *CIE* color space.
- The technique starts with the segmentation of each frame in K subdivisions, using color histograms.

Multimodal audiovisual content description

The combination of audio and image features to extract more semantic features was proposed in [Adami01]:

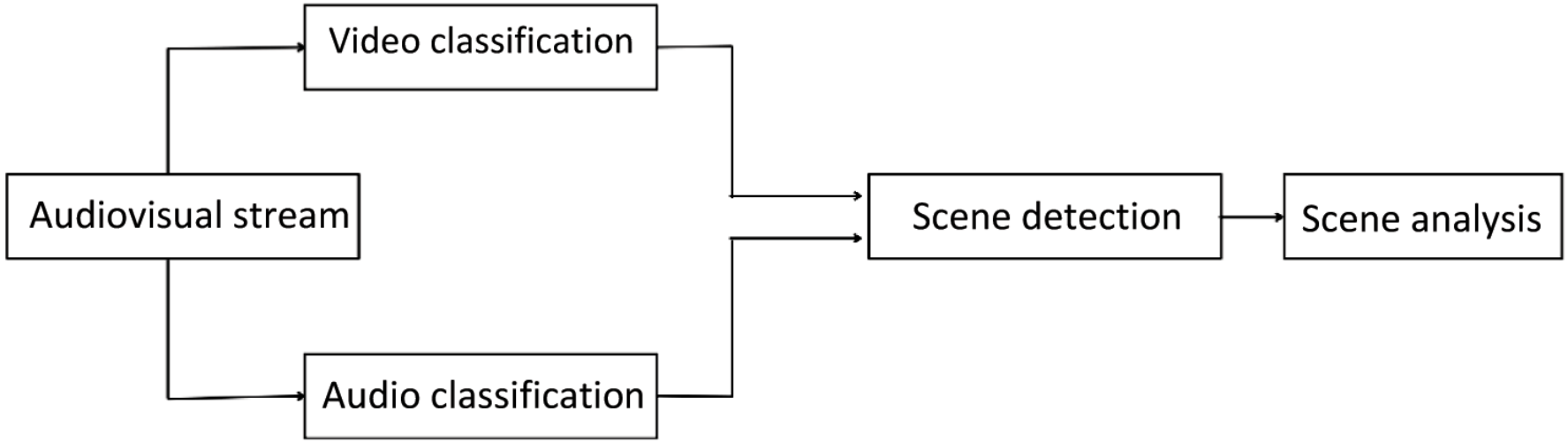
- The mean audio intensity is used as a measure of shot significance
- For audio streams, the audio features are extracted from low-level audio properties.
- For video streams, the visual features are extracted using motion estimation with luminance histograms and pixel differences.

Multimodal audiovisual content description

- Each sequence of features extracted from both audio and video channels is used for the identification of video semantics.

- In case of scenes containing humans, four different shot types can be identified: dialogue, monologue, action and generic video.

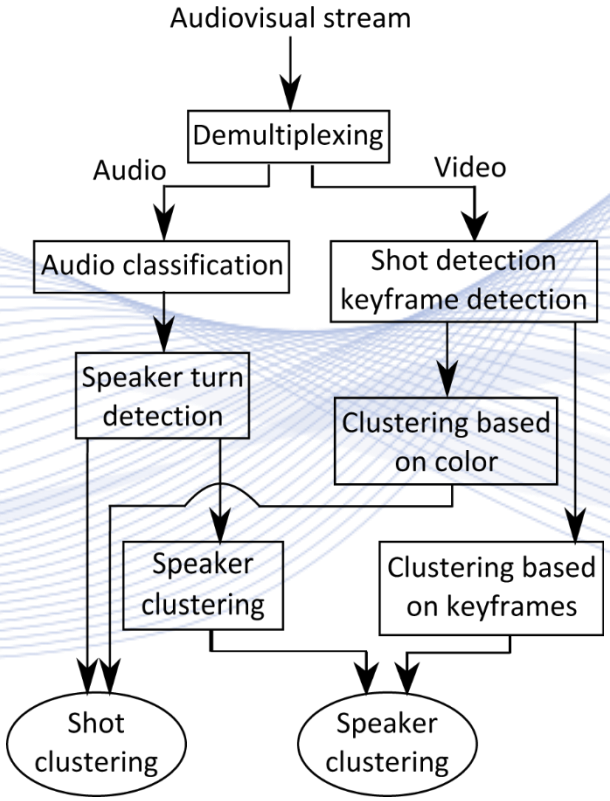
Multimodal audiovisual content description



Combination of audio and video features for the description of audiovisual content.

