

Video Description summary

A. Bolari Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitasp@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.4.1

Video Description

- **MPEG-7 description Standard**
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

Semantic Video Content Analysis & Description



- Video content can be described through state and state transitions of individuals, interactions between humans and physical characteristics of humans:
 - Human presence (face, head, body),
 - Activity/gesture,
 - Facial expressions,
 - Face/body pose,
 - Number of people in a video shot,
 - Which people are in the shot (face recognition)
- Anthropocentric (human-centered) approach: humans are the most important video entity.
- Characteristics and behavior of other foreground entities such as objects can be also used for semantic description.

MPEG-7 content description standard



- Multimedia content semantic analysis results should be stored in a standardized, consistent & structured way.
- Reasons: exchange of information, ingestion in 3D content databases/archives/MAM systems.
- MPEG-7 standard defines a description framework for handling multimedia content annotation and description.

Video Description

- MPEG-7 description Standard
 - **Parts**
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

MPEG-7 parts

The MPEG-7 (ISO/IEC 15938) consists of different Parts. Each part covers a certain aspect of the whole specification.

Part Number	Title
1	Systems
2	Description Definition Language
3	Visual Description Tools
4	Audio Description Tools
5	Multimedia Description Schemes
6	Reference Software
7	Compatibility

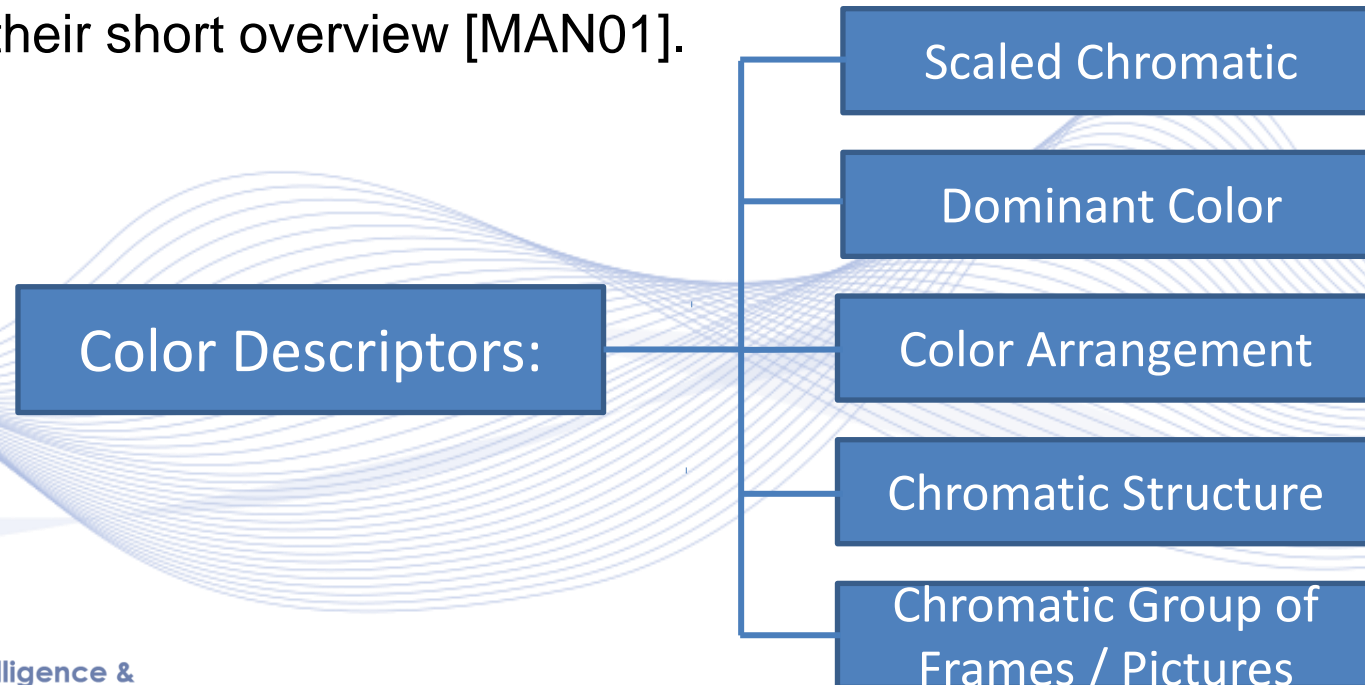
Table 1 : MPEG-7 standard parts

Video Description

- MPEG-7 description Standard
 - Parts
 - **Descriptors**
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

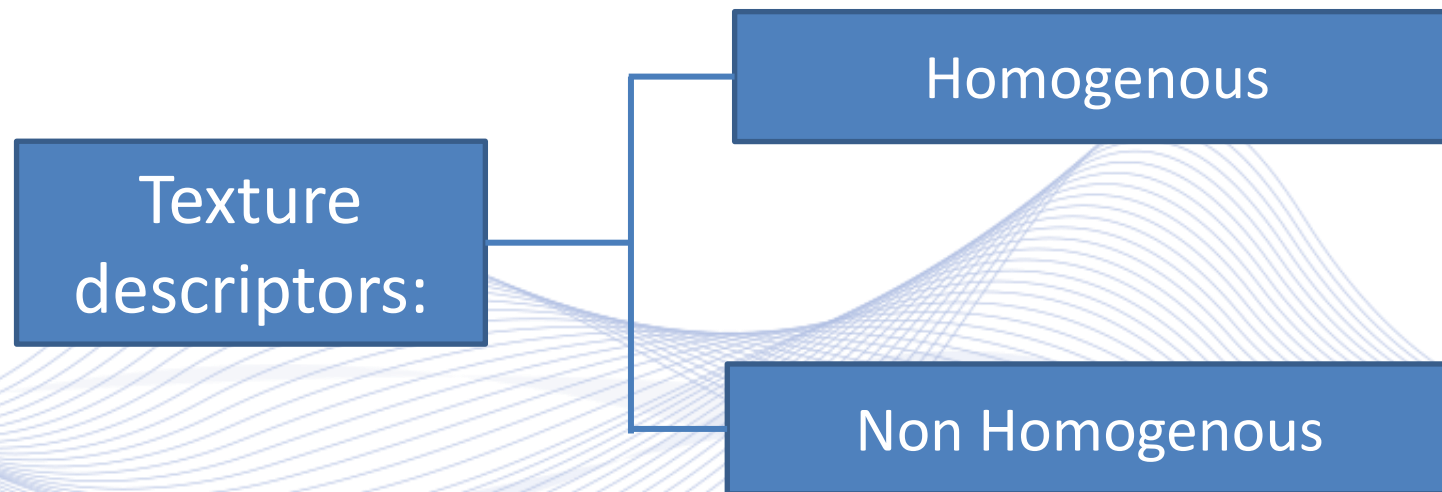
Color Visual Descriptors

- A particular effort has been made in the design of efficient color descriptors and the detection of similar images. There is no generic color descriptor that can be used for all applications. Therefore many generic descriptors were standardized, each one being appropriate for a special visual similarity identification function. Below, we will see their short overview [MAN01].

