Combined Domain Specific and Multimedia Ontologies for Image Understanding

Kosmas Petridis[1], Frederic Precioso[1], Thanos Athanasiadis[2], Yannis Avrithis[2] and Yiannis Kompatsiaris[1]

[1] Informatics and Telematics Institute, GR-57001 Thermi-Thessaloniki, Greece
[2] National Technical University of Athens, School of Electrical and Computer Engineering, GR-15773 Zographou, Athens, Greece

Abstract. Knowledge representation and annotation of multimedia documents typically have been pursued in two different directions. Previous approaches have focused either on low level descriptors, such as *dominant color*, or on the content dimension and corresponding manual annotations, such as *person* or *vehicle*. In this paper, we present a knowledge infrastructure to bridge the gap between the two directions. Ontologies are being extended and enriched to include low-level audiovisual features and descriptors. Additionally, a tool for linking low-level MPEG-7 visual descriptions to ontologies and annotations has been developed. In this way, we construct ontologies that include prototypical instances of domain concepts together with a formal specification of the corresponding visual descriptors. Thus, we combine high-level domain concepts and low-level multimedia descriptions, enabling for new media content analysis.

1 Introduction

Representation and semantic annotation of multimedia content have been identified as an important step towards more efficient manipulation and retrieval of visual media. Today, new multimedia standards such as MPEG-4 and MPEG-7, provide important functionalities for manipulation and transmission of objects and associated metadata. The extraction of semantic descriptions and annotation of the content with the corresponding metadata though, is out of the scope of these standards and is still left to the content manager. This motivates heavy research efforts in the direction of automatic annotation of multimedia content.

Here, we recognize a broad chasm between existing multimedia analysis methods and tools on one hand and semantic description, annotation methods and tools on the other. The state-of-the-art multimedia analysis systems are severely limiting themselves by resorting mostly to visual descriptions at a very low level, e.g. the dominant color of a picture. However, ontologies that express key entities and relationships of multimedia content in a formal machine-processable representation can help to bridge the *semantic gap* [1, 2] between the automatically extracted low-level arithmetic features and the high-level human understandable semantic concept.

Work on *semantic annotation* [3] currently addresses mainly textual resources [4] or simple annotation of photographs [5]. In the *multimedia analysis* area, knowledge about multimedia content domains is a promising approach by which Semantic Web technologies can be incorporated into techniques that capture objects through automatic parsing of multimedia content. In [6], ontology-based semantic descriptions of images are generated based on appropriately defined rules that associate MPEG-7 low-level features to the concepts included in the ontologies. The architecture presented in [7] consists of an audio-visual ontology in compliance with the MPEG-7 specifications and corresponding domain ontologies.

Acknowledging the relevance between low-level visual descriptions and formal, uniform machine-processable representations, we try to bridge the chasm by providing a knowledge infrastructure design focusing both on multimedia related ontologies and domain specific structures. The remainder of the paper is organized as follows: in section 2 we present the general ontology infrastructure design, including a brief description of a tool to assist the annotation process needed for initializing the knowledge base with descriptor instances of domain concepts. A small overview and results from the knowledge-assisted analysis process, which are exploiting the developed infrastructure and annotation framework are presented in section 3. We conclude with a summary of our work in section 4.

2 Knowledge Representation

Based on the above, we propose a comprehensive Ontology Infrastructure, the components of which will be described in this section. The challenge is that the hybrid nature of multimedia data must be necessarily reflected in the ontology architecture that represents and links multimedia and content layers. Fig. 1 summarizes the developed knowledge infrastructure.