Multimodal audiovisual content description

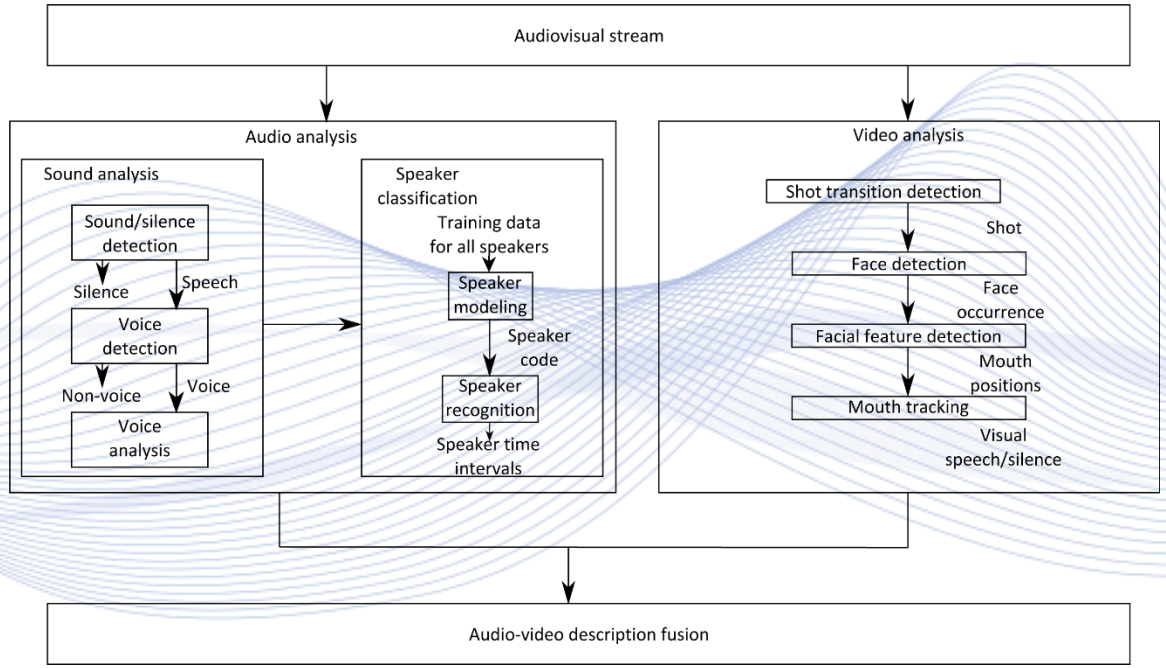
- News video description techniques, using audio have been proposed in [WQI00] and have been further improved with the addition of text processing in [Jiang00]. Figure 3 shows the diagram of the proposed method.



Audio assisted news video processing method.

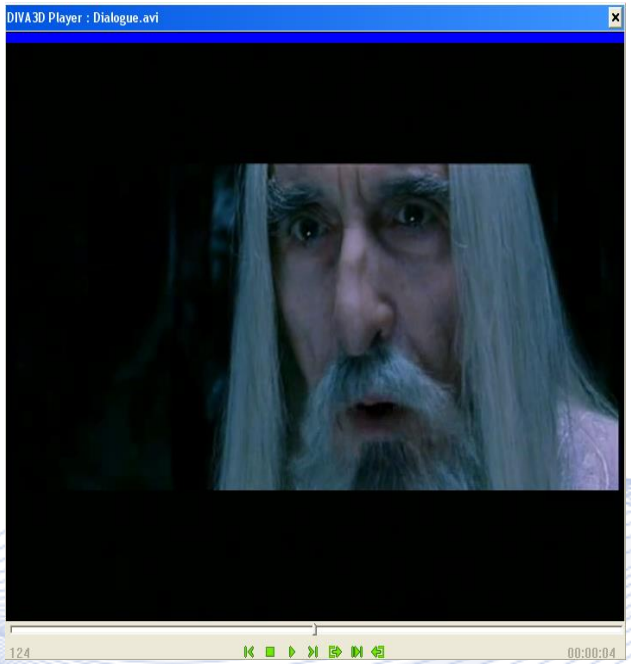
Multimodal audiovisual content description

