Content Quality Assessment and Metadata Interoperability for Preservation of Audiovisual Media

Kurt Majcen, Peter Schallauer, Werner Bailer, Martin Winter, Georg Thallinger, Werner Haas
JOANNEUM RESEARCH
DIGITAL – Institute of Information and Communication Technologies
Steyrergasse 17, 8010 Garz, Austria, {firstName.lastName@joanneum.at}

Summary

There are million hours of audiovisual content in collections of broadcaster archives, libraries and museums. Handling the content is undergoing a major change when migrated to files: invisible, stored “somewhere” and played by black box technology. Long-term digital preservation of the content is a pressing need and ensuring integrity is the basis for access and re-use.

Automatic quality control for audiovisual media is an important tool for broadcasters, archives and content providers in media production, delivery and preservation processes but also for post-production houses when estimating restoration costs. Today mainly technical properties of the material are checked automatically. Other relevant content properties or impairments are manually checked. We work on automatic reference-free visual quality detectors for audiovisual files. Our recent research results are about detecting significant visual distortions, so-called video breakups.

The heterogeneity between workflows and metadata models in audiovisual archives leads to various metadata models and standards. Further models exist for presentation and use of metadata on web portals like Europeana. We give an overview of existing models and recent developments of necessary metadata mappings. Our ontology driven approach eases the mapping for a larger number of models. Further we check validity of metadata against the standards they adhere to by semantic approaches complementing today’s syntactical standard validations.

1 Introduction

The huge amount of audiovisual content available in collections at broadcaster archives, libraries and museums is moved from analogue to digital representation. Significant changes take place when the content is migrated to files: content becomes somehow invisible, it will be stored “in the cloud or somewhere else” and playing it is done by black box technology. Consequently long-term digital preservation of the content is a pressing need and ensuring integrity of content as well of its descriptions is the basis for allowing access on the material and its re-use.

Automatic quality analysis is one important step in several workflows (production, delivery and preservation) for assisting and improving processes. One kind of impairments – video breakup – and a solution how to detect it are shown in chapter 2. Evaluation results are also part of that chapter.

Outcome of the aforementioned quality analysis are metadata (detected defects) which are a small part of metadata existing around audiovisual content. Other information attached to media are created through several processes (production and delivery). They deal with description of the material, technical details and information on rights and provenance. Especially preservation activities and the re-use of material or parts of it create a lot of additional metadata. Various metadata models or standards exist. Those and services like mapping between representations and semantic validation are described in chapter 3.

2 Quality analysis task “Video breakup”

Quality analysis of audiovisual content is an important task for several steps of the media production, delivery and archiving process. Content providers are checking post production content for correct encoding and conformance to required quality and format standards before dispatching to the broadcasters or service providers. Broadcasters are checking audio and video quality of material during ingest, after editing / encoding and before play-out (broadcast or delivery to internet). Within the digital video and film preservation application domain the results of content based quality analysis aim at improving efficiency of various archive related processes. They check the content integrity during ingest processes, perform 'best copy selection' when multiple copies of material are available and provide minimum quality service for archive accesses.

Although many applications are feasible mainly the technical properties of the video material are checked (e.g. stream compliance, GOP structure, play time, aspect ratio, resolution or MXF compliance). Thus automatic content based quality control is currently limited to a few simple measures as e.g. amount of blocking, detecting luma/chroma violation or rough noise level estimation. Other relevant content properties and impairments are usually manually checked (labour intensive and expensive).

Only limited research regarding automatic video breakup detection for quality control was done so far. The most related paper is (Wang & Li, 2009) describing temporal smoothness constraints of consecutive video-frames' wavelet decomposition. So we focus on the detection of the 'severe visual distortions' commonly known as 'video breakup'. Main origins are: analogue errors typically caused by tape transportation, video signal transmission problems and digital errors often caused by broken video streams.