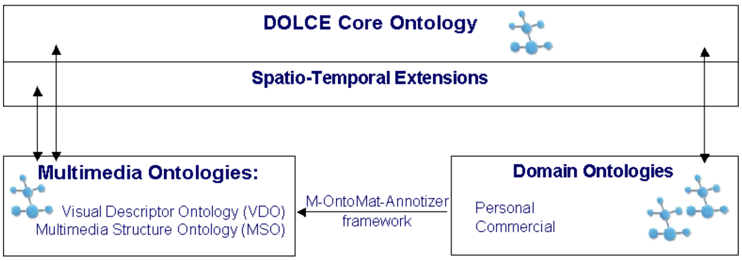
DOLCE Core Ontology
Spatio-Temporal Extensions
Multimedia Ontologies:
Visual Descriptor Ontology (VDO)
Multimedia Structure Ontology (MSO)
M-OntoMat-Annotizer framework
Domain Ontologies
Personal
Commercial

Fig. 1. Ontology Structure Overview

Overview Our framework uses *RDFS (Resource Description Framework Schema)* as modeling language. This decision reflects the fact that a full usage of the increased expressiveness of *OWL (Web Ontology Language)* requires specialized and more advanced inference engines that are not yet available, especially when dealing with large numbers of instances with slot fillers.

Core Ontology The role of the core ontology in this overall framework is to serve as a starting point for the construction of new ontologies, to provide a reference point for comparisons among different ontological approaches and to serve as a bridge between existing ontologies. In our framework, we have used *DOLCE* [8] for this purpose.

Prototype Approach Describing the characteristics of concepts for exploitation in multimedia analysis naturally leads to a meta-concept modeling dilemma. This issue occurs in the sense that using concepts as property values is not directly possible while avoiding 2^{nd} order modeling, i.e. staying within the scope of OWL DL. In our framework, we propose to enrich the knowledge base with instances of domain concepts that serve as *prototypes* for these concepts. This status is modeled by having these instances also instantiate an additional `VDO-EXT:Prototype` concept from a separate *Visual Annotation Ontology (VDO-EXT)*. Each of these instances is then linked to the appropriate visual descriptor instances. The approach we have adopted is thus pragmatical, easily extensible and conceptually clean.

Multimedia Ontologies *Multimedia Ontologies* model the domain of multimedia data, especially the visualizations in still images and videos in terms of low-level features and media structure descriptions. Structure and semantics are carefully modeled to be largely consistent with existing multimedia description standards like MPEG-7.

Visual Descriptor Ontology The Visual Descriptor Ontology (VDO) contains the representations of the MPEG-7 visual descriptors, models *Concepts* and *Properties* that describe visual characteristics of objects. Although the construction of the VDO is tightly coupled with the specification of the MPEG-7 Visual Part [9], several modifications were carried out in order to adapt to the XML Schema provided by MPEG-7 to an ontology and the data type representations available in RDF Schema. The `VDO:VisualDescriptor` concept is the top concept of the VDO and subsumes all modeled visual descriptors. It consists primarily of six subconcepts, one for each category that the MPEG-7 standard specifies. These are: *color, shape, texture, motion, localization* and *basic descriptors*. Each of these categories includes a number of relevant descriptors that are correspondingly defined as concepts in the VDO.

Multimedia Structure Ontology The Multimedia Structure Ontology (MSO) models basic multimedia entities from the MPEG-7 Multimedia Description Scheme [10] and mutual relations like decomposition. Within MPEG-7, multimedia content is classified into five types: *image, video, audio, audiovisual* and *multimedia*.

Domain Ontologies In the multimedia annotation framework, the domain ontologies are meant to model the content layer of multimedia content with respect to specific real-world domains, such as sports events like tennis. All domain ontologies are explicitly based on or aligned to the DOLCE core ontology, and thus connected by high-level concepts, what in turn assures interoperability between different domain ontologies at a later stage.

In the context of our work, domain ontologies are created and maintained by content managers or indexers. They are defined to provide a general model of the domain, with focus on the users´ specific point of view. In general, the domain ontology needs to model the domain in a way that on the one hand the retrieval of pictures becomes more efficient for a user of a multimedia application and on the other hand the included concepts can also be automatically extracted from the multimedia layer. In other words, the concepts have to be recognizable by automatic analysis methods, but need to remain comprehensible for a human.

M-OntoMat-Annotizer framework In order to exploit the ontology infrastructure presented above and annotate the domain ontologies with low-level multimedia descriptors, the usage of a tool is necessary. Our implemented framework is called *M-OntoMat-Annotizer* [1] (M stands for Multimedia) [11]. The development was based on an extension of the CREAM (CREAting Metadata for the Semantic Web) framework [4] and its reference implementation, *OntoMat-Annotizer*[2].