- A method involving both visual and audio content for video indexing was presented in [TSE99], [TSE01]. The block diagram of this method is shown in Figure 4.



Audiovisual content analysis.

Multimodal audiovisual content description



Detection of: a) dialogs in a movie and b) monologues in news broadcasting.

Semiautomatic video description and search approaches

- Semiautomatic approaches for content-based video retrieval were proposed in [LIU00] and [OH00].

Most videos are annotated manually. Different people may have different semantic interpretations for the same video scene.

- To eliminate this confusion, an adaptive and flexible automatic approach was proposed [LIU00].

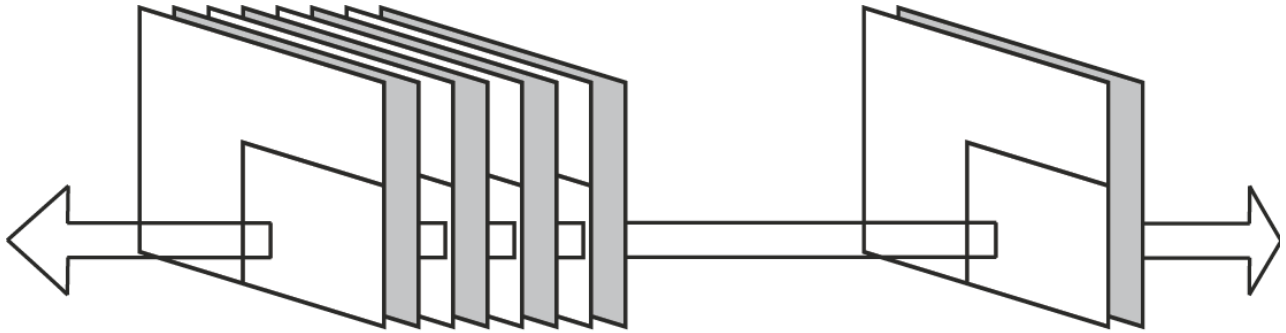
Indexing Techniques

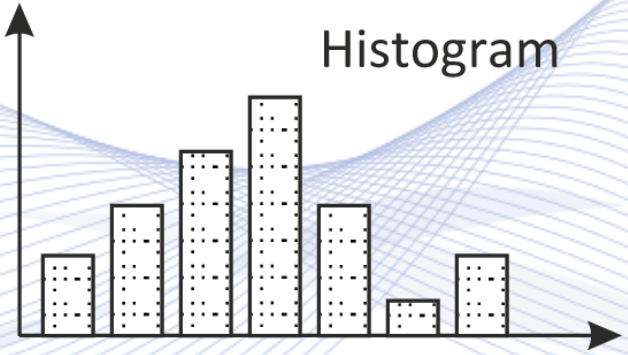
- In the previous sections, techniques were presented for analyzing video streams and extracting semantic information.

- The next step is to present techniques for creating the appropriate indexing structure of the acquired information to facilitate video retrieval. *Hashing* is a widely known technique for data indexing [KNU73].

- An approach for video image (key-frame) indexing based on the edge directions in predefined image regions has been presented in [MOT00].