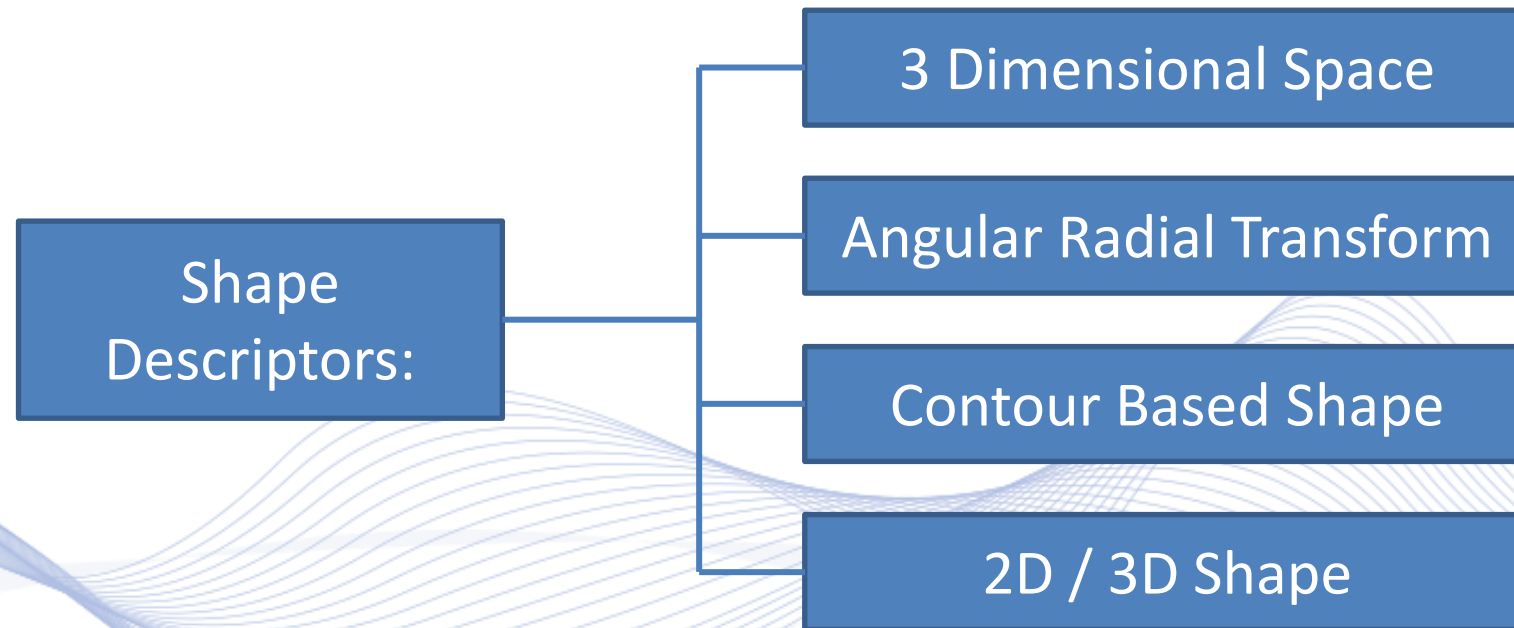


Texture Visual Descriptors

- MPEG-7 has defined appropriate texture descriptors, which can be used in various applications.

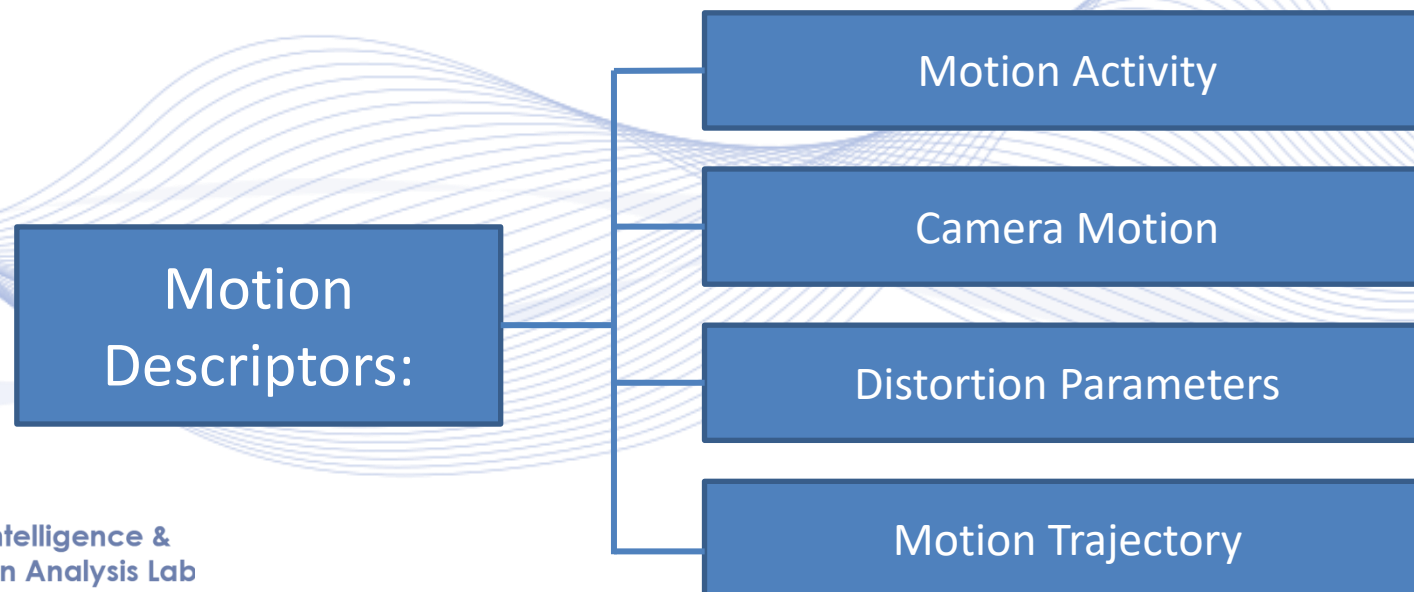


Shape Visual Descriptors



Motion Descriptors in video

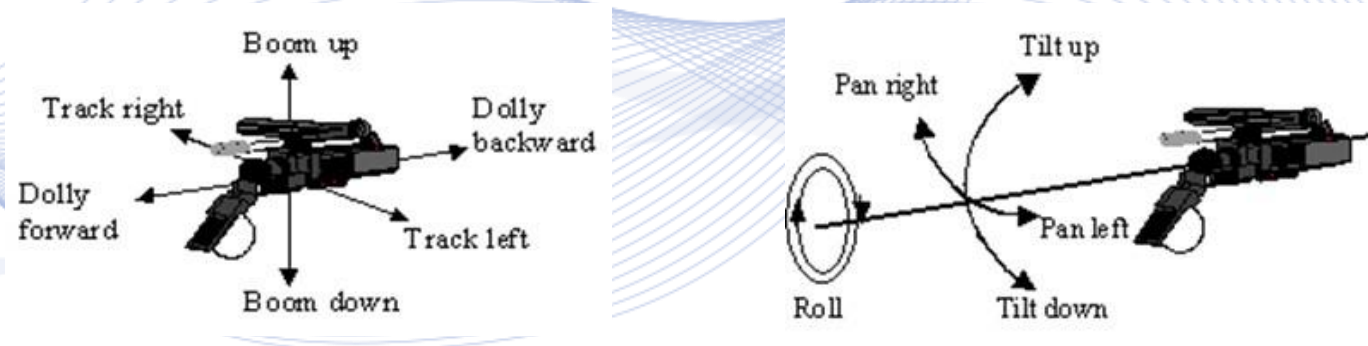
- Motion description in a video sequence by can be particularly heavy in terms of bits per video frame, even if the motion fields are sparse.
- MPEG-7 has developed descriptors which capture basic motion characteristics from the motion field in concise and effective descriptions.



Motion Descriptors in video

- **Camera motion descriptor:**

- Describes the motion of the physical or virtual camera.
- Yields details about the kind of total motion parameters which exist at a given time instance in a shot.
- Used for the search of video sequences based on certain total motion parameters.
- Search which allows matching motion similarities in specific time intervals.



Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - **Audiovisual Description Profile (AVDP)**
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

Audio-Visual Description Profile (AVDP)

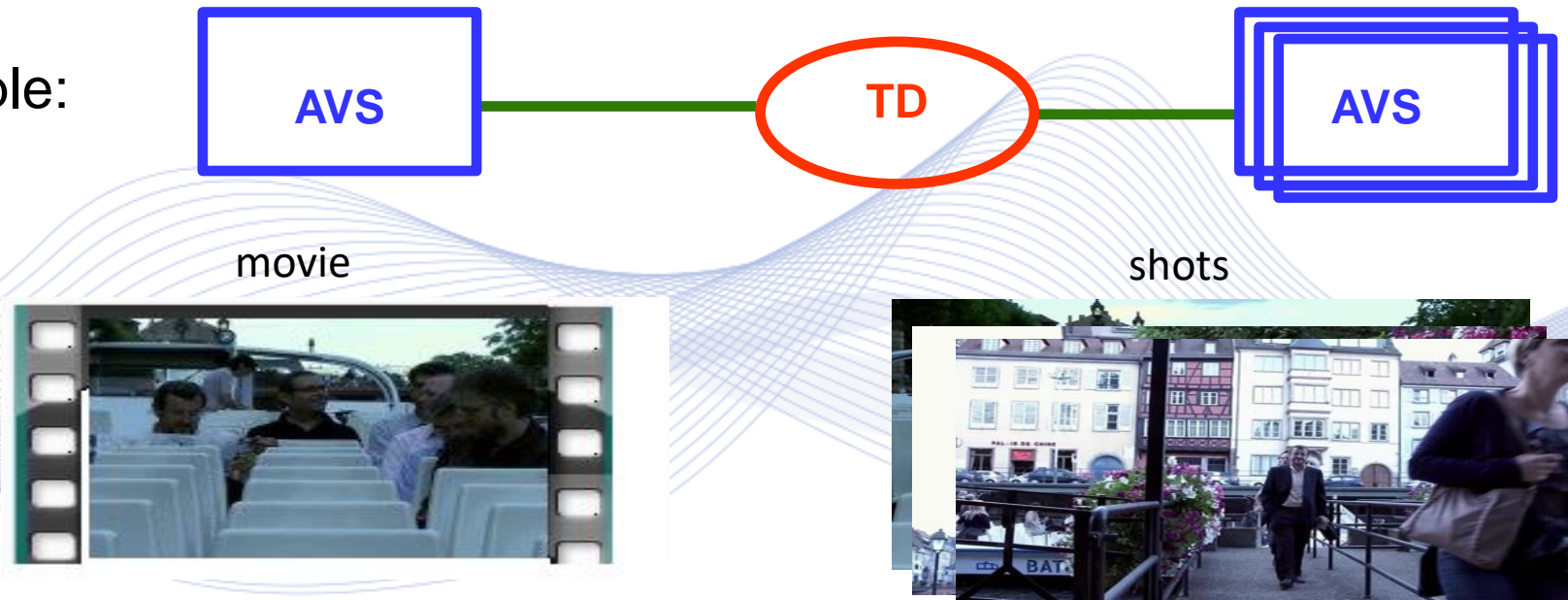


- Audio-Visual Description Profile (AVDP) is an MPEG-7 profile introduced to **describe results** of multimedia analysis algorithms [AVD14]:
 - Person identification, genre detection, keyframe extraction, speech recognition etc.
 - Several results can be described in multiple timelines.
- Descriptions are in XML format, following the AVDP Schema Definition (XSD).

Useful AVDP/MPEG-7 Tools

- **TemporalDecomposition type (TD):** decomposes a VideoSegment type (VS), AudioVisualSegment type (AVS), or AudioSegment type into temporal segments.

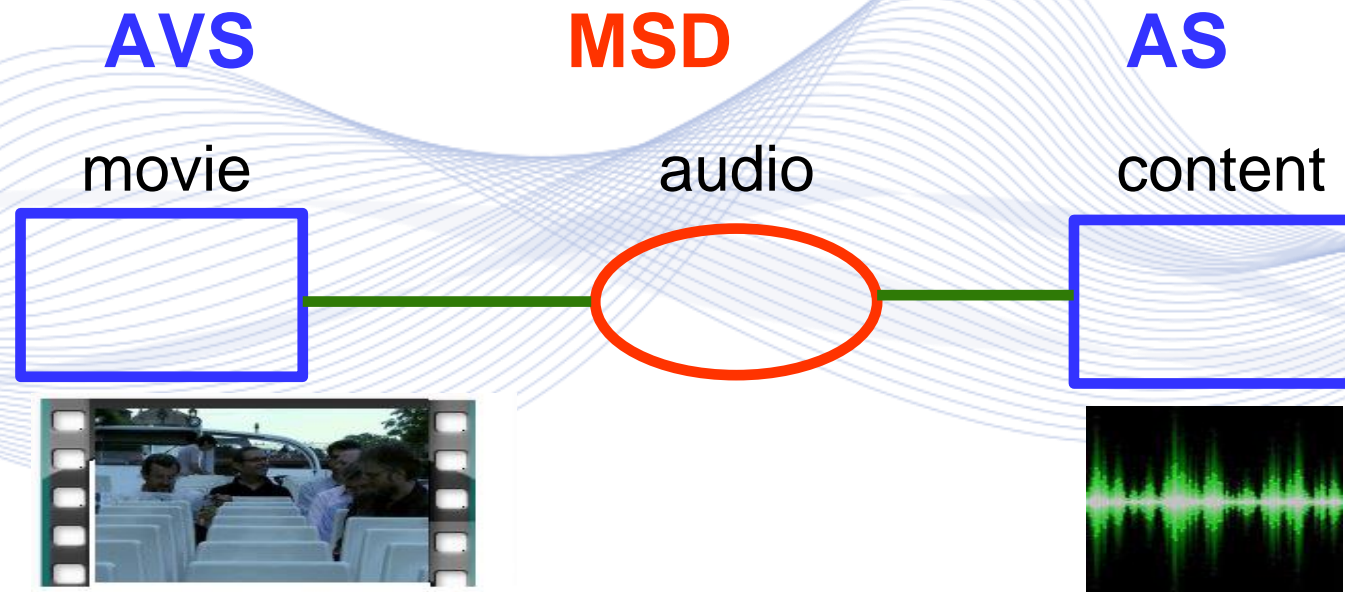
- Example:



Useful AVDP/MPEG-7 Tools

- **MediaSourceDecomposition type (MSD):** Decomposes AudioVisualSegment into different segment types:
 - VideoSegment, AudioSegment,
 - StillRegion (SR): spatial regions within a frame (including entire frames)

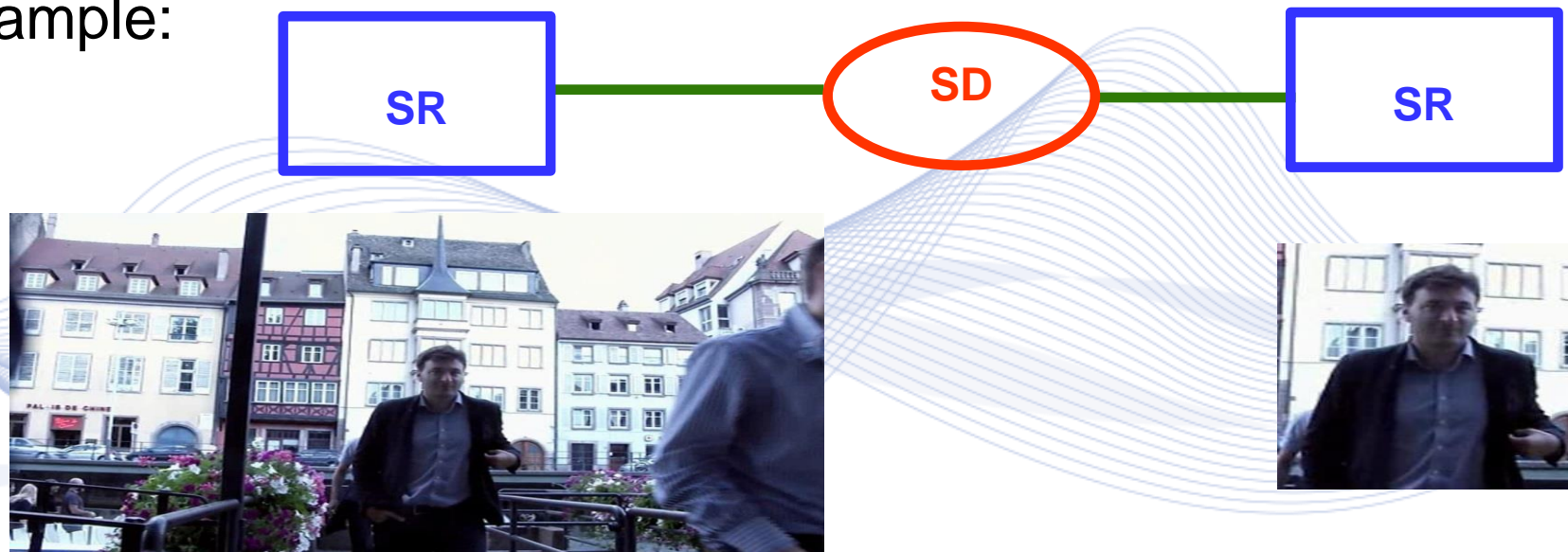
- Example:



Useful AVDP/MPEG-7 Tools

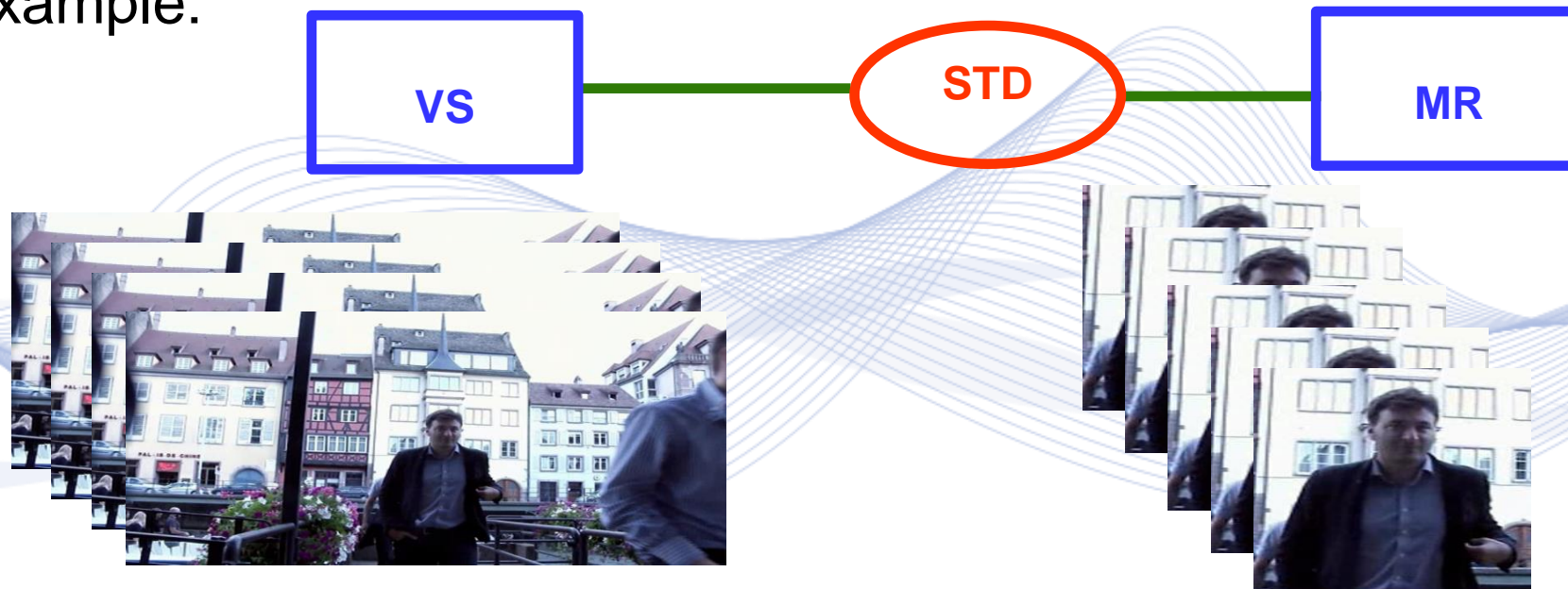
- **SpatialDecomposition type (SD):** Decomposes a frame into spatial regions (StillRegion - SR).

- Example:

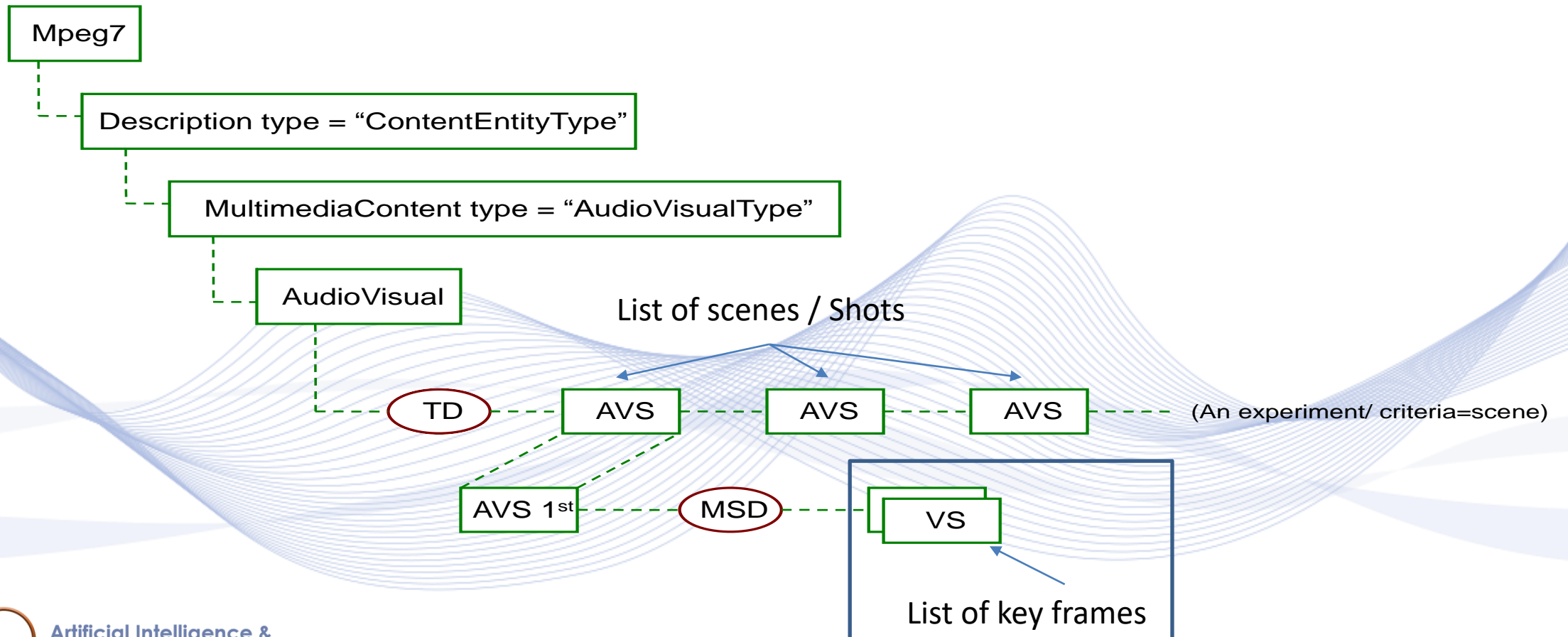


Useful AVDP/MPEG-7 Tools

- **SpatioTemporalDecomposition type (STD):** Decomposes a VideoSegment into MovingRegions (MR) corresponding to trajectories of spatial regions over time.
- Example:

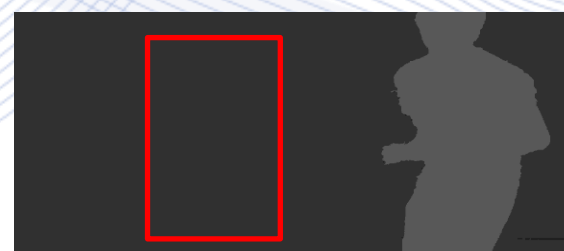


Scene/Shot Boundaries Detection

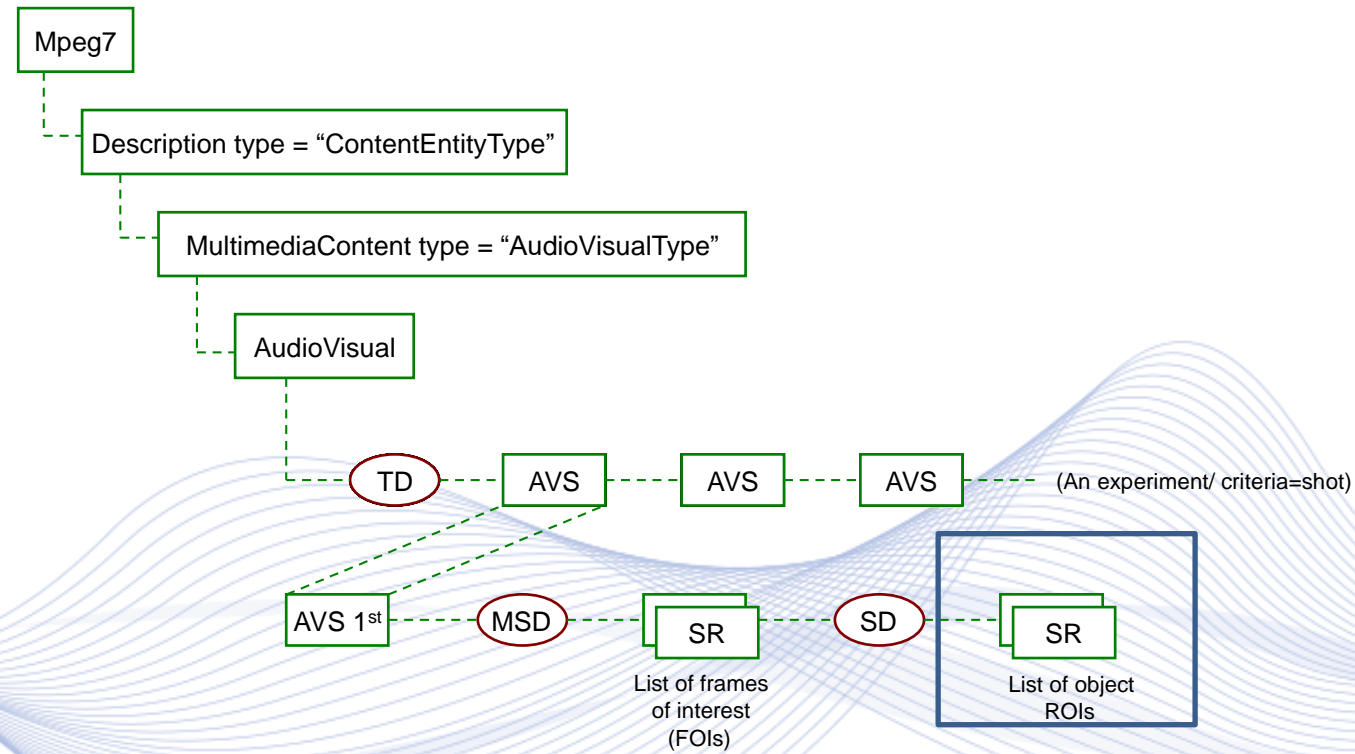


Human/Object Detection

- **Human/object detection:** localizing this entity in a frame.
- Generates a bounding box that includes the detected entity.
- Results are stored in StillRegion types (at least) in the channel(s) where detection takes place.



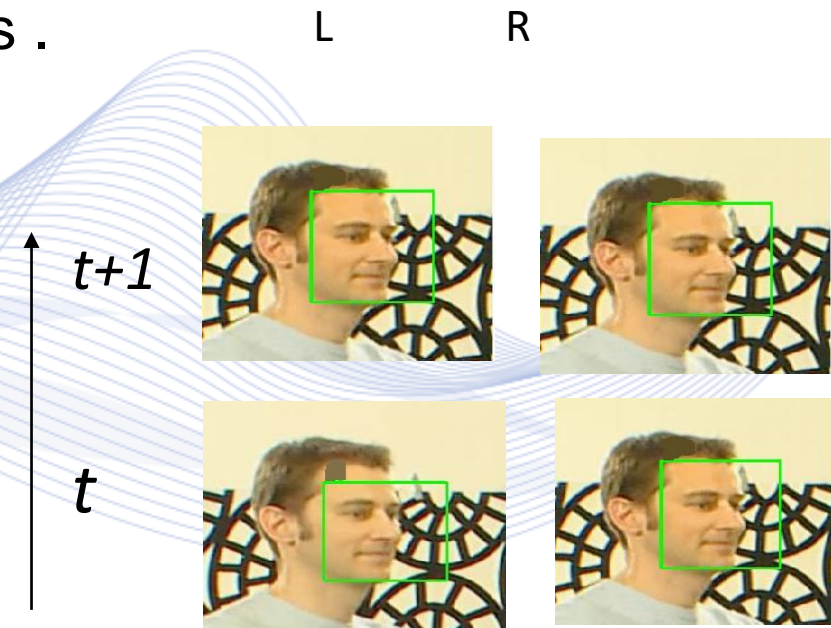
Human/Object Detection



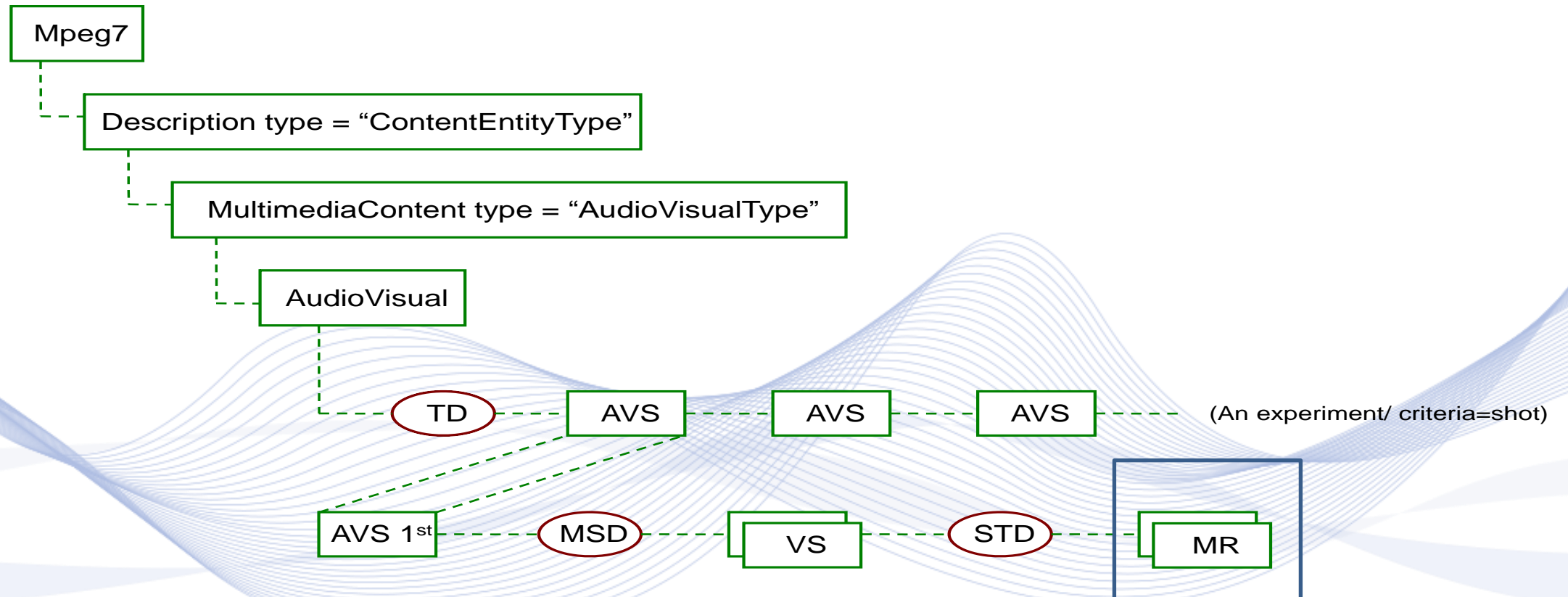
Human/Object Tracking

- **Human/object tracking:** following a detected entity's location in time.
- Generates a series of bounding boxes over time.
- 3D tracking: in both channels, possibly taking into account disparity
- Results are stored in MovingRegion types .