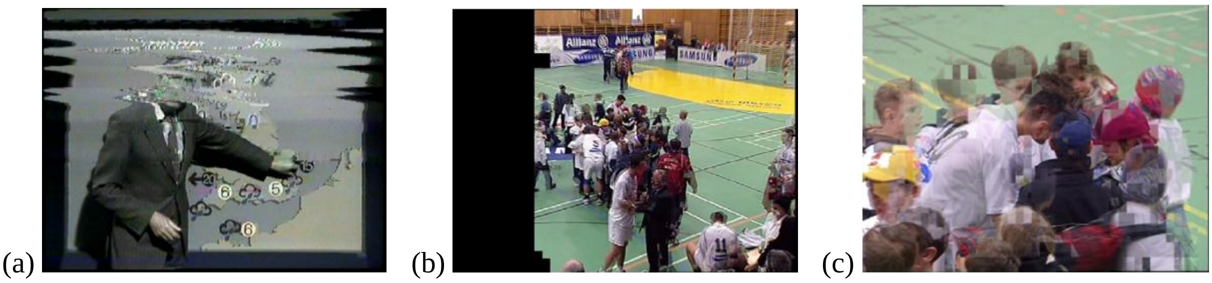


Figure 1: Some typical examples for video breakup impairments caused by analogue (a) and digital (b, c) sources.

Basic application requirements have to be understood when designing impairment detection algorithms for purpose of content based video quality analysis as defined for file based environments below.

- **Algorithms should work (semi-)automatic.** Full manual inspection is at least a real-time process (cost intensive). With human concentration weakening over time, quality of inspection decreases and 'objective' judgment of video quality is not guaranteed over the whole period of inspection.
- **Runtime.** Broadcast delivery services need quality inspection 'on the fly' and therefore strict real-time performance of algorithms. Although most archive and broadcast related processes allow online quality inspection and thus algorithms’ runtime for analyzing single videos being not that critical, the amount of content in audiovisual archives and broadcast production processes requires a high overall throughput. This is either achieved by efficient or by parallelized algorithms.
- **Detection rate should be as high as possible and the number of erroneous detections (wrong detections, 'false positives') low.** Many false alarms annoy the operator and decrease acceptance of the system but also limit the time saving (and thus cost benefit) gained by applying the algorithm.
- **Software solutions where possible.** Seamless integration, extensibility and flexibility are higher than with hardware solutions.
- **Raw video data preferred.** This avoids video encoding dependencies (e.g. encoder specific properties).
- **Abstracted compact analysis results.** E.g. time code of occurrence, defect class and its strength allow compact visualization - a pre-requisite for efficient human inspection of analysis results.

2.1 ‘Video Breakup’ detection algorithm

The appearance of the 'video breakup' defects varies substantially and therefore it is difficult to identify common patterns for reliable, severe distortion detection. Therefore instead of explicitly modelling the defect itself we inverted the problem and model the normal 'video sequence behaviour'. 'video breakups' are detected by violation of its continuous motion constraints.

As mentioned impairment detection algorithms are subject to several design criteria. Especially the run-time constraints prohibit the design of complex and therefore time consuming algorithms. Fortunately 'video breakup' distortions significantly change their location and appearance from frame to frame (usually not the case for regular moving objects).Thus even simple pixel differences between consecutive frames can distinguish between normal object motion within the content and abrupt content changes induced by 'video breakup'. Only exceptions to that are fast motion sequences. This leads to a very simple indicator for the presence of 'video breakup' events with surprisingly good performance results.

