Video Description

A. Bolari, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.5

Artificial Intelligence &
Information Analysis Lab

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.

MPEG-7 content description standard

- Considerable effort has been invested over the last years to improve MPEG-7's ability to deal with semantic content description, resulting in various MPEG-7 profiles.

- The data described in MPEG-7 may include:
  - Static images, 2D graphics, 3D object models.
  - Audio, Speech, Video.
  - Information on how all these elements are combined in a multimedia scenario.

MPEG-7 content description standard

- Does not aim at a specific application. The representation tools do not depend on digital media content encoding and storage modes.

- Aims to standardize [HUN99]:
  - ***Descriptors*** (“D”): used for the description of various features of the multimedia content (color texture and shape).
  - ***Description Schemes*** (“DS”): Predefined structures of Descriptors and their relations.
  - ***Description Definition Language*** (“DDL”): A language for the determination of the descriptors and the description schemes.
  - An encoded representation of descriptors (efficient storage & fast access to data).

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

MPEG-7 parts

Systems:

- Tools required for the preparation of the MPEG-7 descriptions for efficient transfer and storage of digital media, for the terminal architecture and the regulative interfaces.

Description Definition Language (DDL):

- Allows the creation of new description schemes and, possibly, new descriptors.
- Allows the extension and modification of existing description schemes.
- ***Is based on the XML language***.

  XML has not been designed for the description of audiovisual content. As a consequence, this language can be separated in the following **logical components** :
  - Structural & data type components of the XML scheme language.
  - MPEG-7 extensions to XML Scheme language.

MPEG-7 parts

Visual description tools

- Offer basic structures and descriptors for the following basic visual characteristics : **Color**, **texture**, **shape**, **motion**, **locality** and **face description**.

- Each category consists of elementary and advanced descriptors.

MPEG-7 parts

Audio description tools

- Provide structures, in combination with the multimedia description schemes part, for audio content description.
- A set of **low-level** descriptors uses these structures for audio characteristics encountered in many applications.
- A suit of **high-level** description tools (application-dependent). These high-level tools include general audio recognition tools and description indexing tools, speech content description tools, audio signal description schemes and melody description tools to facilitate queries by humming (a song).

MPEG-7 parts

Multimedia description schemes (1/2)

- Set of description tools (descriptors and description schemes) for both basic and multimedia entities.
- The basic entities: general characteristics used in digital media, e.g., vector, time, text description tools, controlled *dictionaries*, etc.
- Complex description tools have been standardized. Used whenever more than one medium has to be described (e.g., video and audio).

MPEG-7 parts

Multimedia description schemes (2/2)

These description tools can be grouped in five different classes, according to their functionality :

- ***Content description***: representation of perceivable information.
- ***Content management***: information about the medium characteristics, the creation, ingestion and use of the audiovisual content.
- ***Content organization***: representation, analysis, and sorting.
- ***Navigation and access***: specifications of summaries and variants of the audiovisual content.
- ***User interfaces***: description of user preferences and usage record related to the multimedia material consumption.

MPEG-7 parts

Reference software (eXperimentation Model, XM):

- Simulation platform for MPEG-7 descriptors, description schemes, coding schemes and the descriptor definition language.
- The simulation platform needs several non-regulatory components for the execution of a program on the data structures.
- The data structures and the program together form the application. Its applications are divided in two types:
  - Server applications (retrieval).
  - Client applications (search, filtering and/or transcoding) [YAM00].

Compatibility:

- Contains instructions and compatibility check procedures of each MPEG-7 implementation.

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)

The MPEG-7 visual standard

- The most important aim of MPEG-7 is to provide standardized descriptions of stored image or video streams and standardized headers (low-level visual descriptors), which help the user or the applications to identify, classify or filter images or video.

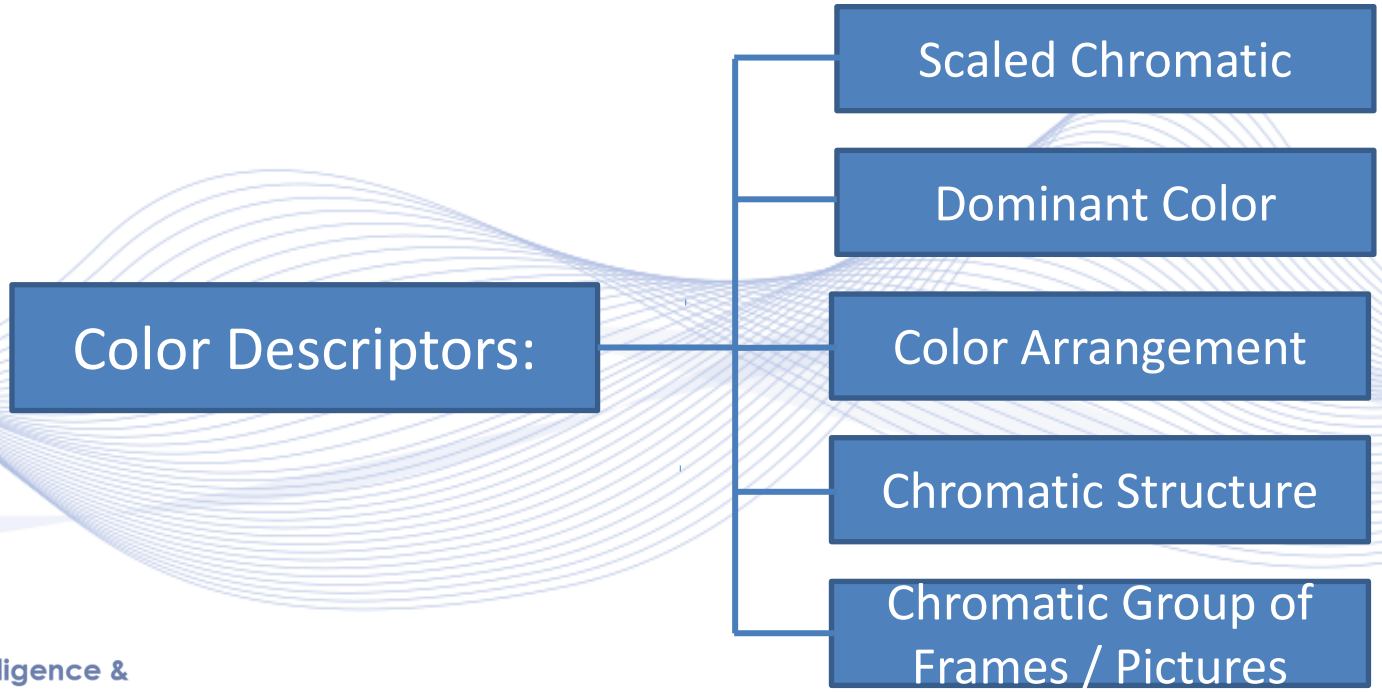
- These ***low-level descriptors*** can be used to compare, filter or display images or video, based on visual content descriptions, or in combination with plain text-based searches.

- MPEG-7 descriptors that have been developed can be grouped in 2 categories:
  - Generic visual descriptors (color, shape and motion characteristics).
  - Specialized visual descriptors. (application independent and include location and identification of human faces).

Color Visual Descriptors

Design of efficient ***color descriptors*** and the detection of similar images.

- There is no generic color descriptor that can be used for all applications.
- Many generic descriptors were standardized, each one being appropriate for a special visual similarity identification function [MAN01].



Color Visual Descriptors

Color spaces:

- ***Hue, Saturation, Value*** (HSV): widely used in image applications.
- ***Hue minimum-maximum difference*** (HMMD): used only in the SCD chromatic structure descriptor.

Color Visual Descriptors

Scaled Chromatic Descriptor (SCD)

- Generic SCD is a color histogram that is encoded using the Haar transform.
- Employs the HSV color space, uniformly quantized in 255 levels.
- To reach a more compact representation, the histogram values are non-uniformly quantized ranging from 16 bits/histogram (for low quality representation of the color distribution) up to 1000 bits/histogram (for high quality applications).
- The match between two SCD descriptions can be performed by matching the corresponding Haar transform coefficients.

Color Visual Descriptors

Dominant color descriptor

- Describes the overall & the local color distribution in images, for fast image retrieval and browsing.
- In contrast to color histograms, this descriptor has a much more compact representation, at the cost of a lower performance in some applications. The colors in a given region are clustered in a small number of representative colors.
- The descriptor consists of the representative colors, their proportion in the region, the spatiochromatic homogeneity and the color variability.

Color Visual Descriptors

Color arrangement descriptor

- Describes a ***spatial color distribution*** in an arbitrarily shaped region.
- The chromatic distribution can be described using the dominant color descriptor.
- The spatiochromatic distribution is an efficient description for contour-based image retrieval, filtering indexed images and for content visualization.

Color Visual Descriptors

Chromatic Structure Descriptor (CSD)

- Exposes the local image features.
- Uses the HMMD space.
- An 8×8 window scans the image. In each window, the number of color occurrences is counted and a chromatic histogram is constructed.

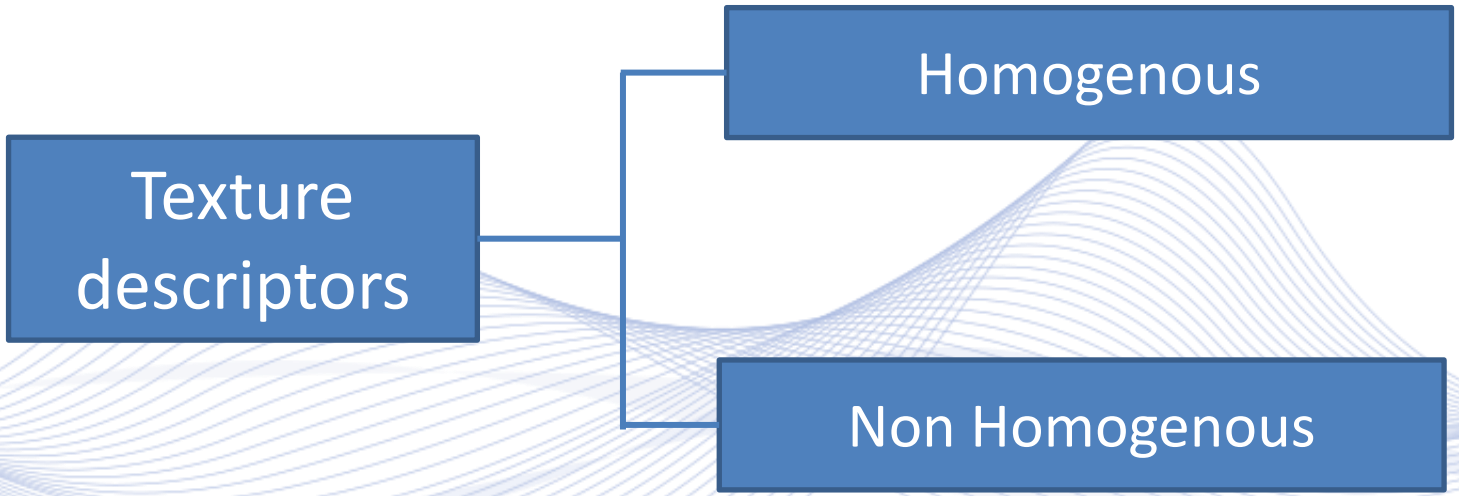
Color Visual Descriptors

Chromatic Group of Frames / Group of Pictures (GoF / GoP)

- Defines a structure, which is necessary for the representation of color features of a collection of similar images or video frames with the use of SCD.
- Useful for image and video retrieval from databases, clustering of video, matching images with image clips and similar applications.
- It consists of mean values, median and histogram intersections of groups of frames, which are calculated based on the histogram of each frame.

Texture Visual Descriptors

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



Texture Visual Descriptors

Homogeneous texture descriptor

- Provides a quantitative characterization of homogeneous texture regions for similarity based image-to-image retrieval.

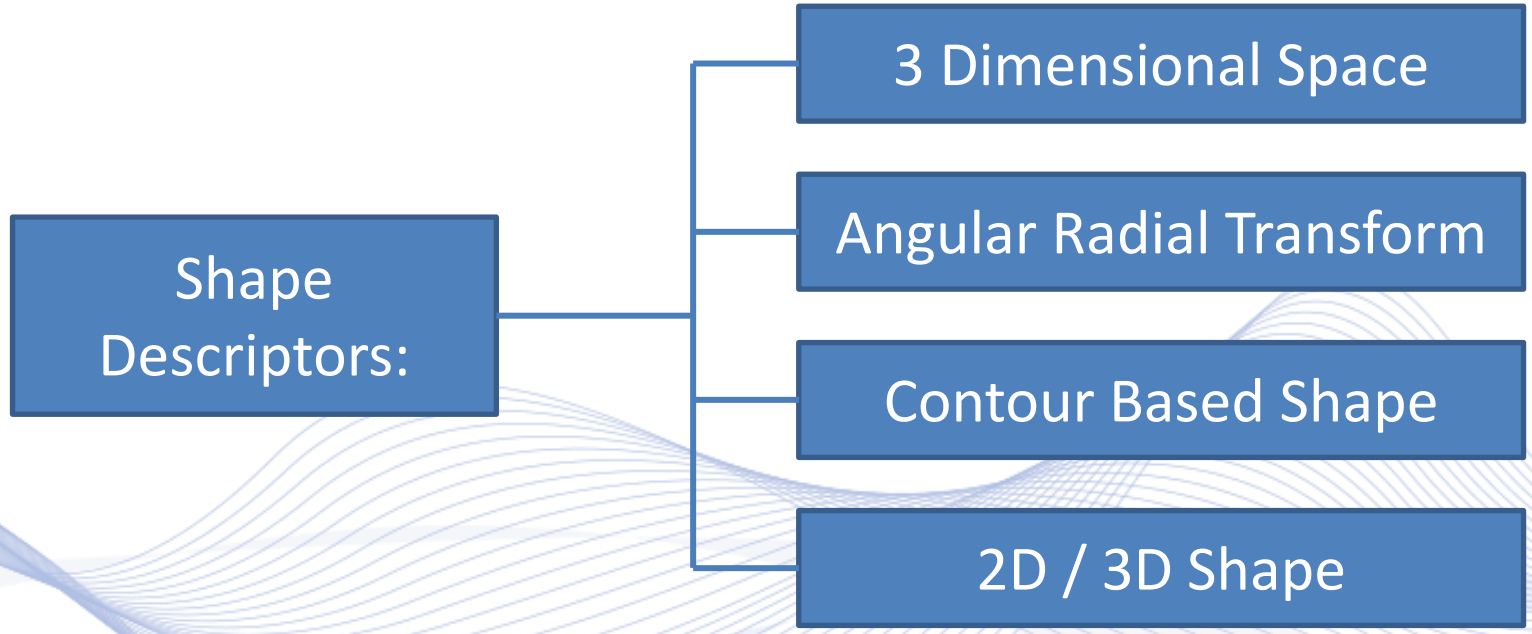
- Is computed by first filtering the image with a ***bank of orientation and scale sensitive filters***, and computing the mean and standard deviation of the filtered outputs in the frequency domain.

Texture Visual Descriptors

Non-homogeneous texture descriptor (***edge histogram***)

- Captures the spatial edge distribution.
- Division of the image in 16 non-overlapping blocks of equal size.
- 5 edge classes: vertical, horizontal, 45°, 135° straight edges and non-directional edges. This is expressed by a 5 bin histogram, one for each edge class.
- The descriptor is scale independent and supports image matching.
- Each histogram bin is non-uniformly quantized using 3 bits, resulting in a 240 bit texture descriptor.

Shape Visual Descriptors



Shape Visual Descriptors

3D shape descriptor (***shape spectrum***)

- Useful in comparing three-dimensional visual objects.
- Based on the concept of shape spectrum, which is defined as the histogram of the ***shape index***, calculated over all the three-dimensional surface.
- The shape index measures the local curvature of each local three-dimensional surface. Histograms with 100 bins are used. Each bin is quantized using a 12-bit word.

Shape Visual Descriptors

Angular Radial Transform (ART)

- Region based descriptor belongs to the ***invariant moment shape descriptors***.
- Suitable for shapes that can be best described by the region they occupy, rather than their contour.
- The central idea is to use moments [PIT01], which are invariant to geometrical transforms. This descriptor performs a complicated angular radial Radon transform, defined on a unit disk in polar coordinates, in order to achieve this goal.
- The ART base function coefficients are quantized and used for image matching. The descriptor is very compact (140 bits per region) and very robust in region segmentation noise.

Shape Visual Descriptors

Contour-based shape descriptor

- Describes objects whose shape features are best expressed by contour information
- Based on ***Curvature Scale Space*** (CSS) contour representations [BGI19]
- Contains ***eccentricity*** and ***circularity*** information of the object contours.
- A CSS index is used for shape matching and indicates the height at the most prominent image chrominance peaks and the horizontal and vertical positions of the rest of the peaks in the CSS image.
- The mean size of this descriptor is 122 bits per contour [MAN01].

Shape Visual Descriptors

2D/3D shape descriptors

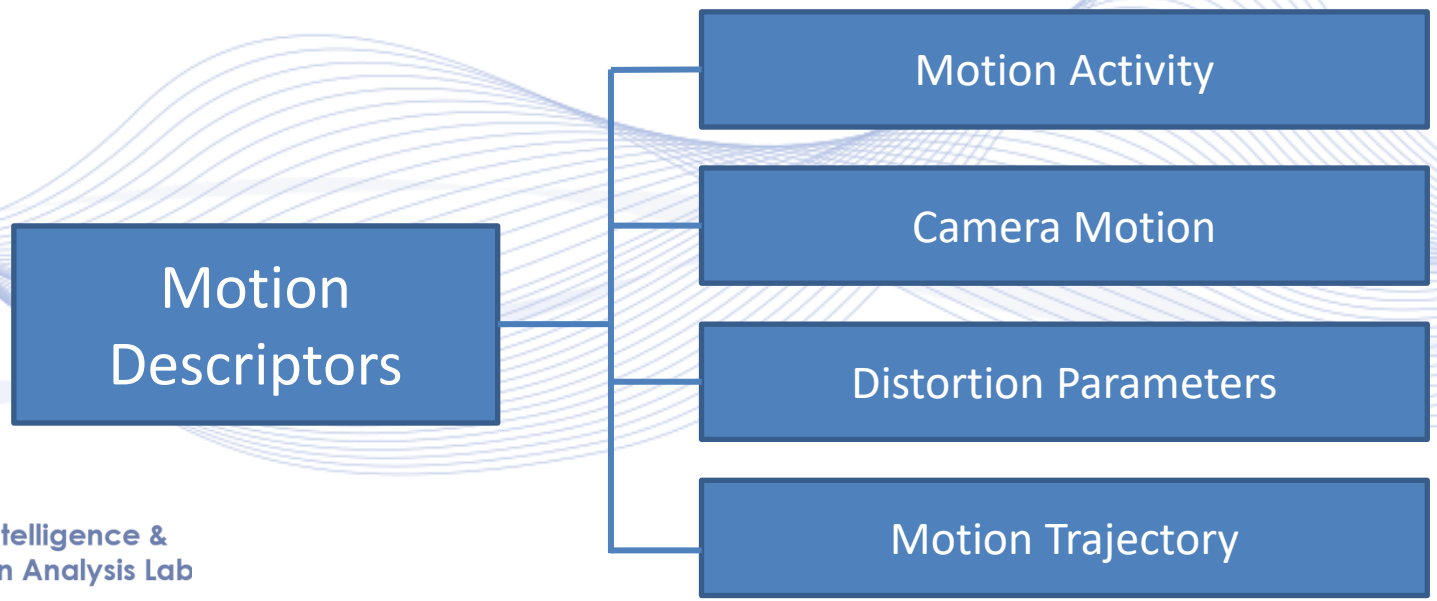
- Detection of similarities between three-dimensional objects can be performed by matching multiple pairs of 2D shots, one for each object.

- Generally, a good search performance for three-dimensional shapes by using contour-based two-dimensional MPEG-7 descriptor has been achieved.

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

Motion activity descriptors

- Captures the overall motion activity level or speed In a video segment (e.g. a scene, a shot or a given set of frames).
- Describes if a scene is slow, fast or very fast (action).
- Measures the ***motion intensity***, based on the standard deviations of the motion vector norms. The standard deviations are quantized in 5 activity values. Optionally, the motion direction, the spatial distribution of the motion activity and the temporal distribution of the motion activity can also be extracted as motion activity descriptors and be used for motion similarity detection.

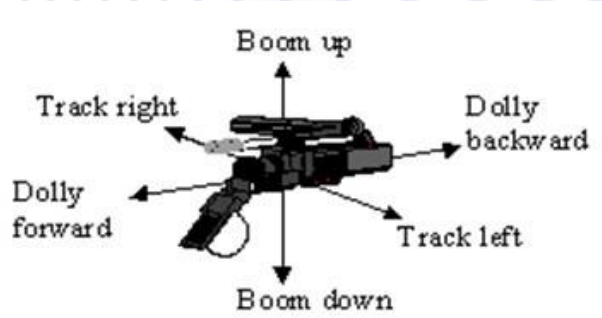
Activity Value	Range of σ
1	$0 \leq \sigma < 3.9$
2	$3.9 \leq \sigma < 10.7$
3	$10.7 \leq \sigma < 17.1$
4	$17.1 \leq \sigma < 32$
5	$32 \leq \sigma$

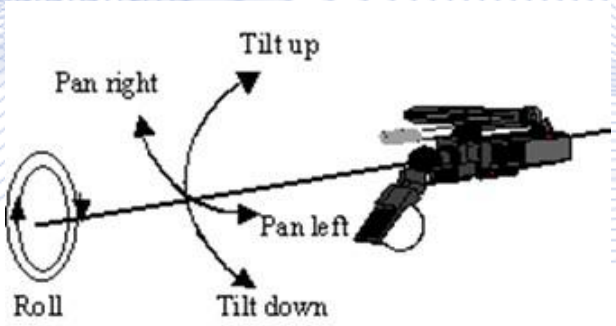
[JEAN01]

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.





[sp1]

Motion Descriptors in video

Distortion parameters

- Provides a way to calculate motion in a scene.
- Here, the overall motion is expressed relative to a general sprite or mosaic of a video scene.

Motion Descriptors in video

Motion trajectory descriptor

- Describes the motion in a video for each independently moving object.
- Describes object displacement as a function of time. It allows the matching of object trajectories for motion-based video search and other relevant applications. The searches can be performed based on descriptors that allow for the fast scene search with the requested object motion.
- Applications:
  - search for objects which move close to a specific region.
  - objects (e.g., cars) which move faster than a speed limit.

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 the 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).

Audio-Visual Description Profile (AVDP)

- AVDP defines a subset of the MPEG-7 description tools needed for storing audiovisual content analysis results.
- AVDP also defines the semantics of some MPEG-7 description tools to suit audiovisual content description.
- AVDP was authored having in mind mainly single-view (“2D”) video & mono/stereo audio
  - Its semantics do not include ways for dealing with certain aspects of 3D video / multichannel audio content description.

Audio-Visual Description Profile (AVDP)

- A framework for using the MPEG-7 AVDP profile for 3D content (e.g., 3DTV audiovisual content) description was proposed as well:
  - A subset of the description tools available in the AVDP has been selected.
  - A description procedure that can be used to store 3D content analysis results has been defined.

- The framework also details procedures for storing 2D content annotations, not foreseen in the AVDP initial guidelines.

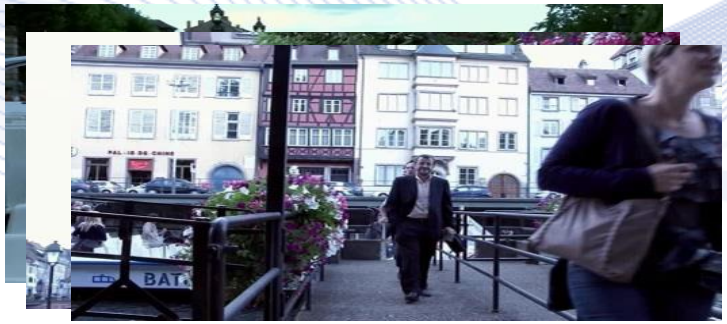
Useful AVDP/MPEG-7 Tools

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

- Example:

movie

shots

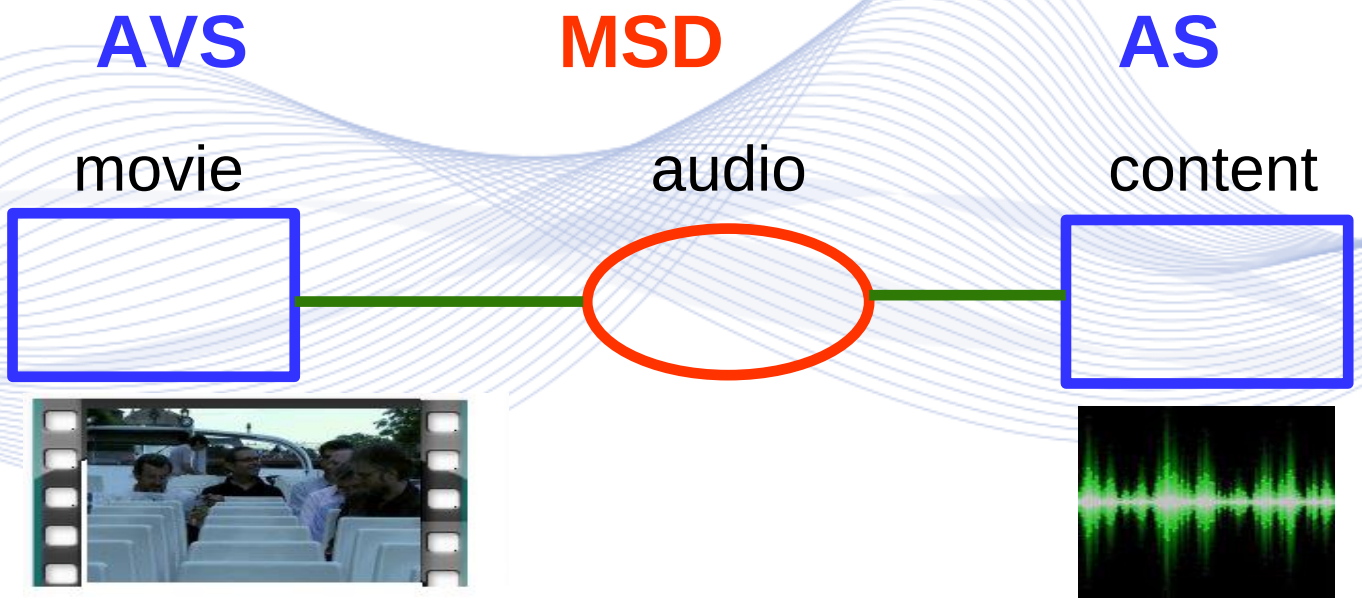


Useful AVDP/MPEG-7 Tools

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

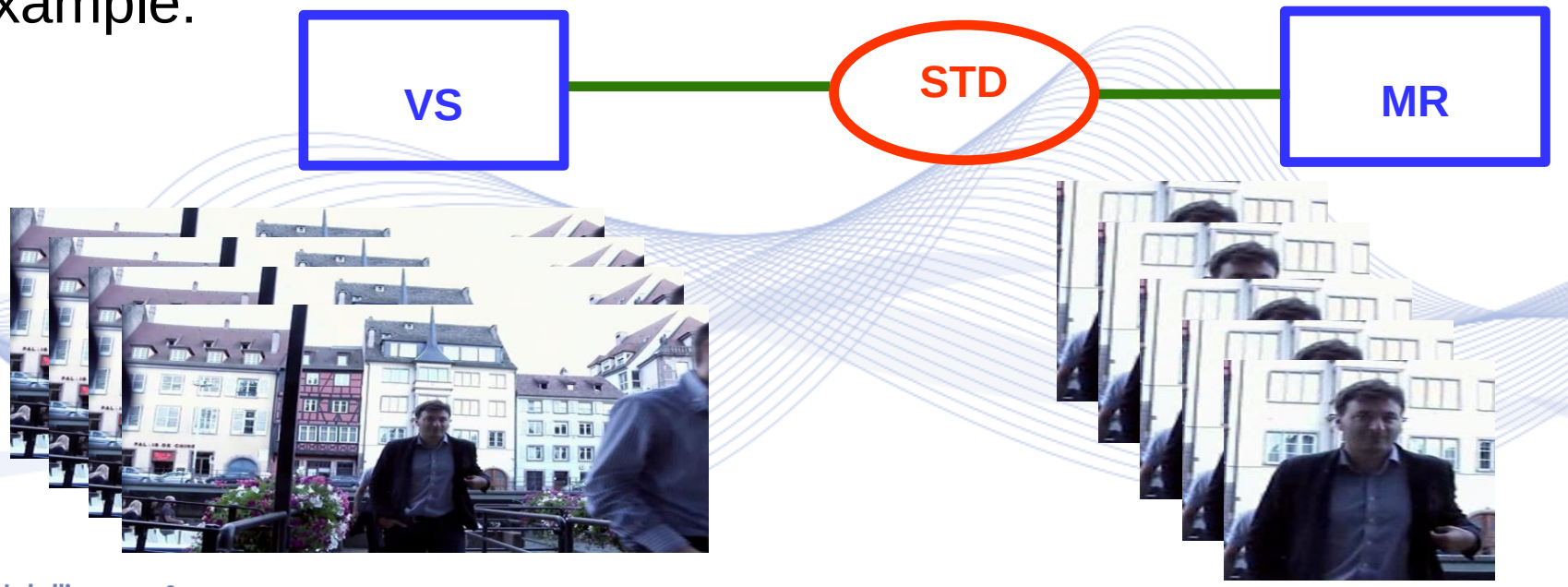
SpatialDecomposition type (SD) decomposes a frame into spatial regions (StillRegion - SR).

- Example:

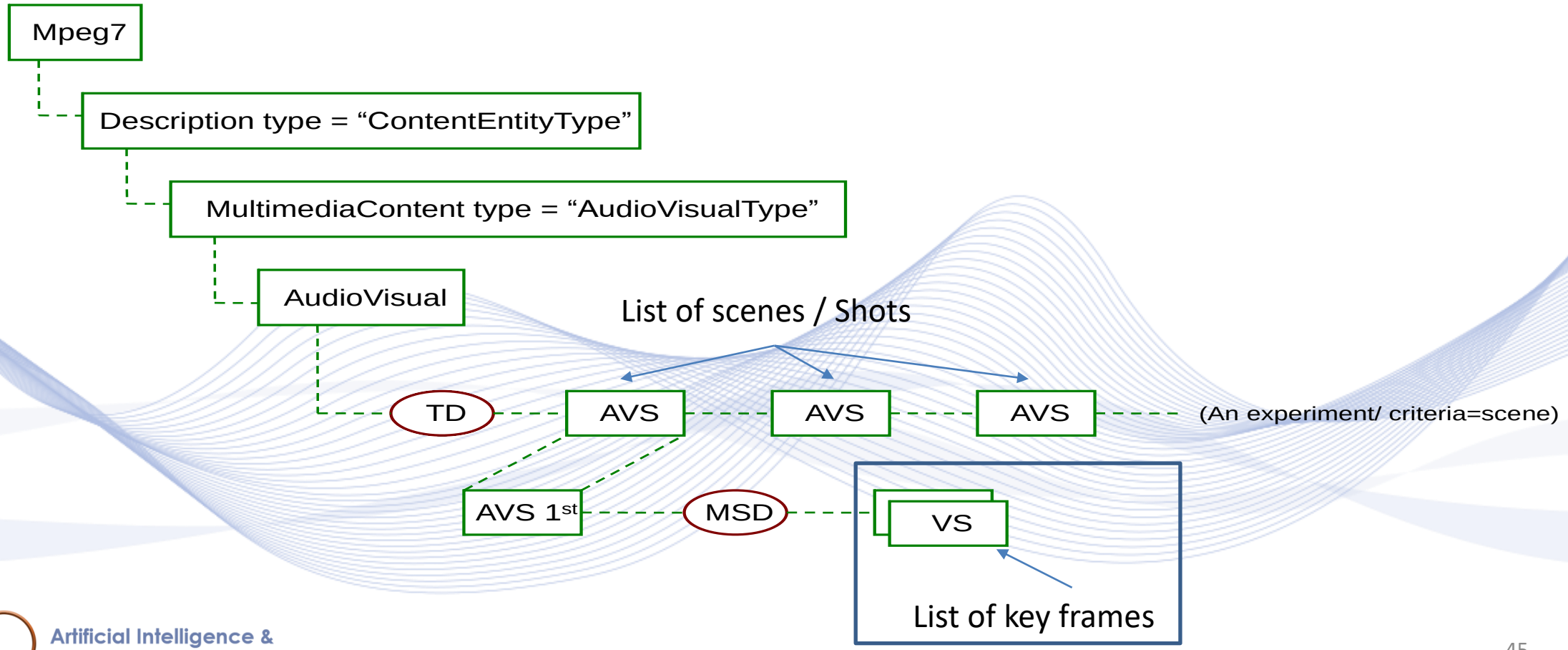
Useful AVDP/MPEG-7 Tools

SpatioTemporalDecomposition type (STD) decomposes a VideoSegment into MovingRegions (MR) corresponding to trajectories of spatial regions over time.

- Example:



Scene/Shot Boundaries Detection

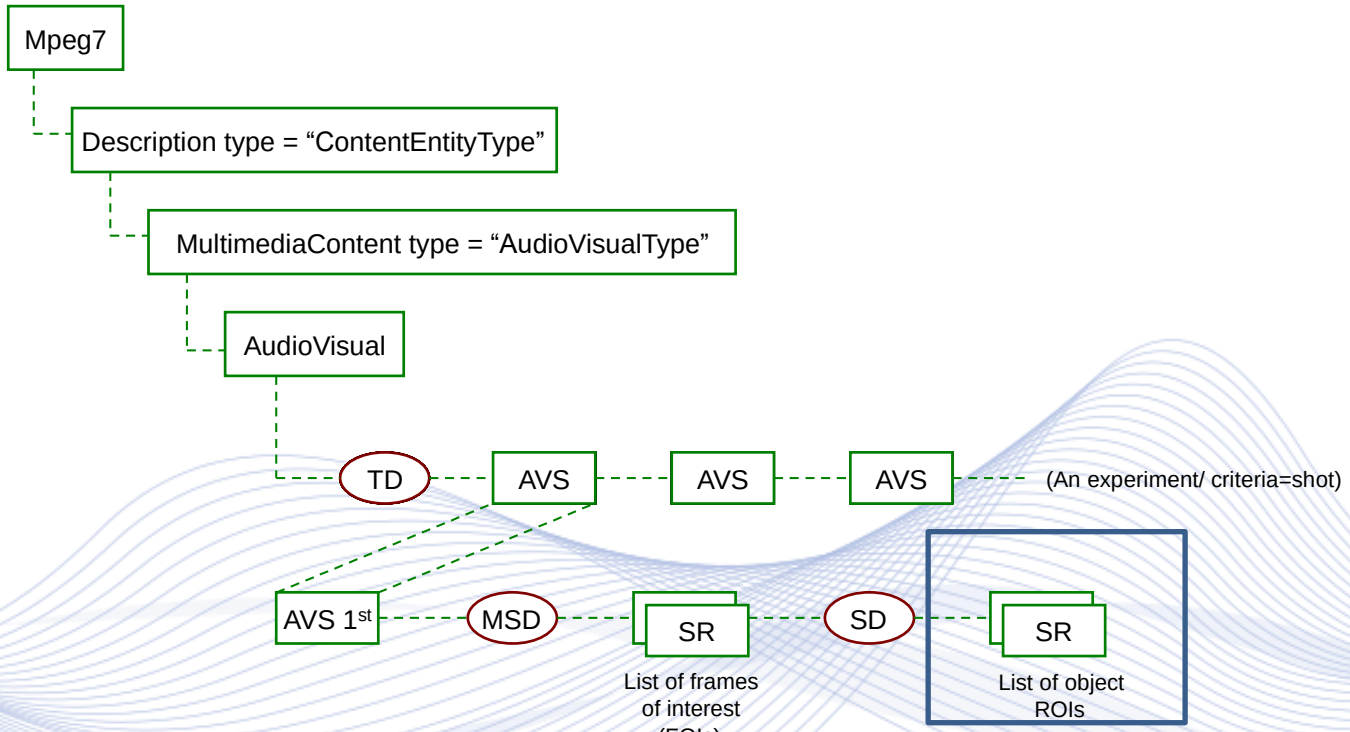


Human/Object Detection

Human/object detection localizes a video 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.

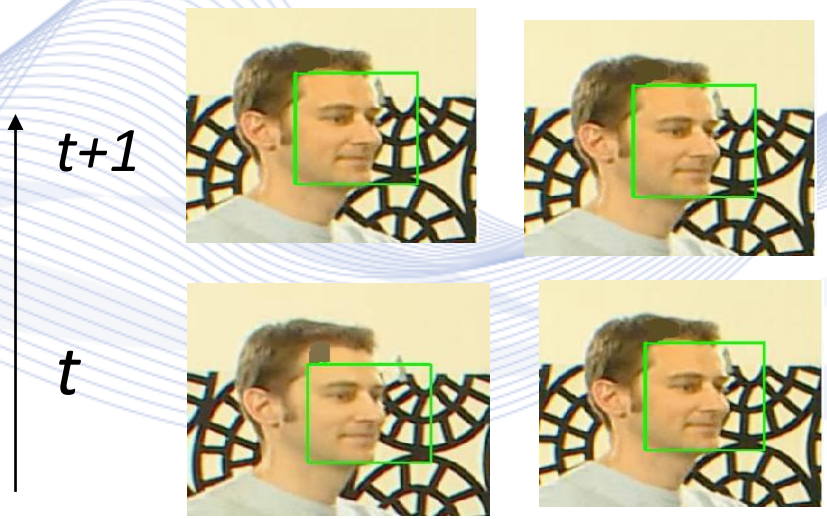
Human/Object Detection



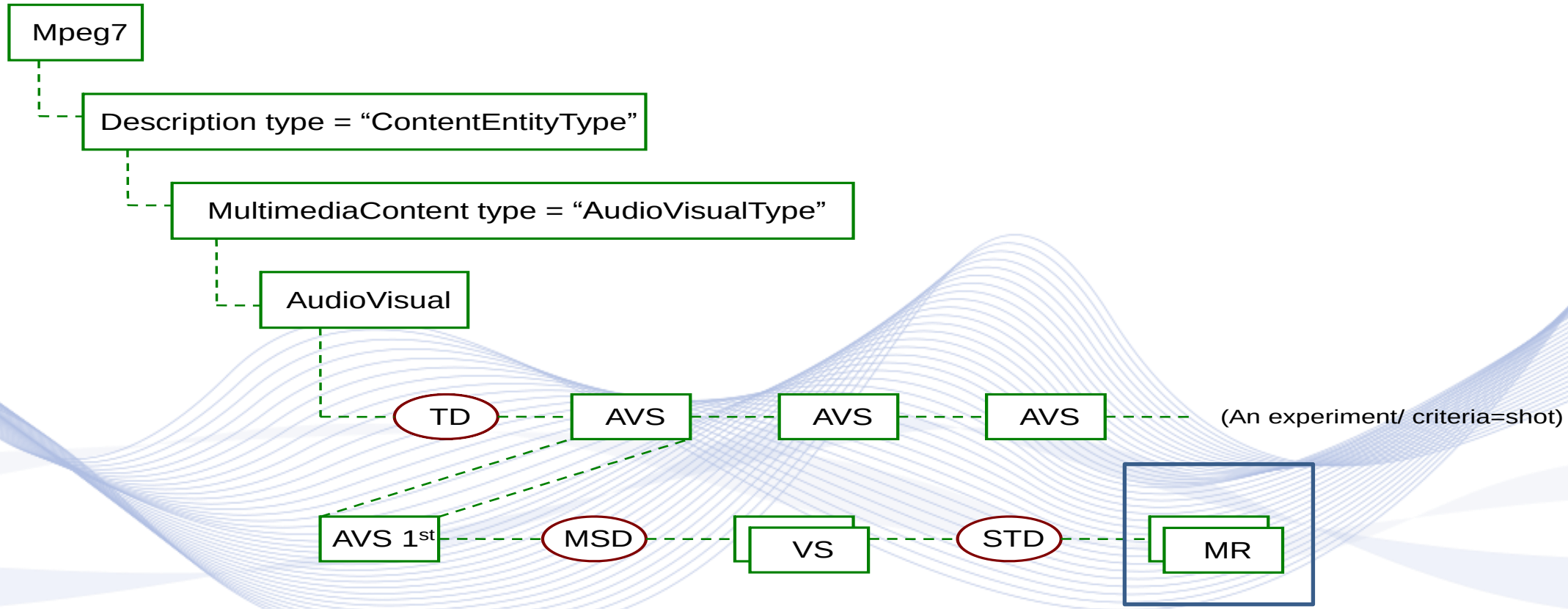
Human/Object Tracking

Human/object tracking follows 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.



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)

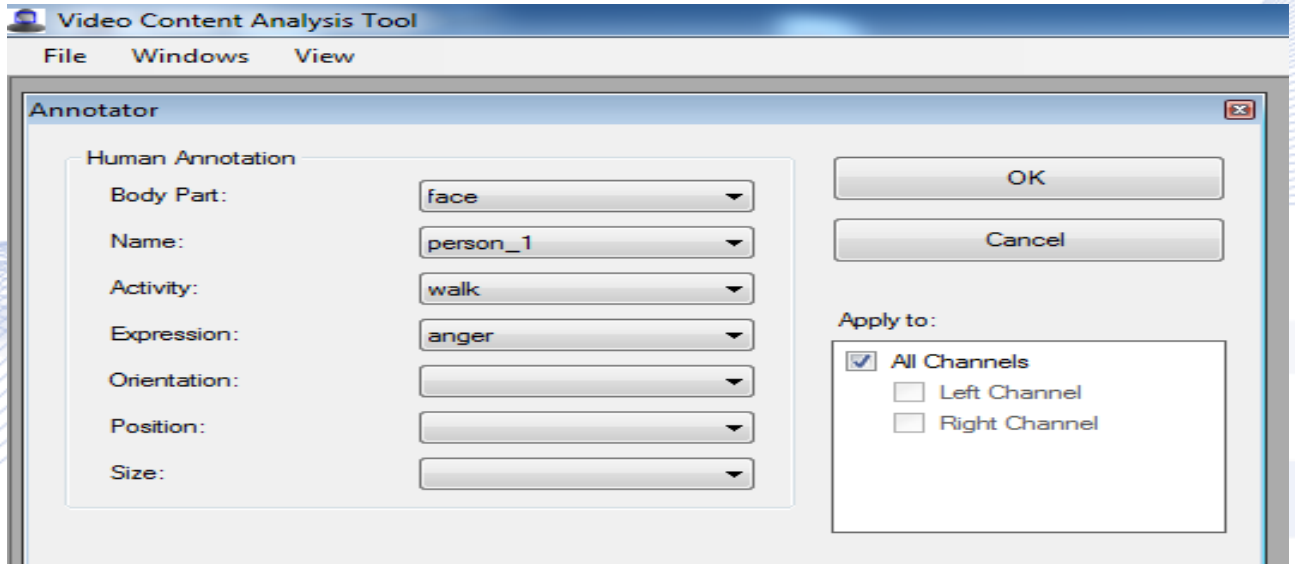
3D video analysis / annotation software

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.

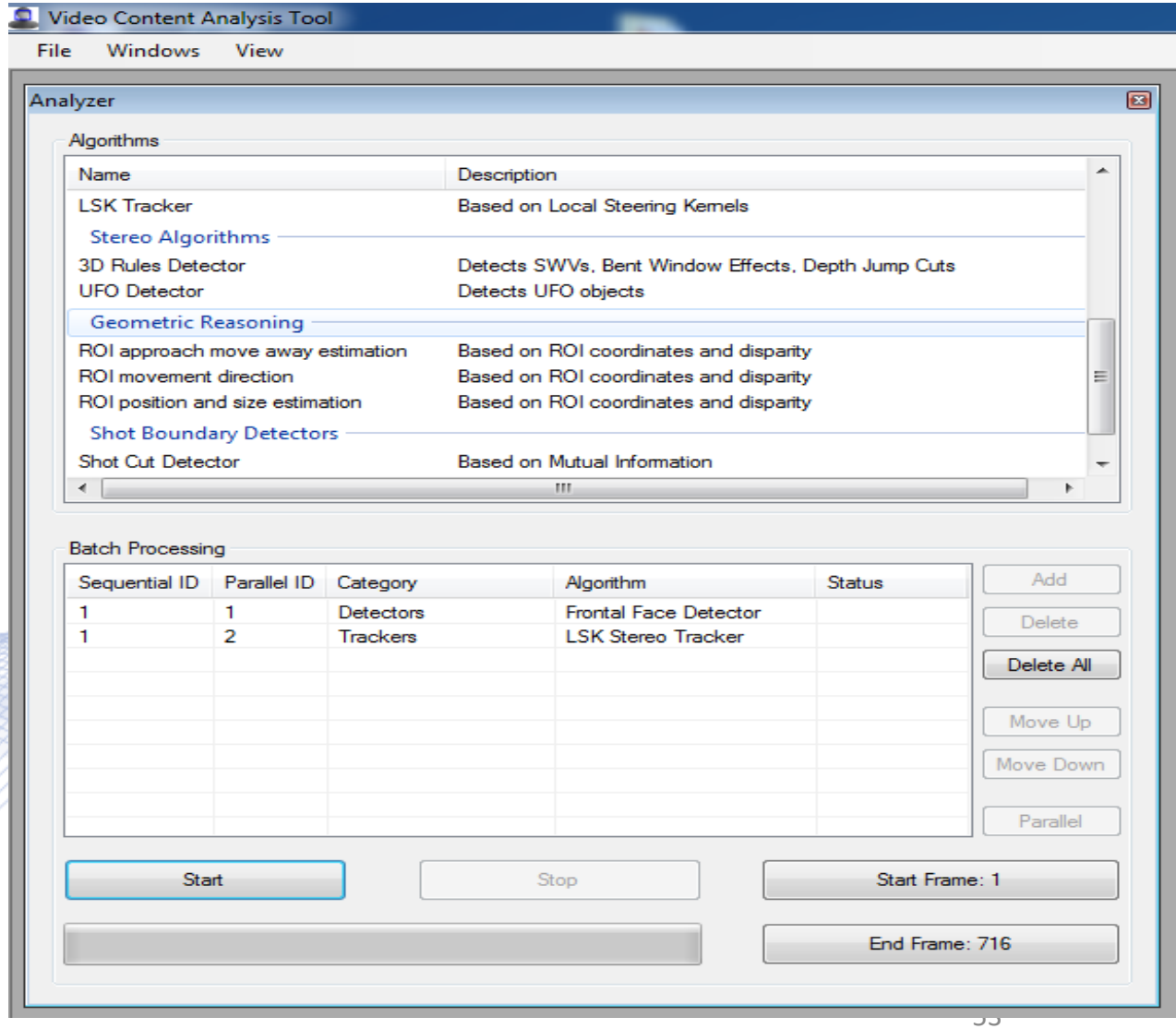
Annotator

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



Analyzer

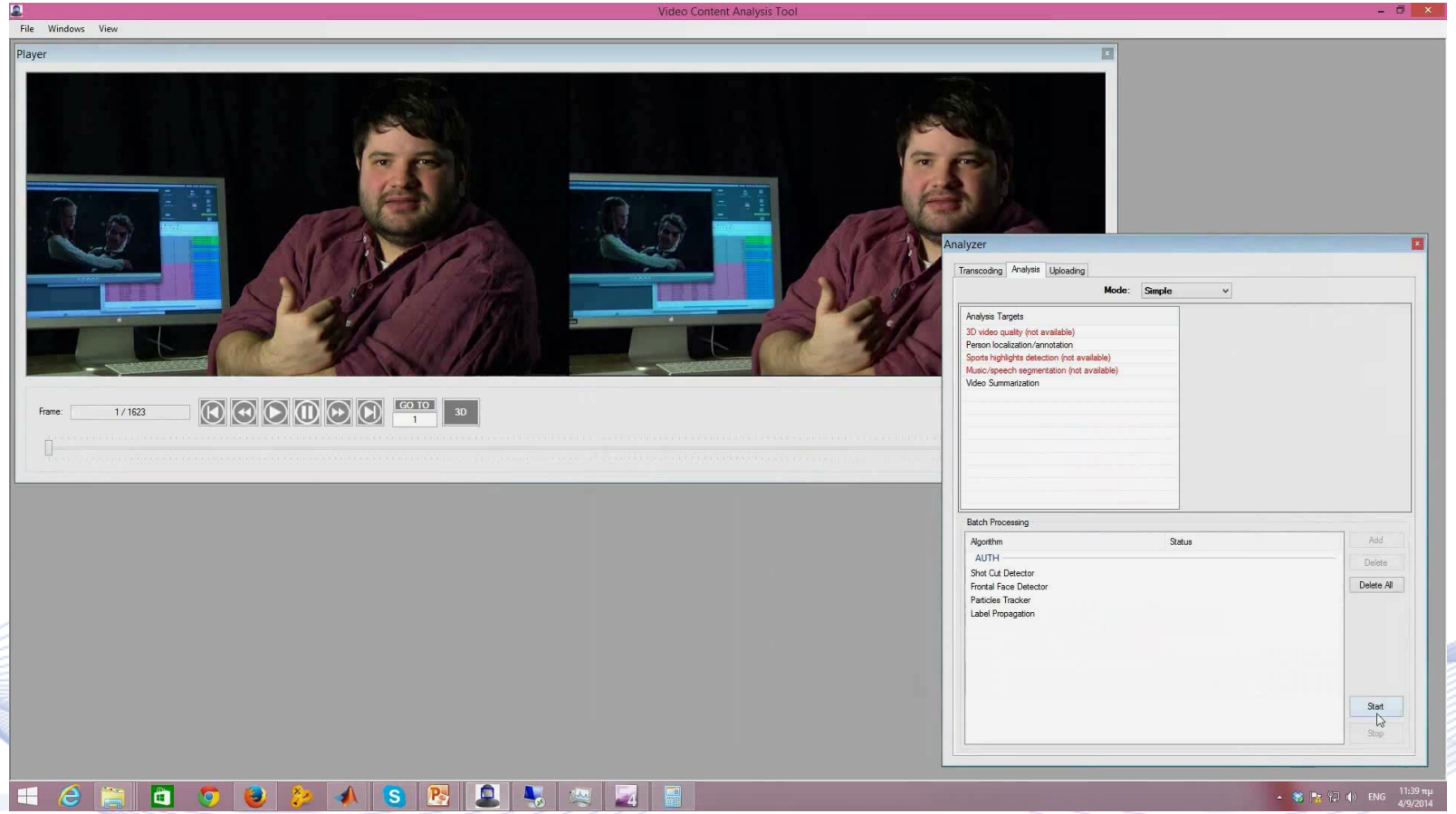
- Enables the user to execute various video analysis algorithms, e.g. face/body detection and tracking.
- Supports importing new algorithms through .dll files.



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.

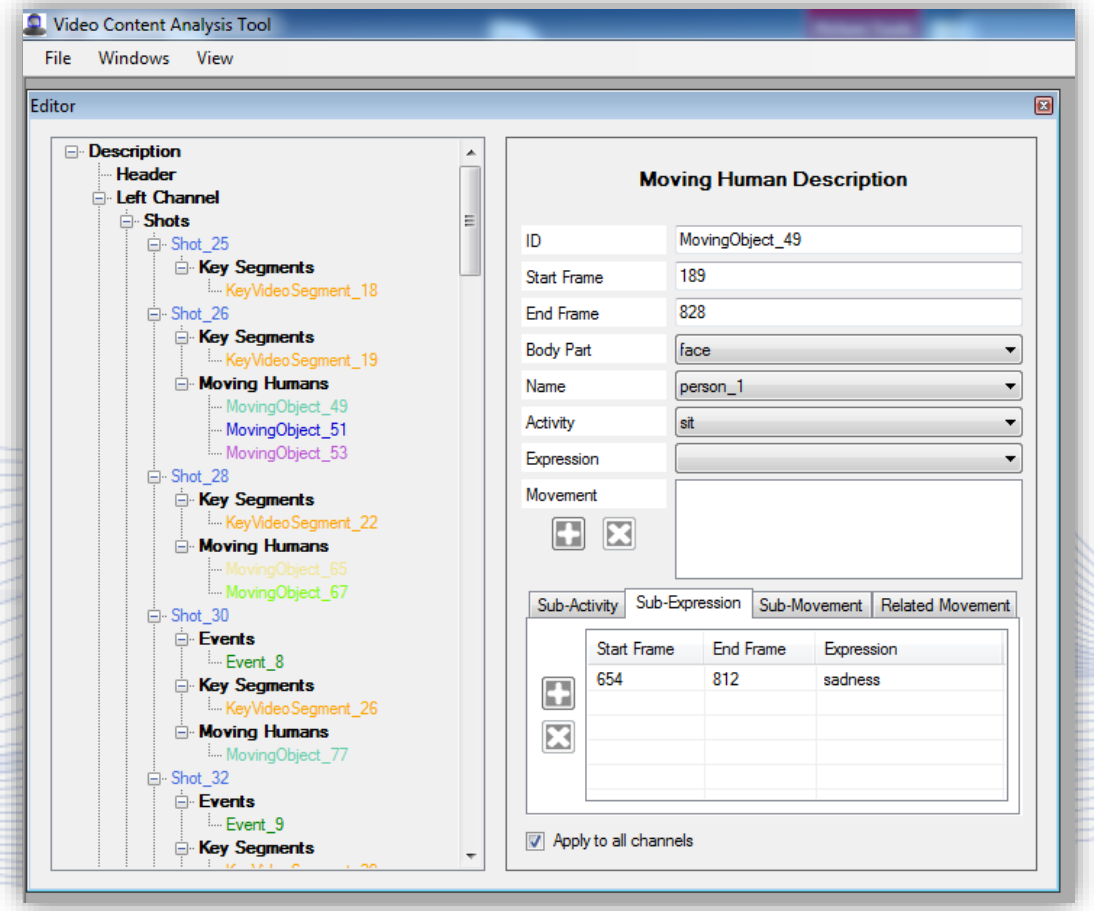
Analyzer


Video Content Analysis Tool
File
Windows
View
Player
Frame:
1 / 1623
GO TO
1
3D
Analyzer
Transcoding
Analysis
Uploading
Mode:
Simple
Analysis Targets
3D video quality (not available)
Person localization/annotation
Sports highlights detection (not available)
Music/speech segmentation (not available)
Video Summarization
Batch Processing
Algorithm
Status
AUTH
Shot Cut Detector
Frontal Face Detector
Particles Tracker
Label Propagation
Add
Delete
Delete All
Start
Stop
ENG
11:39 πμ
4/9/2014

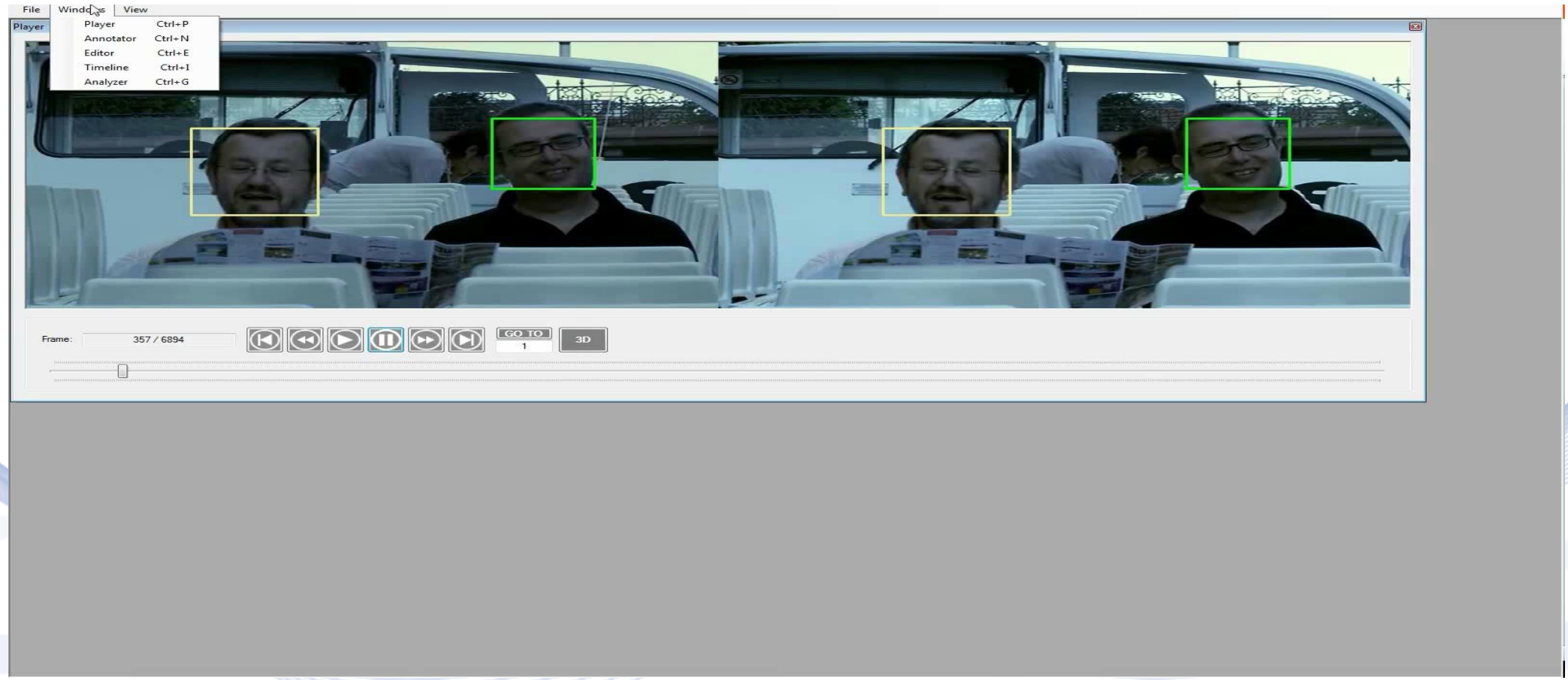
Editor

- 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

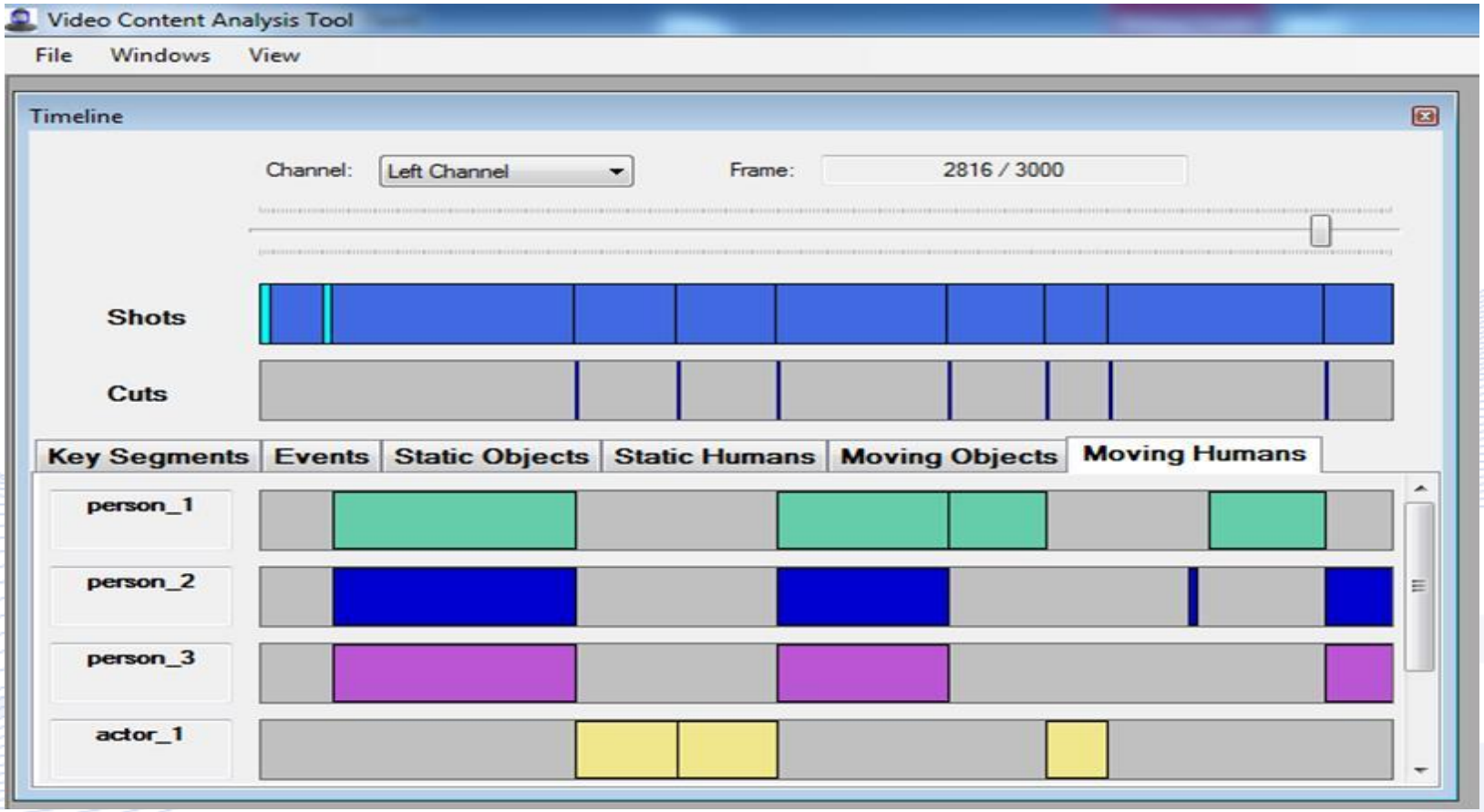

Video Content Analysis Tool
File
Windows
View
Editor
Description
Header
Left Channel
Shots
Shot_25
Key Segments
KeyVideoSegment_18
Shot_26
Key Segments
KeyVideoSegment_19
Moving Humans
MovingObject_49
MovingObject_51
MovingObject_53
Shot_28
Key Segments
KeyVideoSegment_22
Moving Humans
MovingObject_65
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

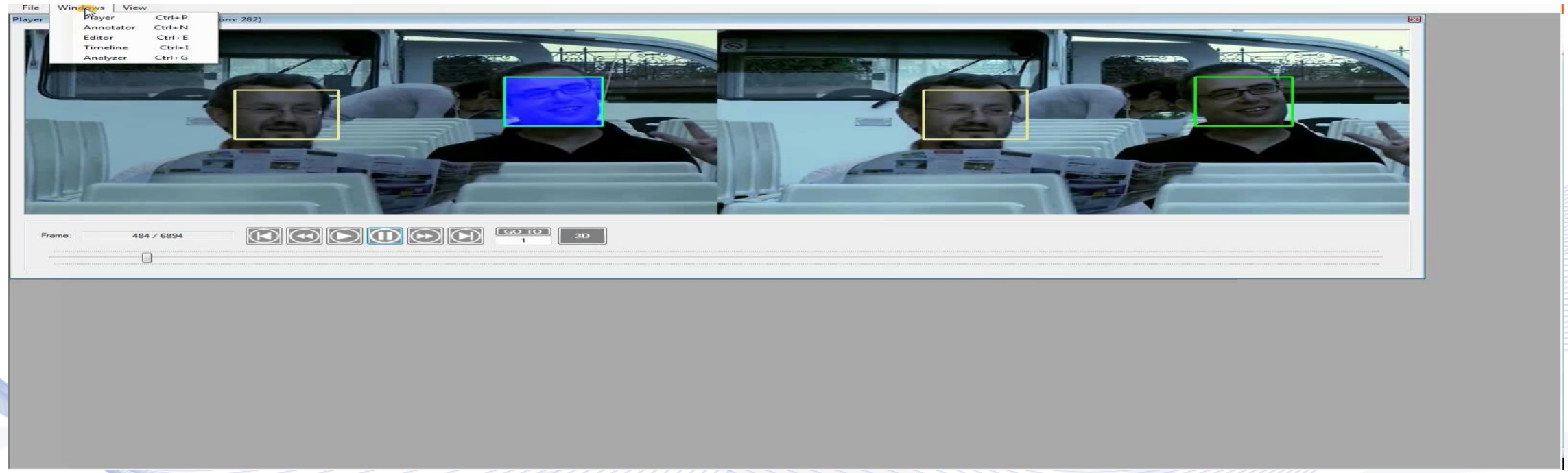

File
Windows
View
Player
Player Ctrl+P
Annotator Ctrl+N
Editor Ctrl+E
Timeline Ctrl+I
Analyzer Ctrl+G
Frame:
357 / 6894
GO TO
1
3D

Timeline

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



Timeline


File
Windows
View
Player
Player Ctrl+P
Annotator Ctrl+N
Editor Ctrl+E
Timeline Ctrl+I
Analyzer Ctrl+G
Frame:
484 / 6894
GO TO
1
3D

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_Shema]
- 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

EBU Core Metadata Set

- Audiovisual objects are described in different languages by:
  - Identifiers (ISAN, EIDR, UMID),
  - Titles (Working title, Original title, program title),
  - Descriptions (summary, Script, synopsis) or Rights (Copyrights, exploitation rights)
- No single standard can cover all user needs.
- Extends Dublin Core with several data types:
  - ***Rating*** (parental or user)
  - Publication history (year, media, format, rights)
  - Planning (distribution schedule)
  - Part (e.g. split and organize chunks of related metadata or to generate timelines of dynamic technical parameters)

EBUCore: root elements

- title
- creator
- description
- publisher
- date
- type
- format
- Identifier
- language
- relation
- isEpisodeOf
- isSeasonOf
- rights
- version
- costume
- food
- emotion
- action

```
<ebuCoreMain xmlns....>
 <coreMetadata>
 <part partId="season1" partName="season a" typeLabel="Season">
 <part partId="e1" partName="episode 1" typeLabel="Episode">
 </part>
 <part partId="e2" partName="episode 2" typeLabel="Episode">
 </part>
 </part>
</coreMetadata>
</ebuCoreMain>
```

Element ebuCoreMainType / coreMetadata
[EBUCore_Shema]

EBUCore: video format

- videoFormat
- width
- height
- framerate
- aspectRatio
- videoEncoding
- codec
- bitRate

```
<ebucore:videoFormat videoFormatName="AVC">
<ebucore:width unit="pixel">1920</ebucore:width>
<ebucore:height unit="pixel">1080</ebucore:height>
<ebucore:frameRate factorNumerator="25000“
factorDenominator="1000">25</ebucore:frameRate> ebucore:aspectRatio
typeLabel="display">
<ebucore:factorNumerator>16</ebucore:factorNumerator>
<ebucore:factorDenominator>9</ebucore:factorDenominator>
</ebucore:aspectRatio>
<ebucore:videoEncoding typeLabel="High 4:2:2 Intra@L4.1"/>
<ebucore:codec>
<ebucore:codecIdentifier>
<dc:identifier>0D01030102106001-
04010202013231 02</dc:identifier>
</ebucore:codecIdentifier>
</ebucore:codec>
EBU Core Metadata Set Tech 3293 v.1.8
22
<ebucore:bitRate>113664000</ebucore:bitRate>
```

EBUCore: audio format

- audioEncoding
- codec
- samplingRate
- sampleSize
- bitRate
- bitRateMode
- channels

```
<ebucore:audioTrackConfiguration typeLabel="EBU R 123:
16c"/>
<ebucore:samplingRate>48000</ebucore:samplingRate>
<ebucore:sampleSize>24</ebucore:sampleSize>
<ebucore:bitRate>1152000</ebucore:bitRate>
<ebucore:bitRateMode>constant</ebucore:bitRateMode>
<ebucore:audioTrack trackId="2001" trackName="A1"/>
<ebucore:channels>1</ebucore:channels>
<ebucore:technicalAttributeString
typeLabel="ChannelPositions">
```

EBU Core Metadata Set

- Users can add extensions based on their needs and define their own specific technical metadata elements for audio, video, pictures and documents.

- EBUCore schema is strictly organized around xml complex types, which can be redefined and customized.

Where is EBUCore?

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

PBCore Metadata Set

- ***Provides a standard for describing media objects***:
  - Intellectual content – Title, Subject, Description, Genre.
  - Intellectual Property – Creator, Distributor, Publisher.
  - Instantiation – technical metadata about physical or digital representation.
  - Extension – allows for the integration of data from other metadata schema.

PBCore: video format

Identifier:12345

Title: Hamlet

Description: Filmed production of Shakespeare Hamlet

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<pbcoreDescriptionDocument xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html
https://raw.githubusercontent.com/WGBH/PBCore_2.1/master/pbcore-2.1.xsd">
  <pbcoreIdentifier source="PBCore Handbook">12345</pbcoreIdentifier>
  <pbcoreTitle>Hamlet</pbcoreTitle>
  <pbcoreDescription>Filmed production of Shakespeare's Hamlet</pbcoreDescription>
</pbcoreDescriptionDocument>
```

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)

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]

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]

YouTube recommendations system and relevant metadata

- ***Click-through rate*** (likelihood of clicking the video after seeing it).
- ***Watch time*** (combined amount of time viewers spend watching a channels videos).
- The number of videos a user has watched from the channel.
- How recently the user watched a video about this topic.
- ***Past user searches***.
- User's previously watched videos.
- User's location and demographic information.

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.

Modern Smart Media Asset Systems

Modern Smart Media Asset Management Systems

- Provide tools for Archival, Retrieval and Dissemination of audiovisual content,
- Integrate Machine Learning models and lower the number of operators needed,
- Provide automated (with or without human feedback) metadata enrichment pipelines,
- Include modules that classify, analyze summarize or describe media content based on visual, audio and semantic feature descriptors.

Media Asset Management Systems are highly sophisticated software platforms that rely on:

- Enterprise software engineering frameworks,
- Scalable SQL, NoSQL databases and text indexers,
- Graph databases,
- Distributed work scheduling and messaging software,
- Graphics engines for rich Visualizations (data understanding and decision making workflows).

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.

Bibliography

[AVD14] K. Papachristou, N. Nikolaidis, I. Pitas, A. Linnemann, M. Liu and S. Gerke, "Human-centered 2D/3D video content analysis and description," *8th International Conference on Electrical and Computer Engineering*, Dhaka, 2014

[HUN99] J. Hunter, “MPEG-7 behing the scenes”, D-Lib Magazine, September 1999

[JEA01] Sylvie Jeannin and Ajay Divakaran, “MPEG7 visual motion descriptors”, IEEE Trans. on CSVT, 2001

[SLPL1] Mpeg-7 Motion Descriptors, https://slideplayer.com/slide/5139017/

[BGI19] A Beginner's Guide to image shape Feature Extraction Techniques, Jyosmita Chaki, Nilanjan Dey, CRC Press 2019

[YAM00] S. Jrsnnin, L. Cieplinski, J.R Ohm, A. Yamada, M. Pickering, Eds. M. Editors, “MPEG-7 visual part of experimentation modelvesion 8,0”, ISO/IEC JTC1/SC29/WG11/N3673, October 2000

[EVA14] FIMS, EBUCore and semantic web key features in broadcasting production, reaserchgate.net, Jean-Pierre Evain, 2014

[EBUCore_Shema] EBUCore Main Schema https://www.ebu.ch/metadata/schemas/EBUCore/documentation/ebucore.html

[Y_API] https://developers.google.com/youtube/v3/docs/videos

[CNN_API] CNN Content Types (Video) http://developer.cnn.com/docs/basics/content/#video

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