3D



Human/Object Tracking



Video Description

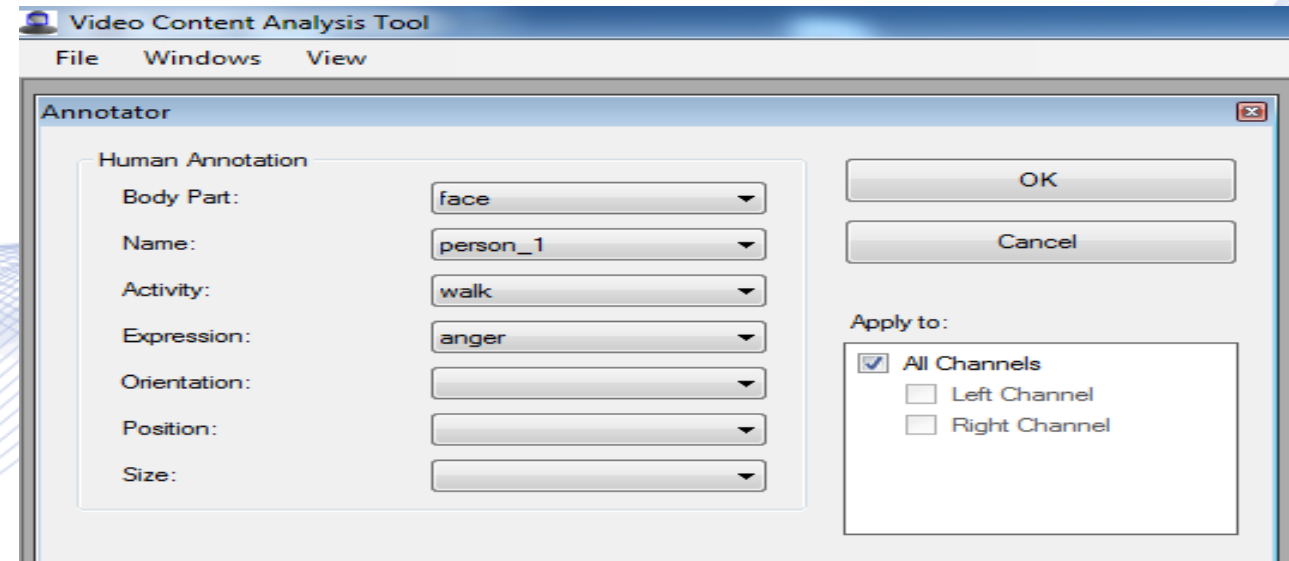
- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - **Video analysis / annotation software**
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

3D video analysis / annotation software

- **3DVideoAnnotator software platform assists a user in the tasks of:**
 - Analyzing and annotating 3D video content.
 - Viewing and editing the relevant description.
 - Saving/reading annotations to/from an MPEG7/AVDP file.

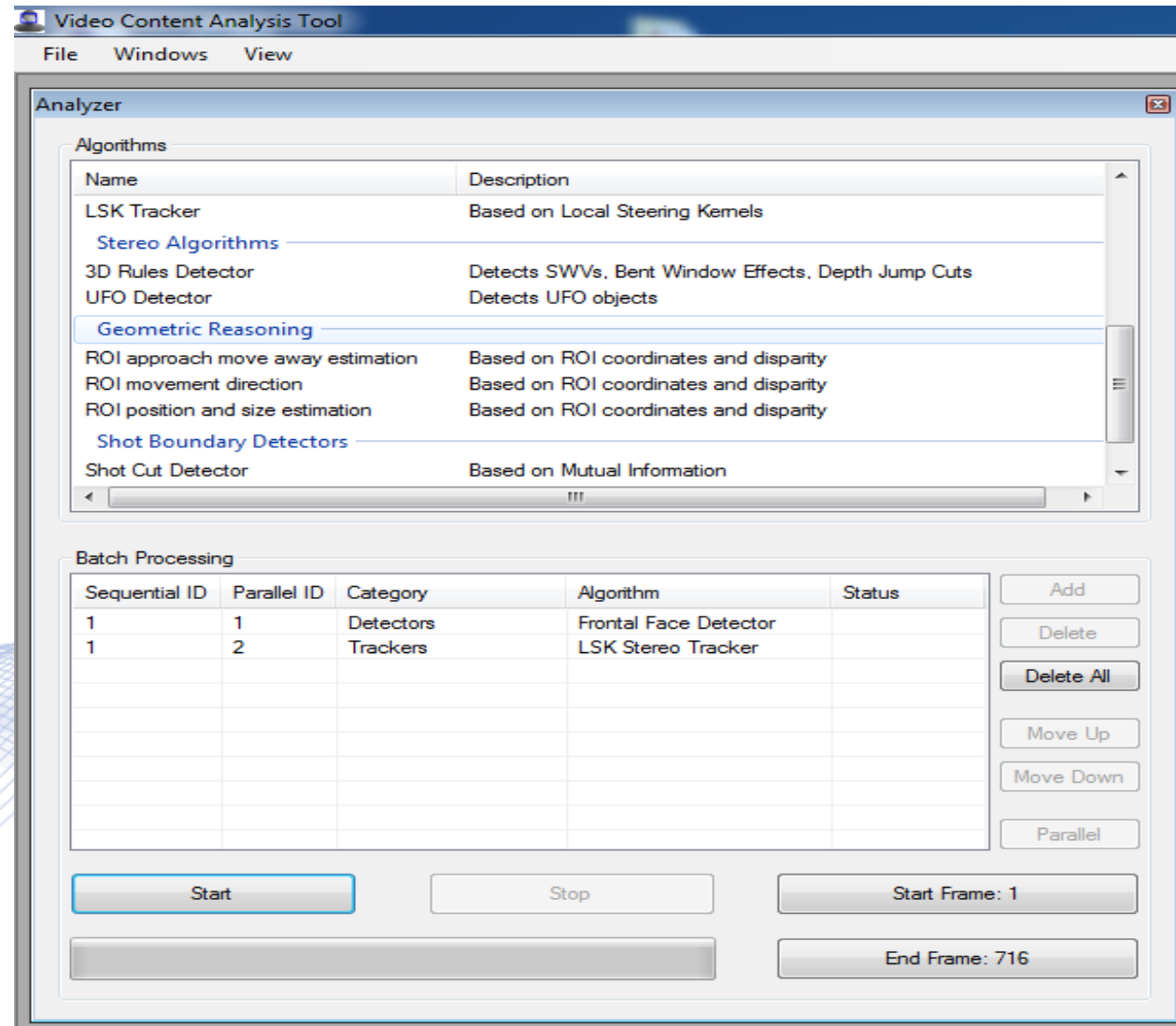
Annotator

- Manual content annotation
- User can annotate:
 - static or moving objects or humans
 - Trajectory, activity, identity etc.
 - shots/transitions
 - key segments
 - events
 - ...