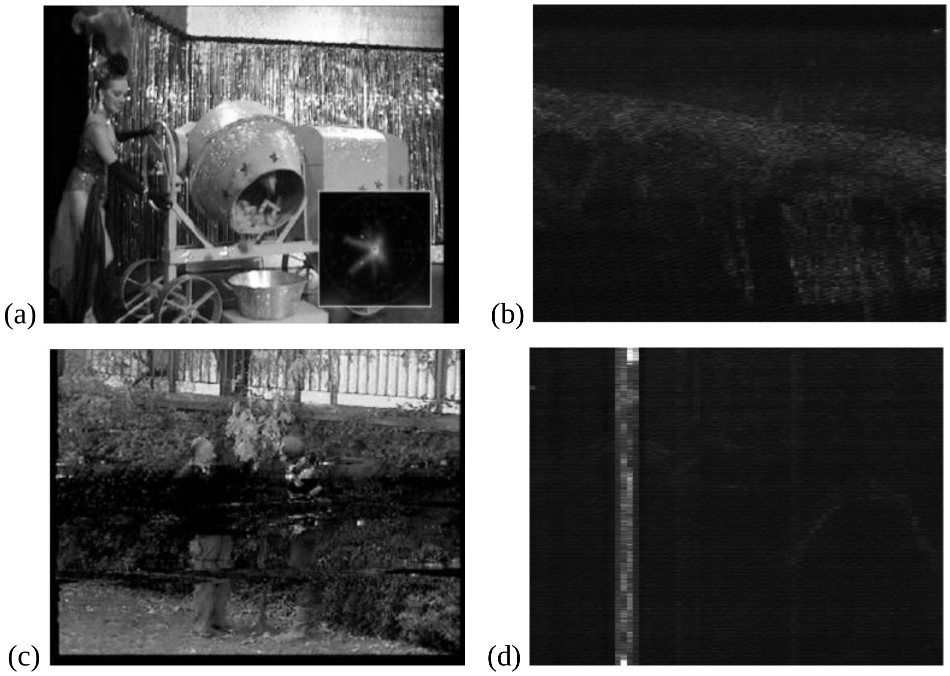


Figure 2: Some examples for 'stripe-difference-images' obtained (b,d) from a scene containing regular motion of a rotating, downwards moving concrete mixer(a) and a scene containing a typical video breakup impairment(c).

A 'stripe-difference-image' representation is obtained between consecutive frames by cumulating the intensity difference values in each row e.g. as shown in Figure 2 (b) and (d). Regular (downwards) motion of the concrete mixer causes only slight changes from column to column (b). 'video breakups' as e.g. shown in Figure 2(c) induces heavy disturbances. Hence, the "*H6*" distance measure (van der Weken et al. 2003), calculated between consecutive columns of the 'stripe-difference-image', allows estimation of the probability of video breakup presence for each frame of the video.

This basic 'video breakup' measure proposed occasionally fails in scenes with heavy, suddenly appearing illumination changes (e.g. caused by ash lights or other strong illumination changes). A second measure based on the ratio *R* between vertical and horizontal edges on the difference image was added. The averaged change of *R* within a 5 frames interval indicates a 'video breakup' event (assuming that the amount of horizontal and vertical edges changes equally for extensive illumination changes, while being not the case for analogue 'video breakups').

To overcome the limitations with fast motion sequences (false detections), the motion between two consecutive frames is estimated and used for warping the image to a motion compensated image. Thus the motion compensated difference image ideally contains only the not motion compensable part between two consecutive frames. This is usually significantly higher if 'video breakup' defects occur. Optical-flow calculations - although very time-consuming tasks – are nowadays used for applications with real-time requirements as well due to the highly parallelized architecture of graphic processors (GPU) providing sufficient run-time capabilities even on full standard definition resolution videos (Zach et al. 2007).

2.2 Evaluation

The performance of our algorithm was evaluated with 51 videos (452 minutes challenging content – extremely fast motion, heavy luminance changes, noise, water etc.) containing several hundreds of video impairments. For an objective judgement, the video breakups have been annotated by experts from a local video producing company. Exact location and a 'subjective' strength (6 levels) have been recorded for each video breakup. The recall-measure $R = TP/(TP+FN)$ was used for evaluation (with TP the number of 'correct detections' and FN the number of 'missed detections'). It estimates the fraction of impairments correctly detected overall. The false positive rate $FPR = FP/t$ was used for estimating the precision of the detection (FP the number of erroneous detections and t the time interval) and reflects the displeasedness felt by an operator, when obtaining many erroneous detections.

mode	working point	recall	FPs per min
STD	1	42.70	0.098
	2	83.33	0.583
	3	93.03	1.044
	4	97.24	2.202
ACC	1	53.85	0.106
	2	82.23	0.437
	3	94.23	0.776
	4	97.29	1.518

Table 1: Preset working points and corresponding recall and precision values for the basic (STD) and motion compensated (ACC) algorithm. (1 = high recall, 2 = default, 3 = low false positives, 4 = extreme low false positives)

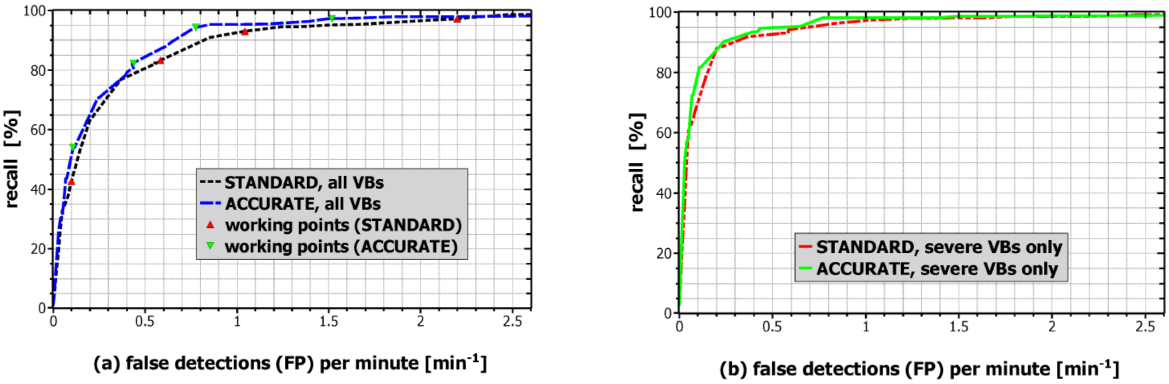


Figure 3: Performance results for the basic (STD) and motion compensated version (ACC) of our video breakup algorithm on a dataset of 51 videos (about 452 minutes). (a) results taking into account all annotated video breakups in ground-truth, (b) the results only for the strongest (= subjective significant) video breakups found

Figure 3(a) shows the overall performance results for the basic (STD) and motion compensated version (ACC) of our Video Breakup algorithm. The overall difference between the basic and motion compensated version characteristics is not that high, but for higher recall rates (above e.g. 95%) the difference in the 'precision' measure and thus the performance gain with respect to the erroneously detected video breakup events is evident (0.8 vs. 1.3 false positives min^{-1}). An operator can easily select desired analysis systems' behaviour with one of 4 preset working points denoted in Figure 3(a) as triangles and explicitly listed in Table 1. Hence it is possible to reach nearly 100% recall (R) but with a significant difference in false detection rate. The performance results when only strongest video breakups are taken into account depicted are shown in Figure 3(b). Results are significantly better, although the differences between basic version and motion compensated version of our algorithms are smaller.

Speed tests done on an Intel Core2 Quad CPU (2.4GHz) with 3GB of RAM and a NVIDIA GeForce GTX 285 (240 unified shaders) on standard (SD, 720 x 576) video material showed that the basic version processing is faster than real-time and thus a candidate for use in real-time requirement scenarios.

3 Metadata for audiovisual content

A large number of standards for representing audiovisual metadata exist. They come from different organisations and cover diverse application areas. Practically all multimedia file formats include technical metadata of the content. This is especially true for container formats that combine essence and associated metadata in one file. The Material Exchange Format (MXF, 2004) is an example for such a container format. Another example for a container format is the Digital Cinema Package (DCP)[1], used to transport digital movies and associated metadata to cinemas, using MXF as essence container. Mapping between metadata models and validation are inevitable in practical applications.

3.3 Models and standards

The *Dublin Core* metadata standard[2] was originally developed to describe electronic text documents but has later been extended to also cover audiovisual material. Focusing on simplicity it contains fifteen elements belonging to three groups (content, version and intellectual property). The full set of elements is an elaborate extension of the original 1.1 version and is called the DCMI Metadata Terms[3].

The *EBU Core* set of metadata aims to define the minimum information needed to describe radio and television content (EBU Core, 2009). The EUscreen[4] project will provide access to distributed audiovisual heritage with mechanisms built on EBU Core and open web standards.

The ISO/IEC standard *Multimedia Content Description Interface (MPEG-7)* (MPEG-7, 2001) is format defined for the description of multimedia content in a wide range of applications. MPEG-7 defines a set of description tools, called descriptors (single properties of the content description) and description schemes (containers for descriptors and other description schemes). A core part of MPEG-7 are the Multimedia Description Schemes (MDS), which provide support for the description of media information, creation and production information, content structure, usage of content, semantics, navigation and access, content organisation and user interaction. Content can be described very flexible and on different levels of granularity. Profiles are defined subsets of the standard which target certain application areas. The “Audiovisual Description Profile” (AVDP) for applications in media production and archiving, developed by the EBU ECM SCAIE work is currently under standardisation. It is based on the earlier proposed Detailed Audiovisual Profile (Bailer & Schallauer, 2006) and the Media Production Framework (NHK, 2008).

The European Broadcasting Union (EBU) has defined the metadata vocabulary *P_Meta*[5] for programme exchange in the professional broadcast industry (business-to-business use case). P_Meta consists of a number of attributes (some of them with a controlled list of values), which are organised in sets.

Beside the representation with a certain metadata format, audiovisual content descriptions often contain references to semantic entities (e.g. objects, events, places, and times). For ensuring consistent descriptions (e.g. persons always referenced with the same name) controlled vocabulary are in use. The simplest controlled vocabulary is a value list for a property (e.g. ISO 3166 list of countries). Taxonomies are defined as (multilingual) tree structured terms and simple relations between them. Thesauri are (poly-) hierarchical structures of terms in a given application domain. They allow more complex relations (e.g. hierarchy of terms, synonyms, related terms) and can be also multilingual. Controlled vocabulary information can also be represented with ontologies and standards such as RDF/OWL or SKOS[6].

[1] http://www.dcimovies.com/

[2] http://dublincore.org/documents/dces/

[3] http://dublincore.org/documents/dcmi-terms/

[4] http://www.euscreen.eu/

[5] http://tech.ebu.ch/docs/tech/tech3295v2.pdf

[6] http://www.w3.org/2004/02/skos/

The *Media Annotations Working Group (MAWG)* of the World Wide Web Consortium[7] (W3C) develops the so called Ontology for Media Resource to become a core vocabulary for describing media resources on the Web. The further defined semantics shall preserve mappings between existing formats. A prototype[8] was implemented for the also provided API.

3.4 Metadata mapping

Exchange of metadata is the key to ensuring access to audiovisual collections. Metadata exchange is often hindered by the diversity of used metadata formats. Thus metadata interoperability is needed. Several scenarios related to audiovisual preservation require mapping between metadata representations:

- Conversion of legacy technical metadata in preservation scenarios
- Access to legacy descriptions for performing mapping between material and editorial entities
- Extracting metadata embedded in file headers and convert it to the data structures needed in a source/or archive information package of a OAIS (CCSDS, 2010) compliant system
- Ingest of metadata from non-/semi-professional content creators
- Content provision to Europeana[9] or similar portals
- Outsourcing of annotation and access services, with potentially different data models involved

Metadata interoperability needs a solution on two levels: on syntactic and on semantic level (metadata interpretation within the semantic context, e.g. metadata's meaning in one archive needs to be linked to metadata from other archives). While the multimedia metadata formats overlap in a core set of covered metadata properties, they are at the same time dissimilar in many ways (e.g. domains, comprehensiveness and complexity). Due to these differences mappings can only be partial in many cases (e.g. properties only supported in one of the formats, for formats not allowing extensions, information can be lost during the mapping steps etc.).

Our goal was that the mapping tool must provide a best practice mapping of the input metadata document (in some supported XML format) to another supported XML representation. The resulting XML must be valid w.r.t. to the standard and must convey the semantics of the input document as far as possible. For each supported format, it is assumed that the mapping tool has access to the following definitions:

- A formal description of the set of system-wide supported metadata elements and their relations.
- A formal description of the set of metadata elements in the format and their relations among them (if applicable) and their relations to the set of elements supported by the system.
- A set of mapping rules to be applied.

There are some difficulties and limitations which can be anticipated:

- For some pairs of formats, mapping will be lossy in one direction, and will have only loosely defined semantics in the opposite direction. Thus bidirectional mappings are incomplete.
- There are ambiguities in mappings that need default or use case specific rules to be resolved.
- If formats use specific data types, their mappings to (structures of) standard data types needs to be defined. This will reduce the saving in the number of mapping definitions needed.

Our approach uses a high-level intermediate representation of metadata elements. Metadata elements from specific metadata formats are described in relation to these generic elements. Further mapping templates are used on data type level. The mapping between a pair of formats is derived from these sources. We extended an existing ontology for our purpose, to represent the formal semantics of the high-level intermediate representation. Deriving mapping instructions from the ontology eases maintenance of

[7] http://www.w3.org/2008/WebVideo/Annotations/

[8] http://mawg.joanneum.at

[9] http://www.europeana.eu/portal/

mapping instructions against hard coded ones. The high-level intermediate representation further serves as a hub for mapping between formats and hand-crafted one-to-one mappings between pairs of formats are avoided. Mappings can be created automatically. Adding new formats does not have side effects.

