A VISUAL DESCRIPTOR ONTOLOGY FOR MULTIMEDIA REASONING

N. Simou, V. Tzouvaras, Y. Avrithis, G. Stamou and S. Kollias

Department of Electrical and Computer Engineering,
National Technical University of Athens, Zographou 15780, Greece
email:{nsimou, tzouvaras, iavr}@image.ntua.gr
{gstam, stefanos}@softlab.ntua.gr

ABSTRACT

In this paper we present the construction of an ontology that represents the structure of the MPEG-7 visual part. The goal of this ontology is to enable machines to generate and understand visual descriptions which can be used for multimedia reasoning. Within the specification, MPEG-7 definitions (description schemes and descriptors) are expressed in XML Schema. Although XML Schema provides the syntactic, structural, cardinality and datatyping constraints required by MPEG-7, it does not provide the semantic interoperability required to make MPEG-7 visual descriptors accessible by other domains. The knowledge representation provided by the ontology can be used to develop tools which perform knowledge-based reasoning. For the construction of the ontology we use the RDFS ontology language. We present the problems that occurred, mainly, due to the RDFS modelling limitations. Finally, we propose a way to apply reasoning using the VD ontology.

1. INTRODUCTION

To enable multimedia content to be discovered and exploited by services, agents and applications, it needs to be described semantically. Generating descriptions of multimedia content is inherently problematic because of the volume and complexity of the data, its multidimensional nature and the potentially high subjectivity of human-generated descriptions. Significant progress has been made in recent years on automatic segmentation or structuring of multimedia content and the recognition of low-level features within such content. However, comparatively little progress has been made on machine-generation of semantic descriptions of audiovisual information.

During the early development stages of MPEG-7, Unified Modelling Language (UML) was used to model the entities properties and relationships (description schemes and descriptors) which comprised MPEG-7. However, the massive size of the specification combined with the belief that the UML models were a development tool which duplicated information in the XML schemas, led to the decision to drop them from the final specification. Although the lack of an existing data models hinders the development of an MPEG-7 ontology, it also means that the generated ontology will be even more valuable, providing both data model and a definition of the semantics of MPEG-7 terms and the semantic relationships between them. Building the ontology should also highlight any inconsistencies, duplication or ambiguities, which exist across the large number of MPEG-7 description schemes and descriptors. Without a data model to build on, the class and property hierarchies and semantic definitions had to be derived through reverse-engineering of the existing XML Schema definitions together with interpretation of the English-text semantic descriptions.

A MPEG-7 ontology has been constructed and is presented in [5]. However, since MPEG-7 is too large, is very difficult to represent all its features with the required detail in one ontology. To simplify the process, they used a core subset of the MPEG-7 specification together with a top-down approach to generate the ontology. The constructed MPEG-7 ontology is very general and unappropriate to dedicated multimedia applications. In this paper we present the representation of the visual part of MPEG-7 in an ontology. This ontology has been constructed with a great detail according to the MPEG-7 visual part. The problems that occurred for the construction of the ontology are presented together with possible ways to overcomes them. Finally, we present a way to apply reasoning using the ontology.

2. VDO IMPLEMENTATION

2.1. Visual Descriptor Ontology

This section provides a description of the Visual Descriptor Ontology (VDO), containing a set of visual descriptors to be used for knowledge-assisted analysis of multimedia content. The term descriptor has been taken from the MPEG-7 that standardizes the core technologies allowing description of audiovisual data content in multimedia environments. This effort was achieved by standardizing descriptors, descriptors Schemes and description definition language. Descriptors are the representations of features that define the syntax and the semantics of each feature representation. Description Schemes on the other hand specify the structure and the semantics of the relationships between their components that can be Descriptors or Description Schemes and description definition language allows the definition of Descriptors and Descriptor schemes. MPEG-7 makes this standardization into seven parts; our ontology is based on the third part of MPEG [8] that specifies the visual content. Hence by the term descriptor we mean a specific representation of a visual feature (color, shape, texture etc) that defines the syntax and the semantics of a specific aspect of the feature (dominant color, region shape etc).

2.2. Implementing the VD ontology

VD ontology was implemented using Ontoedit [9] that is an Ontology Engineering Environment supporting the development and maintenance of ontologies by using graphical means. Using this tool the creation of an ontology of an internal format (oxml) is possible, which can be exported to two of the ontology representation languages that are RDF(S) and DAML+OIL.

2.3. VD ontology Structure

This section describes the structure of the Visual Description ontology and the design decisions that were made. The entire Visual Description follows the specification of the MPEG-7 Visual Part, but several modifications were carried out in order to adapt to the XML Schema provided by MPEG-7 to an ontology.

Ontology as defined by Tom Gruber [3] is an explicit specification of a conceptualization and it consists by several components of which the most important are concepts, relations and attributes, instances and axioms. Concepts which are abstract terms organized in taxonomies building a tree, are the vital part of an ontology.

The tree of the Visual Descriptor ontology consists of four main concepts, which are 'Region', 'Feature', 'VisualDescriptor' and 'Metaconcepts'. None of these concepts is included in the XML Schema defined MPEG-7 but their need was vital in order to create a correctly defined ontology. 'VisualDescriptor' concept contains the visual descriptors as these are defined by MPEG-7. 'Metaconcepts' concept on the other hand contains some additional concepts that were necessary for the Visual descriptor ontology but they are not clearly defined in the XML Schema of MPEG-7.The remaining two concepts that were defined 'Region' and 'Feature' are also not included in the MPEG-7 specification but their definition was necessary for the completeness of the ontology.

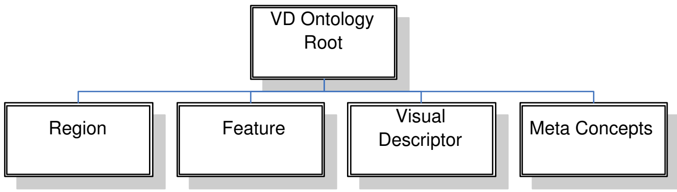

Fig. 1. The Visual Description Ontology.

Visual Description ontology was made so as to make possible the transparently integration of MPEG-7 visual part. Hence a relation of the visual content and the visual descriptors was necessary. So for that purpose concepts 'Region' and 'Feature' were created. These concepts provide an abstract specification of the visual content that is described, relating it with the Visual Descriptors. Therefore concept 'Region' is a concept that represents a region of a visual content and sets relation 'hasVisualDescriptor' of range 'VisualDescriptor'. In other words this relation ensures that a region has a visual descriptor that specifies its description and also defines the meaning of 'VisualDescriptor'. Similarly concept 'Feature' consists of four concepts that are 'Color', 'Shape', 'Motion' and 'Texture' that are relationally connected with the relevant subclasses of 'VisualDescriptor' concept 'ColorDescriptor', 'ShapeDescriptor', 'MotionDescriptor' and 'TextureDescriptor'.

'VisualDescriptor' concept is the top concept of our ontology since it contains all the visual descriptors. Its structure is based on the way that Visual Descriptors are specified by MPEG-7. Hence it consists of six subclasses one for each category that MPEG-7 specifies. These are color, shape, texture, motion, localization and basic descriptors that are needed for the definition of the others. Each of these categories includes the relevant descriptors that are defined as concepts in the VD ontology. The only category that was modified and does not contain all the MPEG-7 descriptors is the 'BasicDescriptors'. This category of descriptors as specified by the MPEG-7 standard consist of five descriptors of which only 'Spatial2DCoordinates' and 'TemporalInterpollation' was used. This restriction was made because the others descriptors (GridLayout, TimeSeries, MultipleView) were of no use for the Visual Descriptor Ontology since it was not possible to be determined.

Except 'BasicDescriptors' the other descriptors were represented as defined by MPEG-7. MPEG-7 defines visual descriptors by using some components that specify them. In order to make this representation into an ontology we defined these components of each a descriptor using relations and defining the range of the relations. In order to understand this procedure the implementation of 'DominantColorDescriptor' will be explained. Firstly dominant color is a visual descriptor and more specifically it is one of MPEG-7 color descriptors. Therefore we defined concept 'DominantColorDescriptor' as a subclass of concept 'ColorDescriptor' and analogously 'ColorDescriptor' as a subclass of 'VisualDescriptor'.

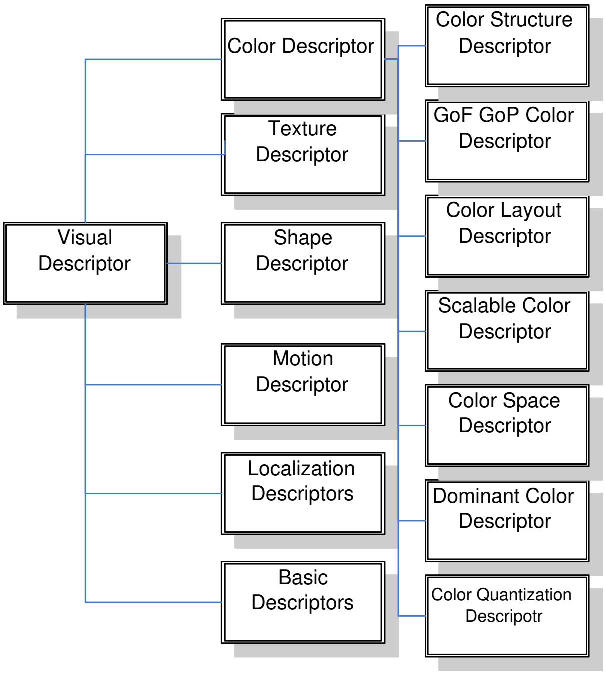

Fig. 2. Dominant Color Hierarchy.

The next step after ensuring the MPEG-7 structure in the Visual Descriptor ontology hierarchy was the components representation. Dominant color is represented by XML Schema of MPEG-7 as complex type and is described by its components. Defining 'DominantColorDescriptor' as a concept we created the complex type and hence the necessary conversion for the Visual Descriptor ontology. The components that specify Dominant Color are 'values', 'spatialCoherency', 'colorSpace' and 'ColorQuantization' that are defined as relations.Among these relations only 'spatialCoherency' was possible to be specified using the datatypes of XML Schema and setting it's range to double. The remaining components that specify Dominant color were of complex type breeding the need of a new concept that would represent them. For this reason concept 'Metaconcepts' was created in order to include the concepts that despite the fact that were not visible at a first sight when examining XML Schema of MPEG-7, there were necessary for the implementation of the visual Descriptor Ontology.

2.4. Problems and restrictions

This section describes the problems that we encounter in the implementation of the Visual Descriptor ontology. Ontoedit that was the tool that was used for the development of the VD ontology allowed us to export to the ontology to RDF(S) or DAML+OIL. One of the problems in either of these languages was datatypes.Examining the Dominant color example, can be observed that we need to specify the datatypes of the components. Dominant Color consist of 'values', 'spatialCoherency', 'colorSpace' and 'ColorQuantization'.The 'spatialCoherency' component is of datatype double that is taken from XMLSchema either exporting the ontology to to RDF(S) or DAML+OIL.

Except this component we had to represent and the remaining of 'DominantColorDescriptor' and since these were of complex type we created additional concepts.But unfortunately this was a partial solution of the problem since RDF(S) is restricted in declaring properties to vary within a certain set of range values, allowing this range to be only a datatype of XMLSchema.

Another important problem that we met was the enumerations of MPEG-7 Schema. The way that we tried to overcome them was by defining the appropriate cardinalities to the relations of the Visual Description ontology. Again this was partially worked since cardinalities restrictions are only available in DAML+OIL.

On the other hand Resource Description Framework RDF(S) despite these restrictions is a recent W3C recommendation designed to standardize the definition and use of meta-data descriptors of web-based resources and it is supported by almost all inference engines.

3. REASONING USING THE VD ONTOLOGY

In this section we describe a method to apply reasoning using the VD ontology. We consider a system that automatically analyses images and extracts visual descriptor from specific region of images. Automatic visual descriptors extraction for images is a significant research area and a wide variety of tools and mechanisms exist for analysing scientific and other types of images. Within the scope of reasoning of this paper, we are not particularly interested in the underlying method by which these programs analyse images but we are interested in the outputs or features that such programs produce and whether they can be used by inferencing engines to infer the occurrence of higher-level semantic concepts e.g., regions or objects of a particular colour, texture and shape.

The data produced by the visual descriptor extraction systems needs to be mapped to the VD ontology and related (using the inferencing rules) to higher-level semantic terms defined in the knowledge. In particular we want to map visual descriptors extracted from specific regions of the images to higher-level semantic concepts, such as object and events.

The main tasks of the reasoning system will be to be able for atom-region merging and to evaluate the plausibility of the initial hypotheses from the set of atom-regions together with their initial labels as resulted. For this purpose, the declaration of rules that will enclose the relevant information concerning the atoms-regions is necessary. Regarding the first task rules will be defined in a manner that will exploit topological information. In other words, when a set of atom-regions initially labeled as belonging to the same object class and their topological information verifies that they are related to one another through adjacency or inclusion then the merging process of these regions to form a single object instance will be made. Similarly for the second task spatiotemporal information will be exploited to reduce the hypotheses corresponding to an atom-region depending on its neighboring atom-regions labels and spatial interrelations. Consequently, the second task treats the expectation about the characteristic locations of the sought objects.

Despite the spatiotemporal information, the integration of low-level features in the reasoning process will further improve the plausibility of the detection results. Low level features will provide information about the shape and the color etc that will be able to specify an object class. Therefore a merging indicated by the defined rules will be performed only if the shape and the color of the resulting segment conform to the shape and the color of one of the plausible object classes corresponding to the merged atom-regions. Obviously, this raises the need for incorporation of low-level feature matching into the reasoning system, which can lead to computational problems and also to reduce the number of eligible reasoning systems, because means to extend the system must be available.

The fuzzy nature of image indexing and retrieval means that users do not require exact matches between the extracted colour , texture or shape descriptors and the prototype model stored in the domain-specific ontology, but only require similarity-based search and retrieval.

This simple example shows the kind of reasoning that we can apply using the VD ontology and a domain specific ontology to detect semantic object or define further semantic relationships, information not explicitly recorded by the system. Such inferencing may be done on the fly or in advance and then stored in the ontology to enable more sophisticated retrieval and query services.

4. CONCLUSION

In this paper we presented the construction of an ontology that represents the structure of the MPEG-7 visual part. The goal of this ontology is to enable machines to generate and understand visual descriptions which can be used to enable multimedia reasoning. The knowledge representation provided by the ontology can be used to develop tools which perform knowledge-based reasoning. For the construction of the ontology we use the RDF-S ontology language. We present the problems that occurred caused by the RDF-S modelling limitations. Finally, we proposed a way to apply reasoning using the VD ontology.

Acknowledgements This research was partially supported by the European Commission under contract FP6-001765 aceMedia. The expressed content is the view of the authors but not necessarily the view of the aceMedia project as a whole.

5. REFERENCES

[1] http://www.daml.org.

[2] I.Horrocks S. Decker M. Erdmann M. Klein D.Fensel, F.Harmelen. Oil in a nutshell. *Proc. Of the 12th European Workshop on Knowledge Acquisition, Modeling, and Management*, 2000.

[3] T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition, 5(2) pages:199-220, 1993* .

[4] J. Hunter. Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description, Volume: 13 , Issue: 1 , Jan. 2003 Pages:49 - 58*.

[5] J. Hunter. Adding multimedia to the semantic web - building an mpeg-7 ontology. *Proceedings of the First Semantic Web Working Symposium*, pages 261–281, 2001.

[6] C. Colombo A. Del Bimbo J. Assfalg, M. Bertini. Semantic annotation of sports videos. *IEEE MultiMedia, April-June 2002 (Vol. 9, No. 2)*.

[7] R.Fikes M.Genesereth. *Knowledge Interchange Format, Version 3.0 Reference Manual. Technical Report Logic-92-1, Computer Science Department, Stanford University, Stanford, US, June 1992.*.

[8] Thomas Sikora. The mpeg-7 visual standard for content description-an overview. *IEEE Transaction on circuits and Systems for video technology*, 11, 2001.

[9] J.Angele D.Wenke A.Maedche Y.Sure, S.Staab. Ontroedit: Guiding ontology development by methodology and inferencing. 2002.