Analyzer

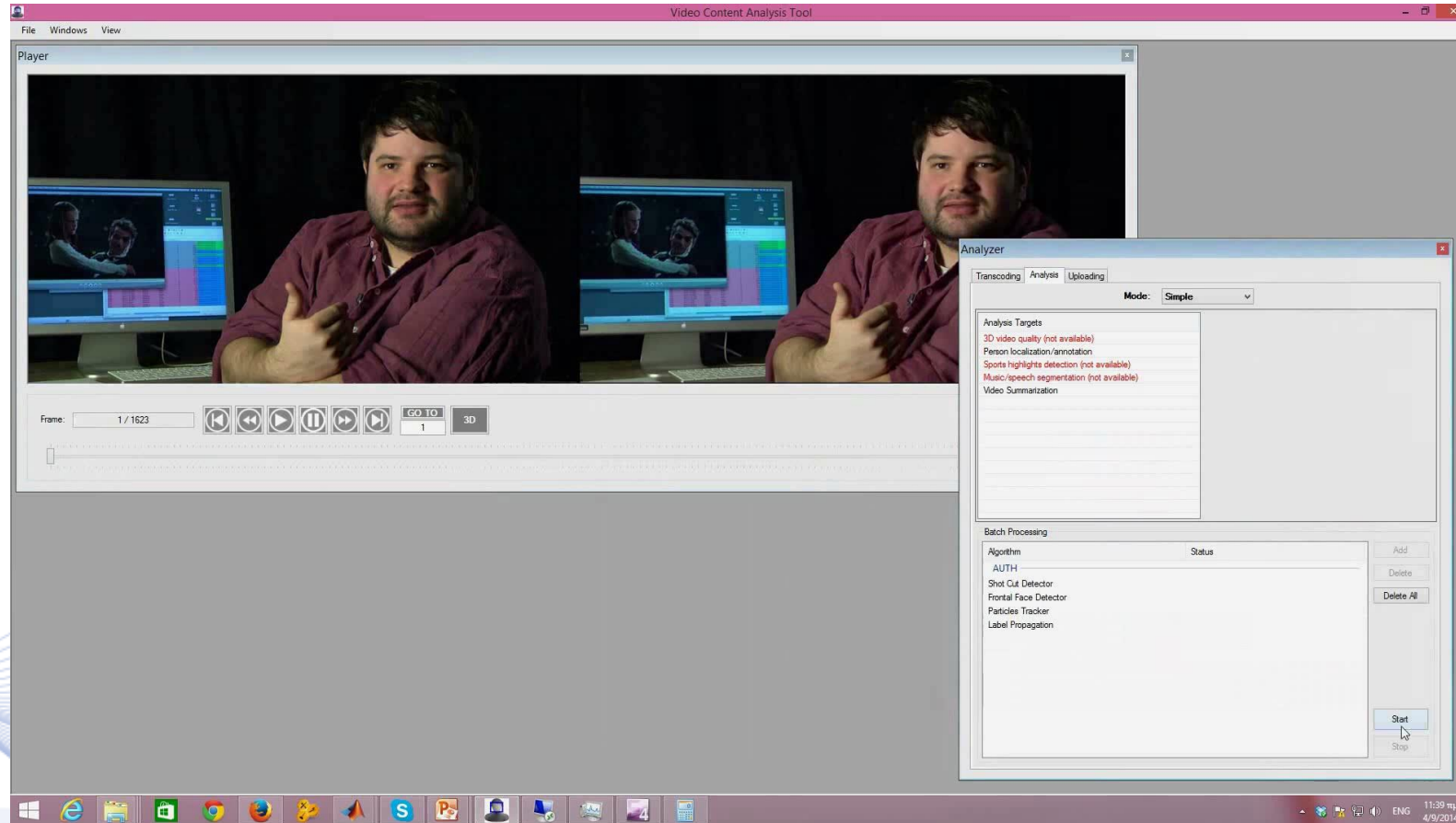
- Enables the user to execute various video analysis algorithms, e.g. face/body detection and tracking.
- Supports importing new algorithms through .dll files.



Analyzer

- **Video analysis algorithms incorporated in the tool:**
 - Shot cut detection.
 - Shot type characterization.
 - Key frame and key segment selection.
 - 3D quality defects (stereoscopic window violation, bent window effect, depth jump cut) detection.
 - Stereo tracking.
 - Person action classification.
 - Object(s)/person(s) characterization in terms of position/movement/size.
 - Label/name propagation.

Analyzer



Editor

- Provides an easy way to navigate and edit the video content AVDP-based description/annotation.
 - The left part displays the description in a structured, hierarchical tree view form structured.
 - Through the right part the user can see and edit the description.

Editor



Video Content Analysis Tool

File Windows View

Editor

Description

- Header
- Left Channel
 - Shots
 - Shot_25
 - Key Segments
 - KeyVideoSegment_18
 - Shot_26
 - Key Segments
 - KeyVideoSegment_19
 - Moving Humans
 - MovingObject_49
 - MovingObject_51
 - MovingObject_53
 - Shot_28
 - Key Segments
 - KeyVideoSegment_22
 - Moving Humans
 - MovingObject_65
 - MovingObject_67
 - Shot_30
 - Events
 - Event_8
 - Key Segments
 - KeyVideoSegment_26
 - Moving Humans
 - MovingObject_77
 - Shot_32
 - Events
 - Event_9
 - Key Segments

Moving Human Description

ID: MovingObject_49

Start Frame: 189

End Frame: 828

Body Part: face

Name: person_1

Activity: sit

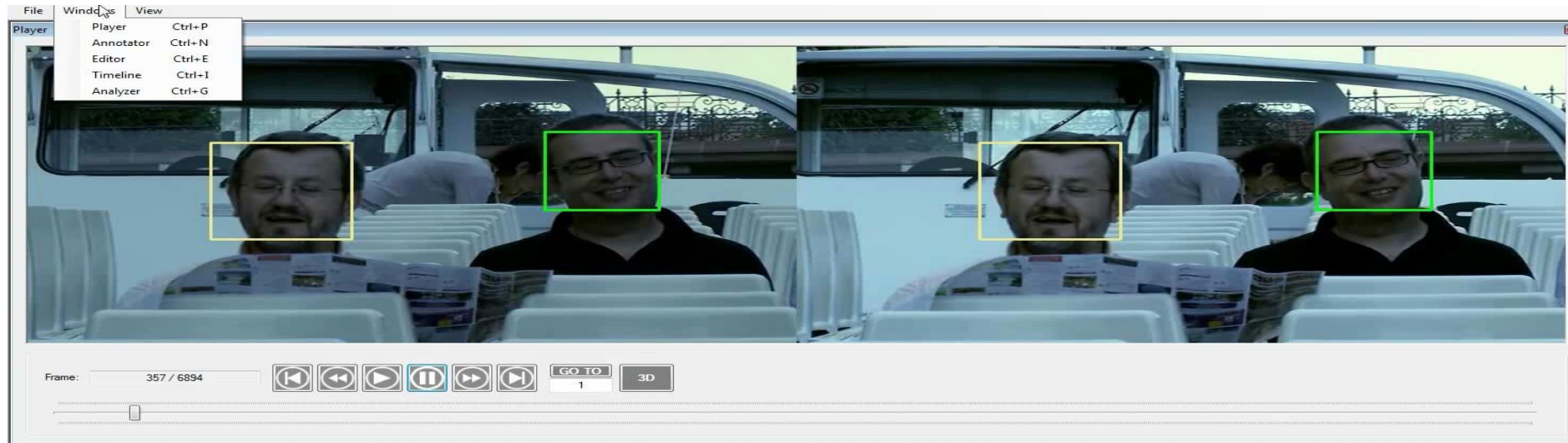
Expression:

Movement:

Sub-Activity	Sub-Expression	Sub-Movement	Related Movement
	Start Frame	End Frame	Expression
	654	812	sadness

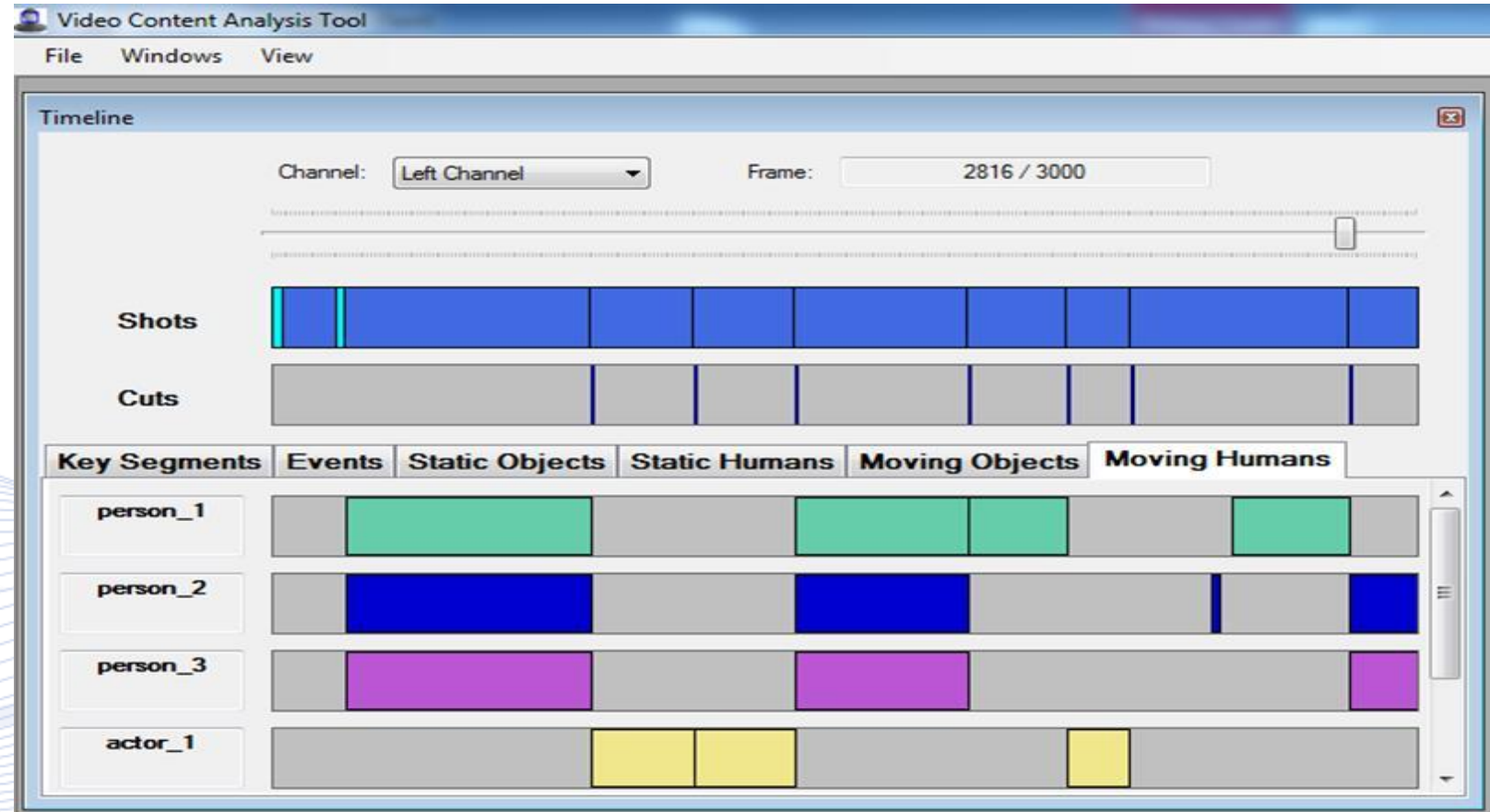
Apply to all channels

Editor



Timeline

- Shots & transitions
- Persons/objects appearances
- Events
- ...



Timeline



Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- **EBUCore description Standard**
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

EBU Core Metadata Set

- Framework for descriptive and technical metadata.[EBUCore_Schema]
- Designed to describe audio, video and other resources for broadcasting applications in the context of a Service Oriented Architecture.
- Facilitates program exchanges between broadcasters or production facilities in distributed cloud environments.
- Is defined by an XML schema but EBUCore metadata instances can easily be converted to JSON.

EBU Core Metadata Set

- Known as the “Dublin Core for Media”.
- Based on well-defined requirements and developer feedback [EVA14].
- Designed to be a metadata specification for users with different needs.
- Used to describe business objects like:
 - Tv & radio programs
 - Clips
 - Series
 - Documents
 - Pictures
 - Locations
 - Events

Where is EBUCore?

- Broadcasters e.g. ERT (Greece), RAI (Italy), France-television (France) and RIA (Ireland),
- Libraries e.g. library of Wales, Singapore,
- The achieves of Eurovision Song Contest.

Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- **PBCore description Standard**
- Standards and Usages
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

PBCore Metadata Set

- Created by the public broadcasting community in 2005 in USA for use by public broadcasters, libraries and archives.
- Set of specified fields used in database applications, designed to describe media, both digital and analog.
- Extends Dublin Core, adds elements that describe audiovisual assets.
- PBCore's primary interest is in data exchange and interoperability, not necessarily in creating a complete metadata model that can be exploited by digital asset management systems for comprehensive, original cataloging and markup of essence.

Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- **Standards and Usages**
- YouTube API & CNN Archive Metadata
- Media Asset Management Systems (MAMs)

Standards and Usages

- Today, thanks to the abundance of audiovisual content in the Internet, the actual use of digital video depends on the development of techniques and systems which will support their efficient indexing and retrieval.
- Content-based retrieval is an active research field. Many research prototypes and innovative techniques have been developed during the last decade.
- Some of these have been incorporated in commercial products. Some of the content description tools have affected the MPEG-7 standardization activities and were incorporated in its current version.
- However the **semantic gap** between the users' needs (mainly semantic video search/description) and the currently available technology (mainly low- to middle-level video characteristics) continues to exist. Strong research and development effort is needed in order to develop fully operational tools for video description and retrieval.

Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- **YouTube API & CNN Archive Metadata**
- Media Asset Management Systems (MAMs)

YouTube API Metadata

Properties	Type	Description
Kind	string	Identifies the API resource's type. The value will be youtube#video.
id	string	The ID that YouTube uses to uniquely identify the video.
snippet	object	The snippet object contains basic details about the video, such as its title, description, and category.
snippet.title	string	The video's title (maximum 100 UTF-8 characters).
snippet.description	string	The video's description (maximum length of 5000 bytes).
snippet.thumbnails	object	Key values : default, medium, standard, high, maxres.
snippet.tags	list	List of keyword tags associated with the video.
snippet.categoryId	string	The YouTube video category associated with the video.
contentDetails	object	Length of the video, indication of whether captions are available.
contentDetails.definition	string	Indicates the available definitions (sd, hd).
contentDetails.regionRestriction	object	information about the countries where a video is (or is not) viewable.

[Y_API]

CNN archive API metadata

- Global Content Fields

id	String	The NewsGraph ID uniquely identifying this piece of content
type	String	The content type as defined by the NewsGraph Index.
url	String	The public canonical URL where this piece of content resides.
language	String	The ISO 639-1 language code that this doc appears to be written in.

- Video Content Fields

title	String	The title of this piece of content
description	Array	The description of this piece of content, a synopsis.
source	String	The source where this piece of content originated, typically a copyright holder.
trt	Float	The length in seconds of this video clip.
topics	Array	The list of controlled vocabulary terms associated with this piece of content.

[CNN_API]

Video Description

- MPEG-7 description Standard
 - Parts
 - Descriptors
 - Audiovisual Description Profile (AVDP)
 - Video analysis / annotation software
- EBUCore description Standard
- PBCore description Standard
- Standards and Usages
- YouTube API & CNN Archive Metadata
- **Media Asset Management Systems (MAMs)**

Media asset management Systems (MAMs)

- A MAM system is part of the A/V production chain and the centerpiece of video or audio workflow utilizing standards like MPEG-7 and PBCore.
- **Primary functions:**
 - Bring media to the database in a controlled manner.
 - Allow content management and enrichment of the assets by the library users.
 - Publish or distribute the media in different ways and allows integration with external applications.

AI and Metadata

- Many Broadcasters are investing significantly in Machine Learning using in-house, third party academic or R&D, or commercial tools.
- Broadcasters tend to train tools that are focused in their domain (e.g. business, sport, news)
- How metadata is used in AI:
 - Ground-truth material
 - Training data set
 - Raw data for processing and analysis

Modern Smart Media Asset Systems



- **Smart Media Asset Modules include:**

- Face/Person recognition,
- Object and Logo Detection,
- Speech Recognition and Translation,
- Scene/shot detection,
- Advertisements Detection,
- Semantic content classification,
- Automatic content segmentation,
- Audio and video analysis,
- Activity detection – recognition,
- Natural Language Processing modules such as named entity detection, part of speech tagging, language detection and text classification.

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
pitass@csd.auth.gr**