The core of this approach is the *meon* ontology (Höffernig & Bailer, 2009) which describes generic metadata elements and the relations between them. *meon* was originally developed to model metadata elements used throughout the audiovisual media production workflow in a format independent way in order to support content exchange and automation. The *meon* ontology has been extended to express mapping relations between metadata formats. In addition to the ontology of generic metadata concepts, specific ontologies are created for each format. They follow the same pattern as the *meon* ontology and include relations between format specific and generic concepts. Mapping instructions for low-level issues (e.g. conversion between data types) are defined as well.

A prototype was implemented following this approach (ontology expressed in OWL-DL and mapping instructions encoded as XSL transformations[10]). The workflow is: metadata concepts and mapping relations in the ontology are annotated with additional information needed for mapping (i.e. format specific metadata, corresponding XPath and information on data type representation). Using reasoning techniques over the ontology data type templates are inferred for the mapping between different format specific concepts. A SPARQL[11] query retrieves all necessary mapping information from the ontology allowing generating XSL templates. Result is an XSL template document with mapping instructions.

3.5 Metadata validation

Validation is important for metadata documents when produced, exported or imported. Standard tools on syntactic level are XML schema validators but the semantic expressivity of XML schema is limited. Thus validation on higher level can be only done with specific application logic. Our approach for validating metadata on a semantic level is based on a description of formal semantics of the metadata format.

The web based application VAMP[12] is one example for validating semantic conformance to MPEG-7 profiles (Figure 4). It is also available as a web service allowing embedding the validation into any application. A REST-style web service interface is provided for the validation service.

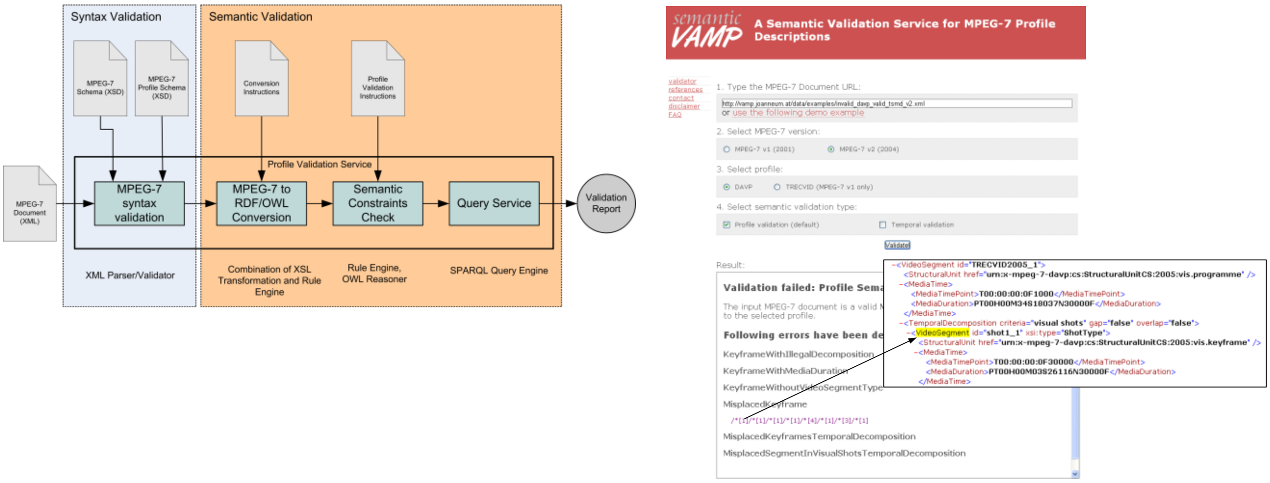
4 Conclusion and future work

We showed a novel algorithm for the detection of severe visual distortions in videos. Depending on the domain of application it is possible to parameterize the algorithm meeting specific real-time requirements in standard definition (SD) resolution. For a lower rate of erroneous detections an improved version of the algorithm uses optical-flow for motion compensation on the GPU. An extensive evaluation on an expert's annotated, huge database shows, that we can detect up to 97.3% of annotated video breakups and we can reach a false detection rate of only 0.1-1.5 per minute. Future work on content based video quality assessment will focus on robustness of this algorithm but we will also concentrate on further preservation related impairment detectors like freeze frame detection or noise level estimation.

[10] http://www.w3.org/TR/xslt20/

[11] http://www.w3.org/TR/rdf-sparql-query/

[12] http://vamp.joanneum.at



A Semantic Validation Service for MPEG-7 Profile Descriptions

1. Type the MPEG-7 Document URL:
http://vamp.joanneum.at/data/examples/invalid_davp_valid_tsmd_v2.xml
or use the following demo example

2. Select MPEG-7 version:
○ MPEG-7 v1 (2001) ● MPEG-7 v2 (2004)

3. Select profile:
● DAVP ○ TRECVID (MPEG-7 v1 only)

4. Select semantic validation type:
☑ Profile validation (default) □ Temporal validation

Result:

Validation failed: Profile Sema

The input MPEG-7 document is a valid M
to the selected profile.

Following errors have been de

KeyframeWithIllegalDecomposition
KeyframeWithMediaDuration
KeyframeWithoutVideoSegmentType
MisplacedKeyframe
/*[1]/*[1]/*[1]/*[1]/*[4]/*[1]/*[3]/*[1]
MisplacedKeyframesTemporalDecomposition
MisplacedSegmentInVisualShotsTemporalDecomposition

```
-<VideoSegment id="TRECVID2005_1">
   <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.programme" />
  -<MediaTime>
     <MediaTimePoint>T00:00:00:0F1000</MediaTimePoint>
     <MediaDuration>PT00H00M34S18037N30000F</MediaDuration>
   </MediaTime>
  -<TemporalDecomposition criteria="visual shots" gap="false" overlap="false">
    -<VideoSegment id="shot1_1" xsi:type="ShotType">
       <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.keyframe" />
      -<MediaTime>
         <MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
         <MediaDuration>PT00H00M03S26116N30000F</MediaDuration>
       </MediaTime>
```

Figure 4: VAMP validation workflow (left) and web user interface (right).

For the great diversity of metadata formats and standards we develop mapping and validation services. Our mapping service allows easy adding of new formats and flexible follow up of potential changes on already implemented mappings. The validation service enables semantic analysis in addition to the traditional pure syntactical validation mechanisms. Future work on metadata services will complete the prototyped mappings, will add new formats and will evaluate the services in other cultural heritage domains.

5 Acknowledgement

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231161 (project PrestoPRIME).

6 References

Bailer W. & Schallauer P. (2006). The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 based Systems, 12th International MultiMedia Modelling Conference, pp. 217-224.

EBU Core (2009). EBU-TECH 3293: EBU Core Metadata Set (EBU Core).

Höffernig M. & Bailer W. (2009). Formal Metadata Semantics for Interoperability in the Audiovisual Media Production Process, Workshop on Semantic Multimedia Database Technologies (SeMuDaTe).

MPEG-7 (2001). Multimedia Content Description Interface, ISO/IEC 15938.

MXF (2004). Material Exchange Format (MXF) – File Format Specification, SMPTE 377M.

NHK Science and Technical Research Laboratories (2008). Metadata Production Framework Specifications (v. 2.0.2E), http://www.nhk.or.jp/strl/mpf/english/index.htm

CCSDS (2010). Reference Model for an Open Archival Information System (OAIS), Blue Book. http://public.ccsds.org/publications/archive/650x0b1.pdf. Last visited: 2010 May 01.

Schallauer P., Bailer W., Mörzinger R., Fürntratt H., Thallinger G. (2007). Automatic quality analysis for film and video restoration. In IEEE ICIP, San Antonio, USA.

van der Weken D., M. Nachtegael, Kerre E. (2003). Using similarity measures for histogram comparison. Lecture Notes in Computer Science, 2715:1-9.

Wang Z. & Li Q. (2009). Statistics of natural image sequences: temporal motion smoothness by local phase correlations. Human Vision and Electronic Imaging XIV, 7240:72400W.

Zach C., Pock Th., Bischof H. (2007). A duality based approach for real-time tv-l1 optical flow. In Proceedings of the 29th DAGM Symposium on Pattern Recognition, pages 214-223.