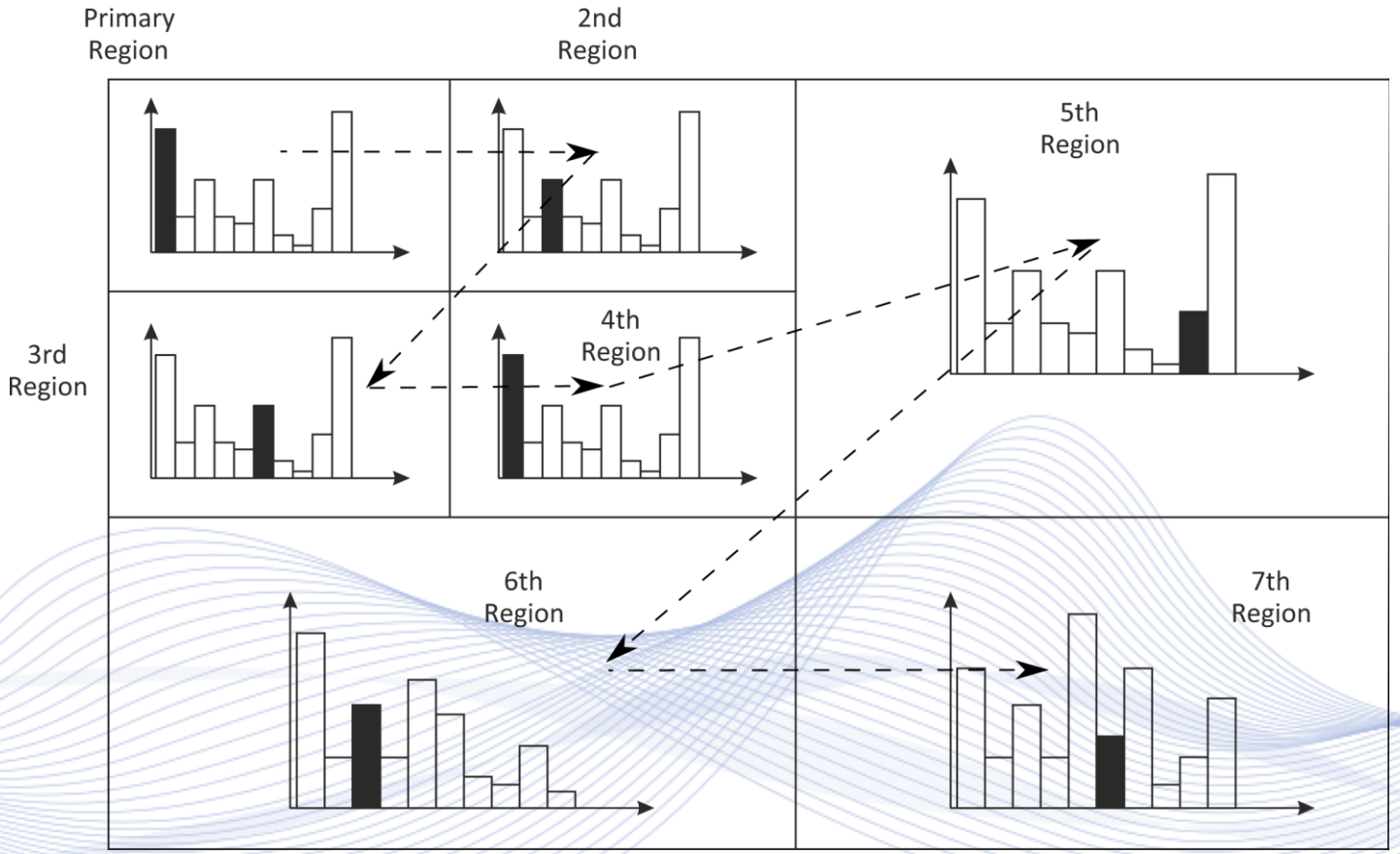
Indexing Techniques





Calculation of the energy histogram in a wavelet band.

Indexing Techniques



Use of the band energy histogram.

Bibliography

[PIT2017] I. Pitas, “Digital video processing and analysis” , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, “Digital Video and Television” , Createspace/Amazon, 2013.
[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.
[NIK2000] N. Nikolaidis and I. Pitas, “3D Image Processing Algorithms”, J. Wiley, 2000.
[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr