

DIGITAL - Institute for Information and Communication Technologies



Using the MPEG-7 Audiovisual Description Profile (AVDP)

Werner Bailer

EBU MIM/MDN Workshop, Geneva, 20 June 2012

Contents

- MPEG-7 and the Audiovisual Description Profile (AVDP)
- Use cases and examples
- Working with AVDP

Part I

MPEG-7 and the Audiovisual Description Profile (AVDP)

Motivation (1)

- Motivated by activities of MIM/SCAIE group
 - Study and evaluation of Content analysis-based Automatic Information Extraction tools in media production
 - facilitate the introduction of new production tools based on automatic extraction of information

 - Relevant tools include
 - Speech Recognition
 - Audiovisual Segmentation
 - Personality Identification
 - Semantic Detection
 - Text Recognition
 - Subject Classification
 - Format Detection
 - Named Entities Detection
 - Copy / Repetition Detection
 - Content Summarisation
 - Keyword Extraction
 - Automatic text Translation
-

Motivation (2)

- Automatic analysis tools generate
 - detailed time-based metadata
 - low-level audio/video features (e.g., color descriptors)
- Common metadata format is needed to
 - represent results
 - exchange between organisations
 - perform automated evaluation

MPEG-7

- Multimedia Content Description Interface
 - ISO/IEC 15938, version 1 (2001) and 2 (2004)
- Metadata of multimedia content
 - descriptors, description schemes
 - definition based on an extension of XML Schema
 - serialised in XML or binary format
- Multimedia description schemes (part 5)
 - content structure (decompositions, segments)
 - descriptive and technical metadata
- Visual and Audio (parts 3 and 4)
 - descriptors for visual and audio low-level features

Why MPEG-7?

- Support for describing arbitrary content structures
 - time segments, regions, moving regions
 - Support for low-level feature descriptors
 - Benefits of MPEG-7
 - designed as a data model, not just as an exchange format
 - broad range of applications
 - flexible, fine-grained description
 - content structuring capabilities
 - Drawbacks
 - complexity: generic types, deep hierarchical structures
 - interoperability problems: several ways to describe the same, conformance to standard can only be checked on syntactic level but not on a semantic level
-

Why profiles for MPEG-7?

- Address issues of complexity and interoperability
- Profiles are a proven concept
 - subsets of the standard for certain functionalities and/or applications
 - cf. MPEG-2, MPEG-4
- Definition of an MPEG-7 profile
 - Selection of description tools → **Profile XML Schema**
 - Constraints on description tools → **Profile XML Schema**
 - Semantic constraints → **Textual description**

Design goals for a new profile

- Fulfill the requirements of MIM/SCAIE
- Harmonise earlier efforts
 - Detailed Audiovisual Profile (DAVP, proposed by JRS)
 - Metadata Production Framework (MPF from NHK)
 - share many design principles, some different implementation choices
- Simplify
 - based on experience from using DAVP and MPF
 - some included elements are never used
 - some constructs can be simplified in terms of cardinality/flexibility
- First profile based on MPEG-7 v2 (2004)
- First profile to include low-level video and audio features (pt. 3&4)

AudioVisual Description Profile – AVDP (1)

■ An MPEG-7 AVDP file contains

- 0 or more descriptions of audiovisual contents
- 0 or 1 summaries of the content(s)

■ Describe audio, video or audiovisual content

- no other types of multimedia content
- no collections of content
- same top-level structure for audio, video or audiovisual content

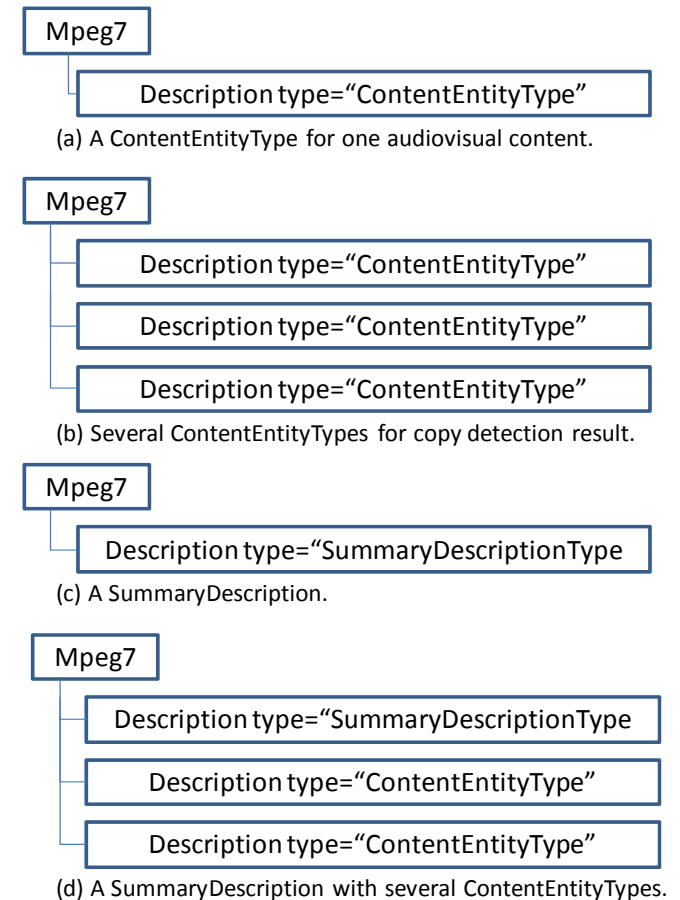
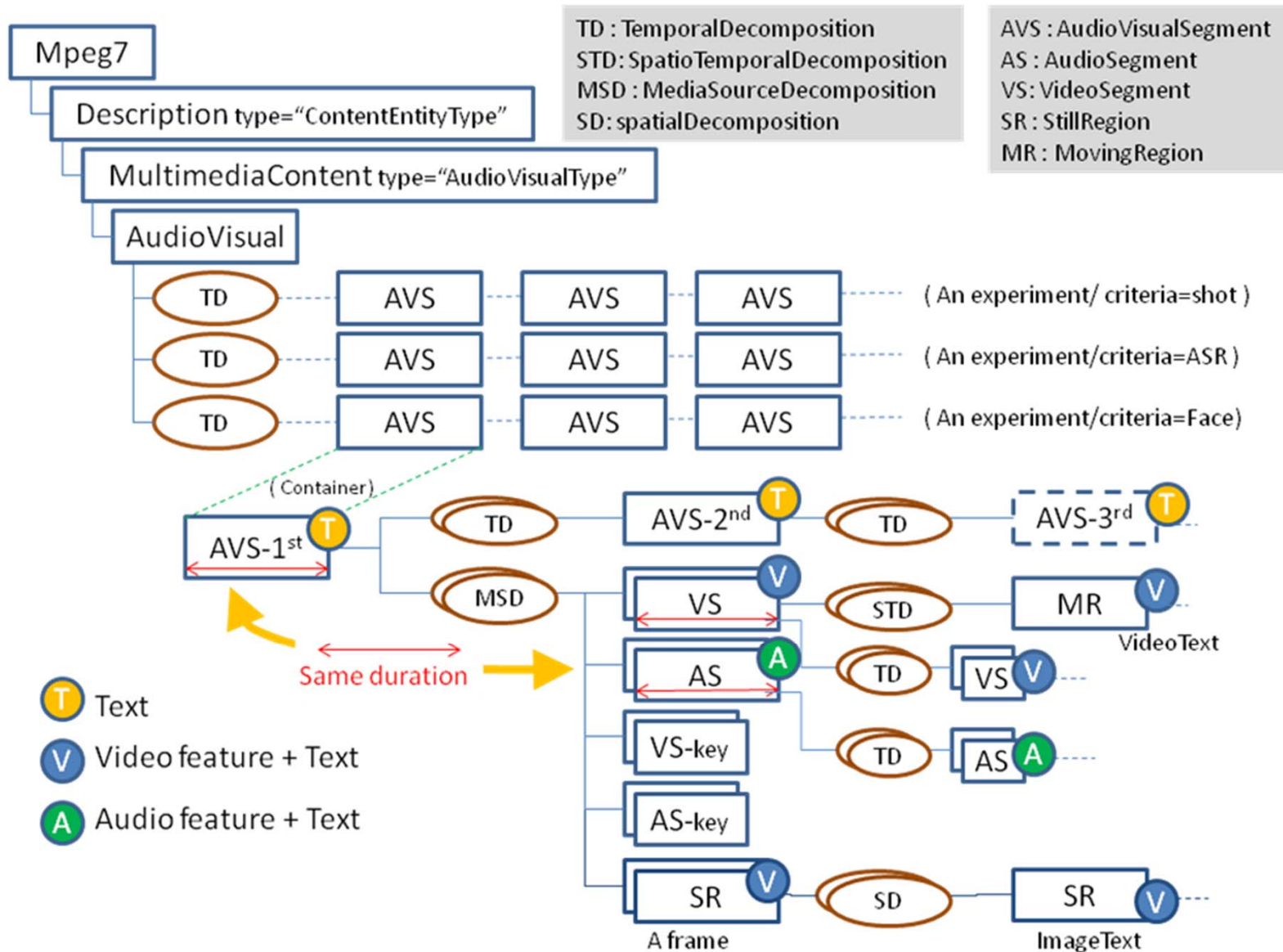


Figure 1 Top level structure

AudioVisual Description Profile – AVDP (2)

- Modularity
 - separate metadata produced by different tools
 - separate metadata on different abstraction level
 - separate metadata specific to one modality or valid for all
 - separate content segmentation and representative elements, e.g. shots and key frames
- Levels
 - 1 temporal/editorial structure
 - 2 modalities or more detailed temporal/editorial structure
 - 3 (spatio)temporal structure within modalities (if applicable)
- Low-level features
 - on levels 2+ in the appropriate parts of the description

AVDP description structure



High-level descriptions in AVDP

- Not using the MPEG-7 Semantic DS
- uses simpler constructs to reference external
 - classification schemes
 - controlled vocabularies
 - ontologies
 - e.g. <http://tech.ebu.ch/MetadataReferenceData>

Profile specification

- Profile specification document
 - list of included types
 - constraints on elements and cardinalities
 - semantic constraints on use of these types
 - semantics of description structure
 - **ISO/IEC 15938-9:2005/Amd1**
- Profile XML schema
 - included types
 - excluded elements, restrictions on cardinality
 - few of the semantic and structure constraints
 - expected to become ISO standard in 2012

Classification schemes

15

- MPEG-7 format for controlled vocabulary
- Addition classification schemes
 - identifying decompositions and segments of the content structure
 - avoid interoperability problems
 - <http://www.ebu.ch/metadata/cs/mpeg/avdp/>

Type	Element/Attribute	Recommended CS
MediaFormatType	FileFormat	ebu_FileFormatCS
	VisualCoding/Format	ebu_videoCompressionCodeCS
	AudioCoding/Format	ebu_AudioCompressionCodeCS
MediaAgentType	Role	ebu_RoleCodeCS
ClassificationType	Genre	ebu_ContentGenreCS
	Format	ebu_EditorialFormatCodeCS
RelationType	Type	SegmentRelationCS
SegmentType	StructuralUnit	StructuralUnitCS
SegmentDecompositionType	Criteria	DecompositionCS

Part II

Use cases and examples

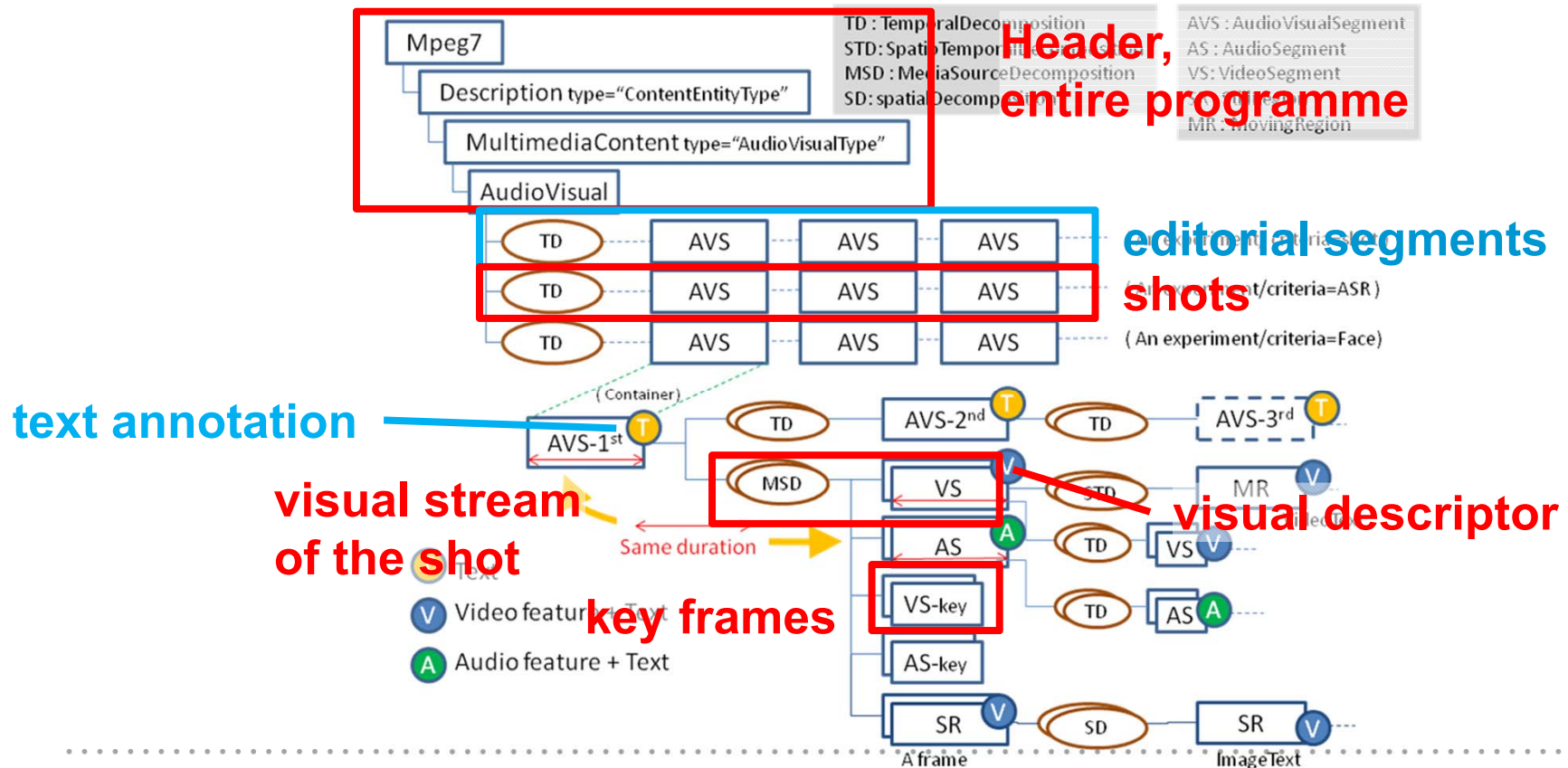
Basic annotation

17

- Shots, key frames
- Camera motion
- Editorial segments
- Text annotation
- SVAS (currently still DAVP based):
 - test version at
<http://www.joanneum.at/en/digital/products-solutions/semantic-video-annotation.html>

Basic annotation

MPEG-7 AVDP structure



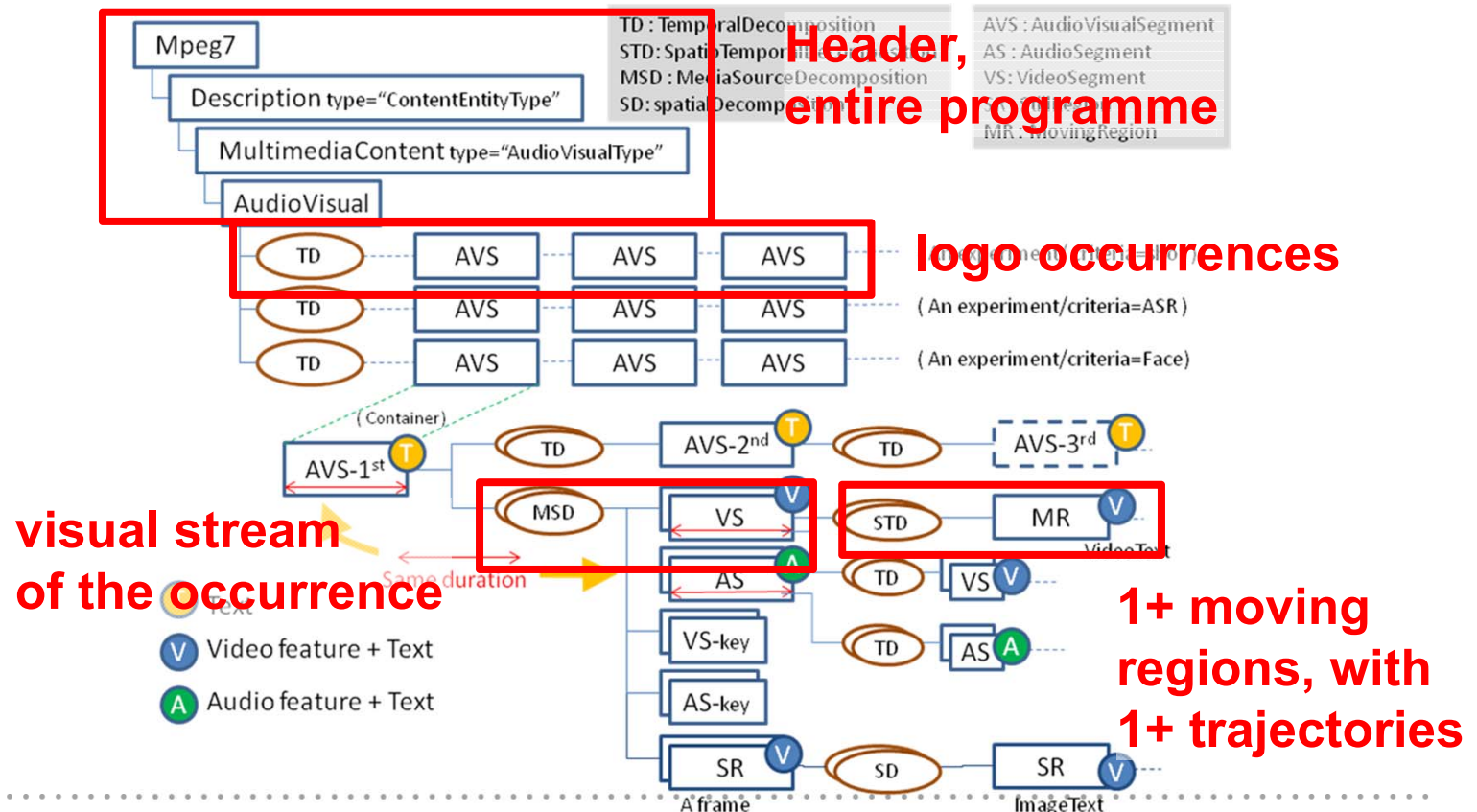
Logo detection

- Detect known logos in video
- Annotate logo occurrence and regions
- Similar to many other applications, e.g. describing
 - Faces, persons
 - Objects
 - Video text



Logo detection

MPEG-7 AVDP structure

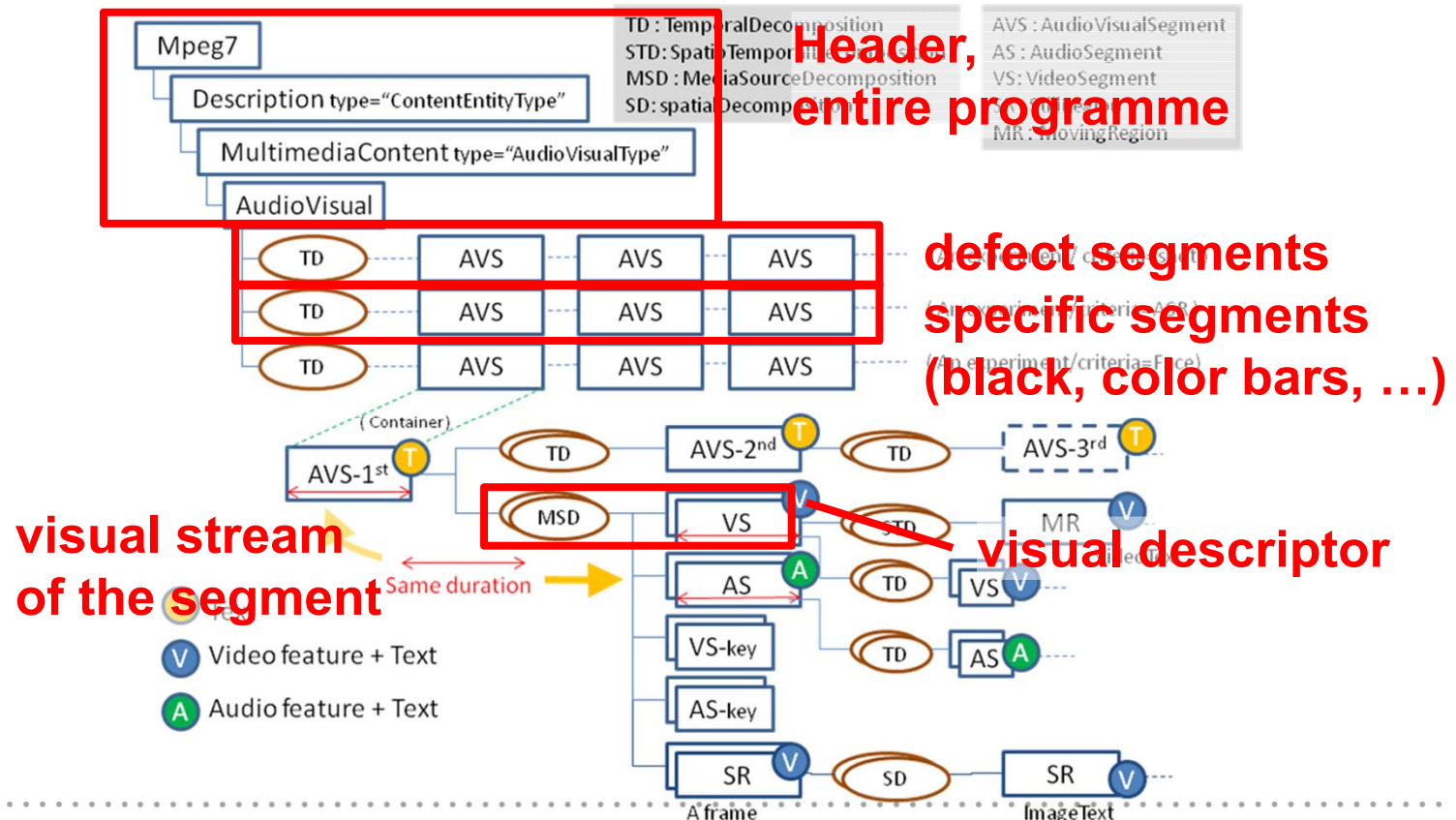


Visual quality analysis

- Detect specific content properties
 - monochrome frames, test patterns
 - Detect defect events
 - drop outs, freeze frames, dust/dirt, ...
 - Describe impairments of segments
 - noise level, blockiness, ...
 - Uses built-in descriptors and extension schema
 - conformance based on MPEG-7 part 7
 - additional descriptors extend VisualDescriptor DS
-

Visual quality analysis

MPEG-7 AVDP structure



Part III

Working with AVDP

Resources

24

- Example documents
- Transforming metadata to MPEG-7 AVDP
- Handling MPEG-7 documents in applications
- Validation

Examples

25

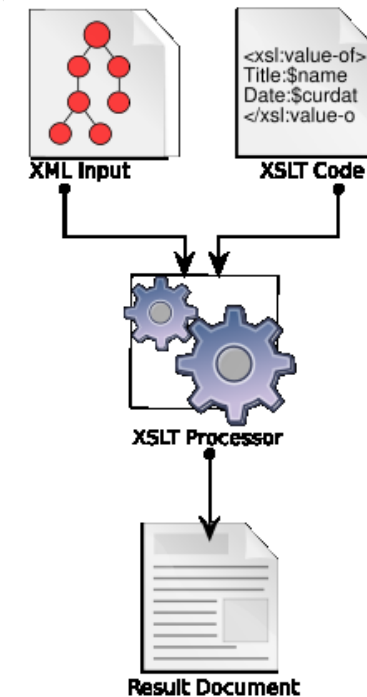
- Example AVDP documents
 - set of basic examples from specification
 - example instances created with tools from different organisations (various content analysis results)
- MIM/MDN Knowledge base
 - <http://workspace.ebu.ch/display/ecmmdn/Knowledge+base+-+contributions>
- MIM/SCAIE guidelines document
 - using AVDP for representing results of automatic content analysis tools
 - to be published by end of 2012

Transforming metadata to MPEG-7 AVDP

- many ways to transform the descriptions
- scripting languages (Perl, Python, etc.)
- writing a program in your favourite programming language
- ...
- if the output of the information extraction tool is XML, use XSLT

XSLT Processing

- XSL = Extensible Stylesheet Language
 - <http://www.w3.org/Style/XSL/>
- XSLT = XSL Transformations
 - <http://www.w3.org/TR/xslt>
 - version 1.0 has limitations, but wide range of tools available
 - version 2.0 supports XPath 2.0, type model based on XML schema, but lacks support in tools
 - apply an XSL to transform a XML document into another XML document, plain text, ...
 - templates match elements/structures in the input document and produce output



http://en.wikipedia.org/wiki/File:XSLT_en.svg

XSLT Processors (1)

■ Saxon

- <http://saxon.sourceforge.net/>
- reference implementation of XSLT 2.0
- Java and .NET

■ Xalan

- <http://xalan.apache.org/>
- Java and C++

■ command line tools and APIs

■ useful for automation

XSLT Processors (2)

- XSLT 1.0 implemented in browsers
- XSLT support in many XML editors
- very basic:
 - XML Notepad, Cooktop (free version no longer supported)
- commercial products, partly with some XSLT debugging support
 - XML Spy, Oxygen, Stylus Studio, ...

XSL Transforms

- Step-by-step example of custom XML → AVDP
 - Tutorial @ MDN workshop 2011
 - Transforming speech to text output
 - Tools, XSLs, intermediate results
 - See MDN workshop 2011 materials:
http://tech.ebu.ch/events/metadata_workshop11
 - Transforms between MPEG-7 DAVP and AVDP profiles
 - MIM/MDN knowledge base
 - Starting from these XSLTs useful for common tasks
 - building/changing hierarchies
 - absolute ↔ relative time, start+end ↔ start+duration
-

Handling MPEG-7 documents

- possible with any API for XML documents, such as SAX or DOM, e.g., Apache Xerces
- large schema, many types, heavy use of derived types
 - i.e., xsi:type attribute defines the actual type
 - cf. static vs. dynamic type in object oriented programming
- inconvenient with APIs unaware of the schema, and no type safety
- need to handle patterns (e.g. time points)

MPEG-7 Library

- C++ class library for Windows, Linux and MacOS X
- Open source (LGPL)
- Parts 2, 3, 4, 5 of MPEG-7 v1 and v2
- XML serialisation/parsing
- typed access to nodes
 - including access to elements of pattern types, e.g. fields of time points and durations

MPEG-7 Library

- XPath support

- as schema aware, also create subtrees by Xpath

- Extensible

- extensions of schema (cf. visual quality use case)
- own implementations of certain types, e.g.
implementing conversion to/from application specific
types
- example: extension for describing visual quality
analysis results

MPEG-7 Library

- <http://mpeg7.joanneum.at>
 - download & documentation
 - guide to generate a similar API for Java using XMLBeans
 - pointers to other MPEG-7 resources
- used in MPEG MXM (MPEG Extensible Middleware) GenericMetadataEngine
 - <http://mxm.wg11.sc29.org/>

Simple example

...

```
Mp7JrsArchive archive;
```

```
archive.SetEncoding("UTF-8");  
archive.SetDoValidation(false);
```

```
Mp7JrsNodePtr root = archive.FromFile("test.xml");
```

```
archive.ToFile("out.xml", root);
```

...

Example: create from XPath

```
Mp7JrsNodePtr root = archive.FromFile("test.xml");  
  
Mp7JrsNodeEnum imageEnum = root->  
  GetOrCreateFromXPath(X(  
    "Description[@xsi:type=\"ContentEntityType\"]/  
    MultimediaContent[@xsi:type=\"ImageType\"]/Image"  
  ));
```

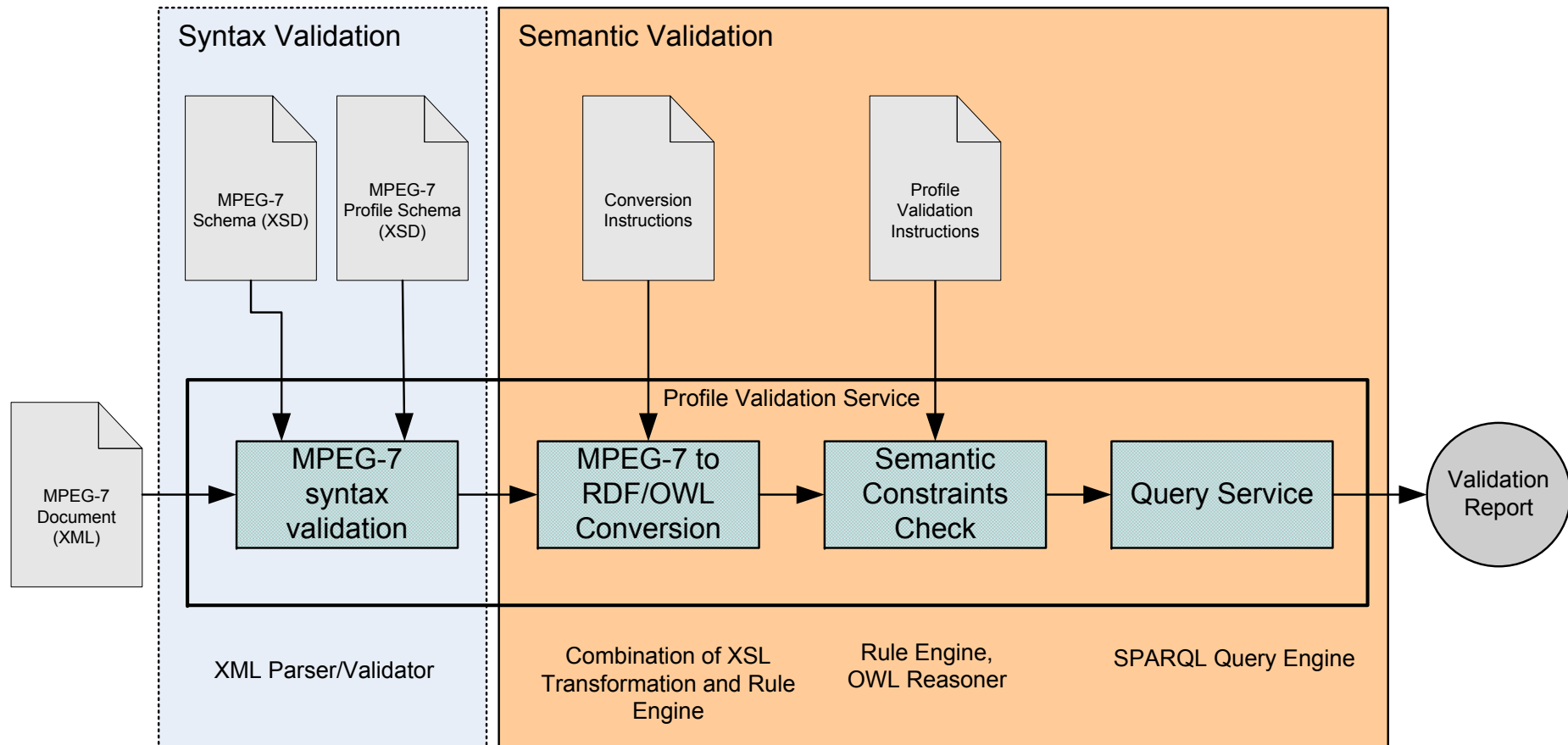
Validation

- Validation against MPEG-7 XML Schema or Profile XML Schema
 - e.g., using validator of an XML editor, Xerces, MS XML, etc.
- many constraints of a profile cannot be represented in XML schema
 - e.g., the AudioVisual element, representing the entire content (root a/v segment) is of type AudiovisualSegmentType
 - the same type is used to represent many audiovisual segments at other places in the description
 - the specific constraints on the elements that need to be present on the root a/v segment but not on others cannot be checked

VAMP: Validation of MPEG-7 Profile Descriptions

- validate metadata documents
 - conforming to complex metadata standards such as MPEG-7
 - semantics of temporal annotations
 - beyond XML schema validation
- prevent interoperability problems
- support of MPEG-7 profiles
 - Detailed Audiovisual Profile (DAVP), PrestoSpace project
 - Audiovisual Description Profile (AVDP)
 - TRECVID format
- <http://vamp.joanneum.at>

VAMP – Workflow



VAMP Web Interface (1)

semantic
VAMP

A Semantic Validation Service for MPEG-7 Profile Descriptions

[validator](#)
[references](#)
[contact](#)
[disclaimer](#)
[FAQ](#)

1. Type the MPEG-7 Document URL:

or [use the following demo example](#)

2. Select MPEG-7 version:

MPEG-7 v1 (2001) MPEG-7 v2 (2004)

3. Select profile:

DAVP TRECVID AVDP

4. Select semantic validation type:

Profile validation (default) Temporal validation

VAMP Web Interface (2)

Result:

Validation failed: Profile Semantic Error

The input MPEG-7 document **does not conform** to the selected profile.

Following errors have been detected:

DescriptionMetadataType header must not be empty

DescriptionMetadataType header that contains only confidence is only allowed in segment

In profile AVDP the element <Header> with type DescriptionMetadataType that contains only <Confidence> may not be used if the parent element of the header is not a segment.

/Mpeg7/Description[2]/MultimediaContent/AudioVisual/TemporalDecomposition[1]/Header

MediaInformation MustBeUsedInRootSegment

mediaTimeBase invalid xpath must reference MediaTime or MediaLocator

mediaTimeBase xpath in root segment must reference MediaLocator in same description

VAMP Web Interface (3)

```

</MultimediaContent>
</Description>
-<Description xsi:type="ContentEntityType">
-<MultimediaContent xsi:type="AudioVisualType">
  -<AudioVisual id="prg02" mediaTimeBase="/Mpeg7/Description[1]/MultimediaContent/AudioVisual/MediaInformation/MediaProfile/MediaInstance/MediaLocator"
    <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/AVDP/StructuralUnitCS:1" />
  -<MediaTime>
    <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
    <MediaIncrDuration>5000</MediaIncrDuration>
  </MediaTime>
  -<TemporalDecomposition criteria="http://www.ebu.ch/metadata/cs/mpeg/AVDP/DecompositionCS:20" id="shotdecomp1">
    -<Header xsi:type="DescriptionMetadataType">
      <Confidence>0.77</Confidence>
    </Header>
  -<AudioVisualSegment id="shot-01">
    <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/AVDP/StructuralUnitCS:10" />
  -<MediaTime>
    <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
    <MediaIncrDuration>1000</MediaIncrDuration>
  </MediaTime>
  -<MediaSourceDecomposition criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS:12">
    <!-- This decomposition is for having general AudioSegment, VideoSegment, StillRegion -->
  -<VideoSegment id="shot-01v">
    <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS:3" />
  -<MediaTime>
    <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
    <MediaIncrDuration>1000</MediaIncrDuration>
  </MediaTime>
  <!-- You can describe video feature of this VideoSegment here -->
  -<VisualDescriptor xsi:type="DominantColorType">
    <SpatialCoherency>31</SpatialCoherency>
  -<Value>
    <Percentage>31</Percentage>

```

VAMP Client

- Java client application
- no need to upload files
- enables batch processing
- <http://vamp.joanneum.at/client.html>
- sample command line for semantic and temporal validation against AVDP

```
java -jar VAMPClient.jar -if mydocument.xml  
-v v2 -pt AVDP -svt both
```

- outputs vamp_summary.html and an HTML file with detail results for each file validated
-

Other useful resources

- MaP7 MPEG-7 schema search engine (v1 only)
 - <http://www.tom.comm.waseda.ac.jp/map7/>
- very useful XSLT reference
 - <http://zvon.org/comp/m/xslt.html>

Questions?



TOSCA-MP

The research leading to these results has received funding from the European Union's Seventh Framework Programme under the grant agreements no. 215475, "2020 3D Media – Spatial Sound and Vision" (<http://www.20203dmedia.eu>), no. 231161, "PrestoPRIME" (<http://www.prestoprime.eu>) and no. 287532, "TOSCA-MP" (<http://tosca-mp.eu>).

