The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems

Werner Bailer, Peter Schallauer
JOANNEUM RESEARCH, Institute of Information Systems and Information Management
Steyrergasse 17, 8010 Graz, Austria
{werner.bailer, peter.schallauer}@joanneum.at

Abstract

MPEG-7 is an excellent choice for the description of audiovisual content due to its flexibility and comprehensiveness. The drawback is that these properties also increase the complexity of descriptions and cause ambiguities which hinder interoperability. In order to partly solve these problems, profiles and levels have been proposed, but the definitions of the adopted profiles lack semantic constraints which are necessary for interoperability. We propose a profile for detailed description of audiovisual content that can be used in a broad range of applications. The profile aims at describing single multimedia content entities, allowing a comprehensive structural description of the content, including also audio and visual feature descriptions. The profile defines a set of semantic constraints on the selected tools, which resolve ambiguities in modeling the description and support system interoperability. The proposed profile has been successfully used in a wide range of applications.

1. Introduction

In many applications and systems dealing with multimedia, content description using different kinds of metadata is an important issue. MPEG-7 [8], formally named Multimedia Content Description Interface, has been designed as a standard for such descriptions in a broad range of applications. In order to cover their diverse requirements, many description tools in the standard have been defined in a generic and flexible way.

This generality leads to a large number of description tools and increases the complexity of descriptions. In some application scenarios, only a small part of the functionality is needed, so that the complexity is experienced as a burden. Profiles and levels have been proposed to define subsets of the standard tailored towards certain functionalities and with different levels of complexity [1]. Three profiles have been adopted for standardization [16].

1-4244-0028-7/06/$20.00 ©2006 IEEE

In this paper we discuss the detailed and fine-grained description of audiovisual content entities, ensuring interoperability between different systems. MPEG-7 is a suitable instrument for this purpose, but there are some drawbacks, that make the use of profiles desirable. We survey whether the adopted profiles are suitable for this purpose. We further analyze if profiles can facilitate interoperability and if this interoperability can also be ensured on a semantic, and not only on a syntactic level.

This paper is organized as follows: First we describe the benefits and drawbacks of MPEG-7 for the description of audiovisual content and review the current state of MPEG-7 profiling. We then define the application areas, requirements and functionalities of a profile for the detailed description of audiovisual content and present the definition of a novel profile for this purpose. We describe a number of applications in which the profile has been successfully applied and conclude with a discussion about the role of semantics in the use of MPEG-7 and profiling.

2. Using MPEG-7 to describe audiovisual content

2.1. Benefits

For many applications, MPEG-7 is an excellent choice for describing audiovisual content, mainly because of its comprehensiveness and flexibility.

The comprehensiveness results from the fact that the standard has been designed for a broad range of applications and thus employs very general and widely applicable concepts. The standard contains a large set of tools for diverse types of annotations on different semantic levels.

The flexibility of MPEG-7, which is provided by a high level of generality makes it usable for many application areas without imposing strict constraints on the metadata models of these applications. The flexibility is very much based on the structuring tools and allows the description to be modular and on different levels of abstraction. MPEG-7

supports fine grained description, and it provides the possibility to attach descriptors to arbitrary segments on any level of detail of the description. The possibility to extend MPEG-7 according to the conformance guidelines defined in part 7 [3] provides further flexibility. Extensibility is necessary, as no standard can foresee all required tools for specific applications.

2.2. Drawbacks

Two main problems arise in the practical use of MPEG-7 from its flexibility and comprehensiveness: complexity and limited interoperability. The MPEG-7 requirements group has early recognized these issues [7].

The complexity is a result of the use of generic concepts, which allow deep hierarchical structures, the high number of different descriptors and description schemes and their flexible inner structure, i.e. the variability concerning types of descriptors and their cardinalities. As a result, learning MPEG-7 is time-consuming and this may cause hesitance in using the standard in products. It also makes it more difficult to implement tools for working with MPEG-7, and the consequent lack of tools and implementations can contribute to the hesitance mentioned before.

The interoperability problem is a result of the ambiguities that exist because of the flexible definition of many elements in the standard (e.g. the generic structuring tools). There can be several options to structure and organize descriptions which are similar or even identical in terms of content, and they result in conformant, yet incompatible descriptions. Full interoperability is only possible with knowledge about how the standard has been used. This means that an additional layer of definitions is necessary to enable full interoperability between systems using MPEG-7.

3. MPEG-7 profiles and levels

3.1. Concept

Based on the experience from other MPEG standards, profiles and levels have been proposed [1] to solve some of the problems discussed above. Profiles are subsets of MPEG-7 tools which cover certain functionalities, while levels are flavors of profiles with different complexity. As the restriction to certain functionalities reduces the number of tools, profiles also contribute to the reduction of complexity, and thus influence some of the complexity measures proposed in [7], e.g. the number of descriptors and description schemes used.

The definition of a profile consists of three steps [1]: the selection of tools supported in the profile (i.e. the descriptors and description schemes used), the definition of constraints on these tools (e.g. the reduction of cardinality of some elements) and the definition of constraints on the semantics. The selected tools and most of the tool constraints can be formalized using XML Schema [17], so that conformance to a profile can be automatically validated using standard tools on this level.

The semantic constraints further limit the use of the selected tools. However, as these constraints are only defined in textual form, conformance to the profile in terms of semantics cannot be validated automatically.

3.2. Adopted profiles

Several profiles have been under consideration for standardization and in [9], a set of proposed and a set of adopted profiles have been collected. The definitions of the three adopted profiles have been revised in [16] and constitute part 9 of the standard.

The Simple Metadata Profile (SMP) allows describing single instances of multimedia content or simple collections. The profile contains tools for global metadata in textual form only. The proposed Simple Bibliographic Profile is a subset of SMP. Mappings from ID3, 3GPP and EXIF to SMP have been defined.

The functionality of the User Description Profile (UDP) consists of tools for describing user preferences and usage history for the personalization of multimedia content delivery.

The Core Description Profile (CDP) allows describing image, audio, video and audiovisual content as well as collections of multimedia content. Tools for the description of relationships between content, media information, creation information, usage information and semantic information are included. The profile does not include the visual and audio description tools defined in parts 3 and 4 of the standard.

3.3. Profiles and interoperability

Interoperability is an important reason for standardization, and it has been an important goal in the design of MPEG-7 [8]. This interoperability is ensured on a syntactic level by the XML Schema definition, while on a semantic level only a textual explanation in the standard is available.

By selecting tools and defining additional constraints on both tool and semantic level, profiles can support interoperability, which is a key requirement for widespread use of MPEG-7 and thus an important reason for the definition of profiles. To achieve this goal, interoperability has to be considered in the process of profile definition. This can be done mainly by avoiding ambiguities in the description and by restricting the use of the standard, so that there is only one way to model semantically identical descriptions. The first step in the profile definition process, the selection of tools, is only of limited relevance in this

context, as it mainly limits the capabilities of the description. The second step, the definition of constraints on the selected tools is in some cases sufficient to formulate the intended restrictions, i.e. only a limited set of restrictions can be expressed with tool constraints.

The third step, the definition of the semantic constraints is a crucial one for defining a profile. These constraints describe the use of the standard most precisely and are most powerful in avoiding ambiguities.

3.4. The need for a new profile

The profiles that have been defined in part 9 of the standard [16] are not sufficient for many applications in the area considered in this work (cf. Section 4.1). For example, none of the currently defined profiles includes the tools for visual and audio feature description (part 3 and 4), which are necessary for a detailed description of audiovisual content in a number of applications.

As pointed out before, profiles could be an important means to ensure interoperability between applications and systems producing or processing descriptions. It must however be noted, that the adopted profiles have been designed with complexity reduction and not with interoperability in mind. For none of these profiles, constraints beyond tool selection and tool constraints have been defined, so that the problem of ambiguities in the use of some tools is not solved by using the profiles.

It is thus necessary to define a profile for the detailed description of audiovisual content. As this profile shall serve a broad range of applications, interoperability is an important issue. Thus the profile definition must include a sufficient set of semantic constraints. A proposal for such a profile is presented in the following sections.

4. A profile for detailed audiovisual description

4.1. Application areas

The intended use of this profile is the detailed description of image, audio, video and audiovisual content entities. By “detailed” we mean the support of content structuring down to a fine level of granularity and the capability to annotate and describe features on different abstraction levels. The profile is intended to be complementary to metadata standards and container formats that are restricted to global technical and descriptive metadata.

There is a broad range of applications where such a detailed description of audiovisual content is required. This encompasses all kinds of applications that deal with the analysis, description, retrieval, summarization and exchange of audiovisual content. The profile is designed to support the use of automatic and manual annotation tools and content-based query paradigms such as query by example. Possible application areas include audiovisual archives, image and video databases, media monitoring applications, audiovisual content production and educational applications.

4.2. Requirements

One of the main requirements for the envisaged application area is the ability to describe arbitrary fragments of media items. The scope of an annotation may vary from the whole media item to spatial, temporal or spatiotemporal fragments of the media item down to a region of a single image. The definition of these fragments must be flexible enough to allow fragments that are based on audiovisual features (such as image regions representing objects or shots of a video), any higher-level features (e.g. scenes in a video) or user-defined ones. This includes descriptions of different kinds of modalities, descriptions produced with different tools, such as results from automatic content-analysis, semantic interpretation and manual annotation. As there are many possible criteria for structuring, which are based on different input data and which are on different semantic levels, the description should be as modular as possible. The description tools shall not enforce dependencies between semantically independent parts of the description.

Media, creation and usage information are commonly used in the envisaged application area and often the only ones available in legacy information systems. The description of low- and mid-level audio and visual features is required on the corresponding segments of the structural decomposition of the content. The profile has to support efficient browsing, visualization and sonification of descriptions of multimedia content. Summaries, used in connection with the full content descriptions, are a very valuable tool for this purpose.

4.3. The role of semantic constraints

Apart from the selection of tools to be included in the profile and the definition of constraints on these tools, the definition of the semantic constraints plays a crucial role for this profile. Firstly, as a consequence of the requirement for almost unconstrained support of the structure description tools, all the generality and flexibility of these tools as well as the resulting ambiguities are taken over into the profile. Secondly, in the envisaged application area, the exchange of content descriptions between different systems is an important issue. Thus this profile is defined with interoperability in mind, which can be only achieved in part by tool constraints.

The semantic constraints can be compared to the explanations on description tool semantics in the text of the

standard. As the application area of MPEG-7 is wide, these descriptions on tool semantics cannot be restrictive, because this could exclude some application scenarios. In a profile, a reduced set of supported functionalities is defined, and the semantic constraints can be as restrictive as this set of functionalities allows. Thus the semantic constraints in a profile definition can be used to resolve ambiguities caused by the openness of the tool semantics in the standard, and support interoperability.

The need of a semantic constraint on a certain description tool in a profile depends on two factors. The first is the generality of the definition of the tool, i.e. the strictness of the tool semantics in the standard. For example, a descriptor for a certain visual feature has a strictly defined semantic meaning, while the semantics of a temporal segment are not defined a priori, but depend on the context in which it is used. The second is the semantic information that can be conveyed by the tool. For instance, the semantics of both text annotations and media locators depend on the segment of the description where they are used. While the meaning of the media locator is clear when the context is known, the semantics of the text annotation is not so strictly defined. Consequently, most of the semantic constraints in this profile definition constrain the top-level types and the structure description tools.

5. Definition of the Detailed Audiovisual Profile (DAVP)

Based on the considerations discussed in the previous section we present the definition of a profile for the comprehensive and fine grained description of single audiovisual content entities, called Detailed Audiovisual Profile (DAVP). The XML schema of the profile and a comprehensive description of the semantic constraints are available on the web [6].

5.1. Structure of descriptions in DAVP

Many of the constraints of the DAVP definition aim at structuring the description and defining the semantics of the parts of the description. A basic principle of DAVP is to describe a single audiovisual content entity per MPEG-7 description. The description includes the detailed description of the content, consisting of a structural description and associated textual or feature descriptions on arbitrary levels, and a summary of the content description. The profile allows the use of a wide range of spatiotemporal structuring tools. A further design principle is to keep content decompositions and the related annotations as modular as possible, i.e. to separate decompositions based on different modalities, on different levels of abstraction, or created with/without the use of domain knowledge. An overview of the structure of descriptions conforming to DAVP is shown in Figure 1.

5.2. Parts of MPEG-7 included in DAVP

In contrast to the three adopted MPEG-7 profiles, part 3 and 4 (visual and audio feature descriptions) have been completely included in DAVP. We are convinced that a comprehensive description of audiovisual content must allow the use of low- and mid-level feature descriptions. There are two important uses of this kind of descriptions within the intended application area of DAVP, namely query by example and the extraction of semantic information based on low-level features. In both cases the metadata description must include the low-level audiovisual feature descriptions.

Many of the tools defined in part 5 (Multimedia Description Schemes, MDS [2]) are included in DAVP as discussed in detail in the following section.

5.3. Selection and constraints of MDS tools

5.3.1. Top-level types. The scope of DAVP is restricted to the description of audiovisual media items. The main type for the description of audiovisual content is thus ContentEntityType. Except for ContentEntityType only three types of descriptions are included. ClassificationDescription may be used for stand-alone classifications schemes (cf. 5.3.2.), SemanticDescription to describe semantic entities that are shared between several content descriptions, and SummaryDescription for a summary of the complete content description.

For ContentEntityType descriptions, two types of multimedia content are supported. ImageType is used for the description of still image content, AudioVisualType for the description of audio, video and audiovisual content. We have decided not to use AudioType and VideoType in order to unify the description of content in which either the visual modality or the audio modality or both are present.

5.3.2. Basic description tools. Most of the types of textual annotation are included. Affective content description tools as well as the tools for agents, places, graphs and relations are fully included, especially for the use in semantic descriptions.

Classification schemes are supported with the exception of GraphicalClassificationSchemes. We recognize the need for controlled vocabularies, which can be expressed using simple classification schemes. For any classification schemes, that include further relations between the terms, we advocate the use of an ontology representation like OWL [13].

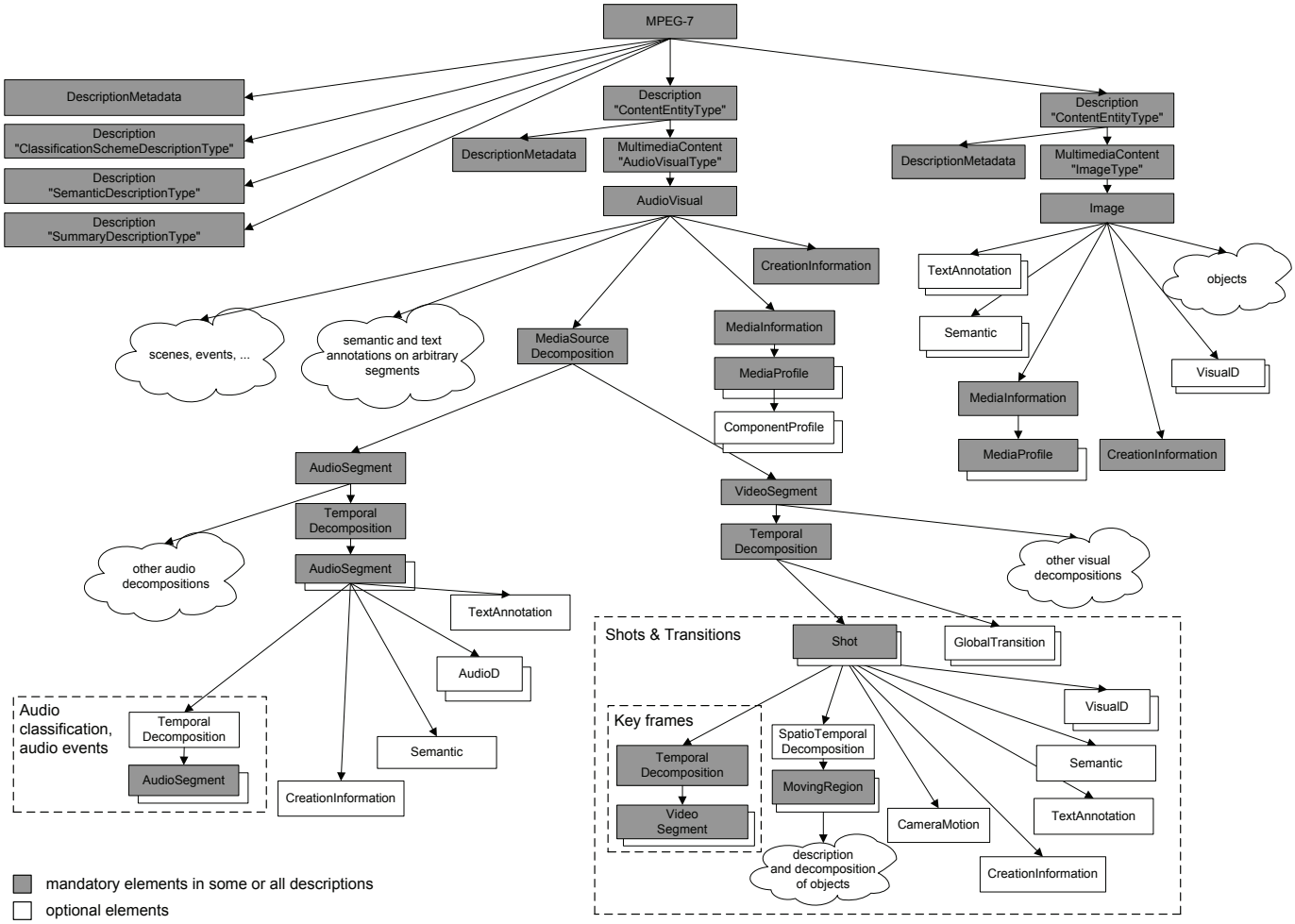

Figure 1: Overview of the structure of descriptions conforming to the Detailed Audiovisual Profile.

5.3.3. Media, creation and usage description tools. These tools, which are mainly used for global metadata, are fully included in the profile, only some tool constraints apply.

5.3.4. Structure description tools. The structuring tools are one of the most powerful groups of tools in MDS and thus all tools for structural description have been included. This also includes tools for describing moving regions, visual text in images and video, and most of the description tools for edited video segments.

Tools for the description of multimedia, ink and handwriting segments have been excluded, as the scope of DAVP is limited to the description of audiovisual and image content.

As structure description tools may be used in different ways, some tool constraints have been introduced to make the decompositions and segments identifiable. These constraints include mandatory use of IDs and StructuralUnit descriptors on all segments, and the mandatory use of the criteria attribute on all decompositions. For commonly used decompositions and segments, a classification scheme for StructuralUnit elements and of a set of recommended values for criteria attributes have been defined.

5.3.5. Navigation and access tools. Summarization tools have been included to allow the description of a summary aside the detailed content description, which can be used for efficient browsing and navigation of the content description. Multiple summaries of one audiovisual content entity may be described.

5.4. Semantic constraints

While the use of structuring tools is nearly unconstrained on the tool level in DAVP, a number of semantic constraints deal with the structure of the description. The main design criteria of these semantic constraints is to keep the description as modular as possible. This is mirrored by the fact that description fragments which are based on different modalities are not put into structurally dependent parts of the description, but are kept independently side by side. The same applies to descriptions expressing different levels of semantic

content, such as descriptors directly derived from features of the audiovisual content (e.g. shot boundaries) and descriptions on a higher semantic level (e.g. scene structuring), which may have been generated by combining other descriptions on a lower semantic level, by using prior knowledge or other external sources.

These constraints do not aim to limit the flexibility of structural descriptions, but to organize them in a way, that their mutual dependencies are reduced as much as possible. This shall enable applications or application components to access or modify certain structural decompositions, which are semantically independent, without having to deal with other, unrelated parts of the description. Especially for annotations on a higher semantic level, such as textual or semantic annotations, the constraints still allow a very high degree of flexibility. The semantic constraints just ensure that parts of the descriptions do not have more interdependencies than necessary and that the semantics of the parts are identifiable.

There is a set of common structural descriptions of audiovisual content, which is used very frequently in the envisaged application area, e.g. shot lists of video content and description of the associated key frames. For these types of structural descriptions, DAVP defines a description structure that ensures the exchangeability of these common structural descriptions, while leaving other, additional structural annotations unaffected.

Many of the semantic constraints can also be seen in the visualization of the DAVP description structure shown in Figure 1.

DAVP describes single audiovisual content entities. If the description is an audiovisual content description (ContentEntityType), only one multimedia content may be described. The content entity type element has either a AudioVisualType or a ImageType element, containing a single AudioVisual or Image element respectively.

A summary of an audiovisual content is modeled as a separate description in MPEG-7. In DAVP, a detailed description and a summary description of the same content is the only case, where two descriptions of the same content may be used. In all other cases, such as standalone summaries, only one description is permitted.

For the description of audiovisual content, the visual and audio parts are described as elements of a media source decomposition of the root audiovisual segment. Annotations concerning both visual and audio information, such as a scene structure generated by combining visual and audio hints, are modeled as temporal or spatiotemporal decompositions of the AudioVisual root segment with an arbitrary recursive structure of AudioVisual and TemporalDecomposition elements below.

Further semantic constraints restrict the use of media description tools on the root segment, where a full media description may be used, and on other segments, where just limited media description tools are allowed (e.g. reference to key frame images).

6. Application of the profile

The profile presented here has been used for content description in a number of projects in different application areas. The experiences gained within two years have provided valuable feedback and have helped to improve the profile definition. The use in these applications has proven that the concept of the profile is feasible, and the diversity of the applications shows, that a restriction to certain functionalities does not necessarily imply a restriction of the possible application areas. In the following, we discuss the application areas in which the profile has been employed.

6.1. Search and retrieval

The profile has been used for content description in a system called Multimedia Mining Toolbox [10], which includes components for automatic content analysis, manual annotation and content based search on multimedia data (video, still images). During ingestion of content, automatic analysis is performed (shot boundary detection, camera motion estimation, keyframe extraction, moving object segmentation, extraction of low-level visual features). The search and retrieval component provides fast access to the digital archive by supporting the formulation of combined text and content-based (similarity based) queries for visual content. The tool enables efficient search for all features, which are automatically extracted by the content analysis component. A media summary viewer visualizes an entire video in terms of a temporal summary/overview and by providing efficient navigation functionality using a temporally condensed video representation.

This application scenario has shown that low-level audiovisual features are required in a profile for detailed content description. Also the semantic constraints on the structuring tools play a crucial role, as the decompositions created during content analysis correspond to the units of the content, that can be retrieved as query results (e.g. shots, moving objects).

Within this project, several tools for working with MPEG-7 descriptions have been developed, which are not application specific and have also been used in the other projects discussed in the following. These tools include a freely available MPEG-7 library in C++ [5], which provides an object-oriented API for the description tools defined in part 3, 4 and 5 as well as a client/server infrastructure for sharing MPEG-7 descriptions in a distributed system.

The profile is also used in the project PrestoSpace [14], which aims at developing solutions for the preservation of

audiovisual archives. In this project, MPEG-7 will also be used for the description of audiovisual defect and quality measures.

6.2. Media monitoring

In many media monitoring applications, content description is the link between the audiovisual content analysis components, which process usually a large amount of multimedia data (e.g. a broadcast stream) and components performing higher level analysis.

In the DIRECT-INFO project [15], which aims at creating a basic system for semi-automatic sponsorship tracking for decision support, 24/7 monitoring of audiovisual streams is done and relevant segments are detected. On these segments, a number of automatic content analysis tasks, such as genre classification, logo detection and event modeling are performed. The generated MPEG-7 metadata are collected in a central repository, which is the basis for fusion and report generation.

In this application scenario, the content description needs to support annotations on different semantic levels and several annotations, with some of them defining their own decomposition of the content. Examples for such diverse annotations with different and independent decompositions are shot structure, genre classification and automatic speech recognition (ASR) results.

The use of the profile in media monitoring applications has shown that the modularity of the description and the possibility to add further annotations and decompositions without modifying the existing description structure are necessary.

6.3. Content production

The application of content description in a content production environment has partial overlap with search and retrieval applications. Automatic content analysis tools and manual annotation can be used to index source content (e.g. rushes) and content-based retrieval tools can help to find content needed in the scope of the production.

Currently most of the metadata available during the production process is lost. To avoid this, it is desirable to create content descriptions in addition to the audiovisual essence as a result of the production process. This content description can be much richer than a description that has to be reconstructed only by subsequent analysis of the audiovisual data.

Examples of projects in this application area, which are using MPEG-7 DAVP for content description, are IP-RACINE [4] and NM2 [11]. The experiences from this application scenario are similar to that from the search and retrieval scenario. Moreover, as it is often important to gain a quick overview over the structure and content of available material, summarization capabilities are very important.

7. Conclusion

In this work, we have studied the use of MPEG-7 for the detailed description of single audiovisual content entities and we have proposed the Detailed Audiovisual Profile (DAVP) to overcome the main drawbacks faced.

The main drawbacks when using MPEG-7 for audiovisual content description are the complexity of MPEG-7 descriptions and interoperability problems due to multiple possibilities for representing semantically identical descriptions. MPEG-7 profiling is a possible solution for these problems by constraining the functionality according to the requirements of the profile's application area. Defining a subset of functionalities will only partly solve these problems, as this does not always reduce the corresponding set of tools. If the functionalities of the profile require the use of more generic description tools, such as structuring tools, there is less complexity reduction. This problem is partly caused by the fact, that a description tool cannot be used independent of its semantics. As pointed out in [12], the approach chosen in MPEG-7 fuses the definition of data types and that of their semantics. Thus including a description tool in the profile also means including its semantics.

If a set of generic description tools has to be included in the profile, it is necessary to define their semantics for each context in which they are used in the profile. In [1], semantic constraints are briefly mentioned as one step of the profile definition process, which can be used to define the use of an included tool in the context of the profile. However, they are not discussed in that document like the other steps, tool selection and tool constraints. Also none of the currently adopted profiles makes use of semantic constraints, they only define subsets of tools. From the currently defined profiles it can be concluded, that MPEG-7 profiling has been mainly used to reduce the number of constructs that must be supported by applications creating and processing descriptions conforming to these profiles.

Profiles can be used as a means to ensure interoperability between systems. Using semantic constraints in the profile definition ensures exchangeability of descriptions on a semantic level and this makes profiles an important and powerful concept for the use of MPEG-7.

In this work we have presented a profile for the detailed description of single audiovisual content entities. In contrast to other profiles, we have included the tools for audio and visual feature description. The profile also includes many of the MDS tools, such as a wide range of structuring tools, as well as tools for the description of media, creation and production information and textual and semantic annotation. Furthermore, summarization tools are included in the profile.

As the profile includes many of the MDS tools to cover the required functionality in its application area, tool selection and tool constraints will not drastically reduce the complexity of the descriptions. Moreover, tool constraints cannot resolve possible ambiguities in the use of the description tools, as they are just syntactic, and not semantic constraints.

The novelty of this work is that DAVP is the first MPEG-7 profile, where semantic constraints play a crucial role in the profile definition. They define the use of the MPEG-7 description tools in the context of the profile and allow describing the relations between the description tools included in the profile and constraining their use depending on the context. This means that only semantic constraints can facilitate system interoperability by ensuring exchangeable MPEG-7 descriptions.

The Detailed Audiovisual Profile (DAVP) has been successfully used for audiovisual content description in a wide range of applications. The semantic constraints, which especially ensure the modularity of descriptions, have proven to be the key enabler for interoperability.

8. Future work

To ensure conformance to profiles on a semantic level, and thus use profiles as a means for interoperability between systems using MPEG-7, the definition of future profiles must include a definition of their semantic constraints. But not all tool constraints and hardly any semantic constraints can be formalized in the XML Schema definition of the profile and must thus be described in textual form like the description of tool semantics in the standard. If the validation of conformance in terms of semantic constraints shall be done automatically—and this is probably the only practical way, an appropriate formalization of these semantic constraints has to be found.

9. Acknowledgements

The authors would like to thank Georg Thallinger, Herwig Rehatschek and Werner Haas as well as several other colleagues at JOANNEUM RESEARCH, who contributed to the definition of the profile and provided valuable feedback. This work has been funded partially under the 6th Framework Programme of the European Union within the IST projects "PrestoSpace" (IST FP6-507336, http://www.prestospace.org), "DIRECT-INFO" (IST FP6-506898, http://www.direct-info.net) and "IP-RACINE" (IST-2-511316-IP, http://www.ipracine.org).

10. References

[1] Definition of MPEG-7 Description Profiling. ISO/IEC JTC 1/SC 29/ WG 11 N6079, Oct. 2003.

[2] Information Technology—Multimedia Content Description Interface, Part 5: Multimedia Description Schemes. ISO/IEC 15938-5, 2001.

[3] Information Technology—Multimedia Content Description Interface, Part 7: Conformance. ISO/IEC 15938-7, 2001.

[4] IP-RACINE—Integrated Project Research Area Cinema. URL: http://www.ipracine.org

[5] Free MPEG-7 Library. URL: http://mpeg-7.joanneum.at

[6] MPEG-7 Detailed Audiovisual Profile (DAVP). URL: http://mpeg-7.joanneum.at

[7] MPEG-7 Interoperability, Conformance Testing and Profiling. ISO/IEC JTC1/SC29/WG11 N4039, Mar. 2001.

[8] MPEG-7 Overview (version 7). ISO/IEC JTC1/SC29/WG11 N4674, Mar. 2002.

[9] MPEG-7 Profiles and Levels under Consideration. ISO/IEC JTC 1/SC 29/ WG 11 N6039, Oct. 2003.

[10] Multimedia Mining Toolbox. URL: http://www.joanneum.at/cms_img/img2287.pdf

[11] NM2—new media for a new millennium. URL: http://www.ist-nm2.org

[12] F. Nack, J. van Ossenbruggen, and L. Hardman, “That obscure object of desire: multimedia metadata on the Web”, part 2, IEEE MultiMedia, vol. 12, nr. 1, Jan. 2005, pp. 54-63.

[13] OWL Web Ontology Language. W3C Recommendation, URL: http://www.w3.org/TR/owl-features

[14] PrestoSpace—Preservation towards storage and access. Standardized Practices for Audiovisual Contents in Europe. URL: http://www.prestospace.org

[15] H. Rehatschek, N. Diakopoulos, G. Kienast, V. Hahn and T. Declerk, “DIRECT-INFO: A Distributed Multimodal Analysis System for Media Monitoring Applications”, Proc. EWIMT—European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, London, UK, Dec. 2004.

[16] Study of MPEG-7 Profiles Part 9 Committee Draft. ISO/IEC JTC1/SC29/WG11 N6263, Dec. 2003.

[17] XML Schema Part 1: Structures, Second Edition. W3C Recommendation, Oct. 2004.