For this reason, the *Visual Descriptor Extraction (VDE)* tool was implemented as a plug-in to OntoMat-Annotizer and is the core component for extending its capabilities and supporting the initialization of domain ontologies with low-level multimedia features. The VDE plug-in manages the overall low-level feature extraction and linking process by communicating with the other components. Using this tool, we manage to build the knowledge base that will serve as the primary reference resource for the multimedia content analysis process presented in the next section.

3 Knowledge-Assisted Multimedia Analysis

The Knowledge-Assisted Analysis system (KAA) includes methods that automatically segment images, video sequences and key frames into areas corresponding to salient semantic objects (e.g. cars, road, people, field, etc), track these objects over time, and provide a flexible infrastructure for further analysis of their relative motion and interactions, as well as object recognition, metadata generation, indexing and retrieval. Recognition is performed by comparing existing semantic descriptions contained in the multimedia-enriched domain ontologies to lower-level features extracted in the signal (image/video), thus identifying objects and their relations in the multimedia content.

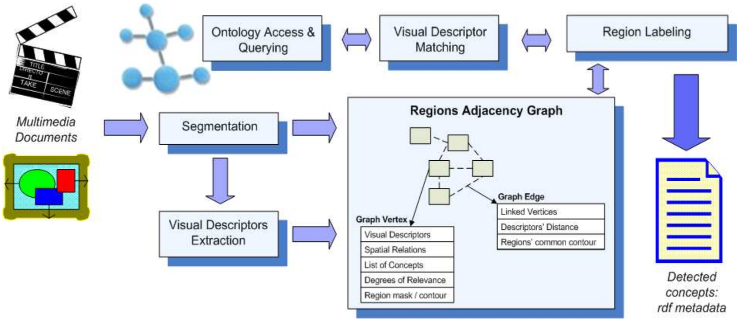

Fig. 2. Knowledge-assisted analysis architecture

A more precise description of the KAA general architecture scheme is given in Fig. 2. The core of the architecture is defined by the region adjacency graph. This graph

[1] see `http://www.acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html`
[2] see `http://annotation.semanticweb.org/ontomat/`

structure holds the region-based representation of the image during the analysis process. During image/video analysis, a set of atom-regions is generated by an initial segmentation. Each node of the graph corresponds to an atom-region and holds the *Dominant Color* and *Region Shape* MPEG-7 visual descriptors extracted for this specific region. The next step for the analysis is to compute a matching distance value between each one of these atom-regions and each one of the prototype instances of all concepts in the domain ontology. This matching distance is evaluated by means of low-level visual descriptors. In order to combine the current two modalities, Dominant Color and Region Shape, in a unique matching distance, we use a neural network approach that provides us with the required distance weighting. Finally, a unique semantic label is assigned to each region corresponding to the concept with minimum distance. Spatial relations (such as “above”, ”below”, ”is included in”...) are extracted for each atom-region. Such information can be further used in a reasoning process in order to refine the semantic labeling. This approach is generic and applicable to any domain as long as new domain ontologies are designed and made available.

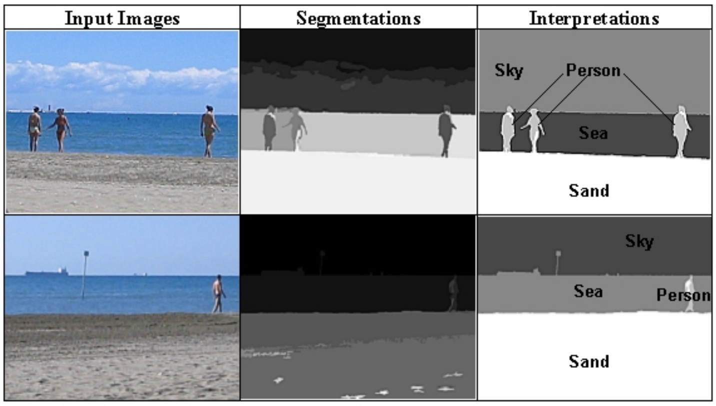

Fig. 3. Holiday-Beach domain results

As illustrated in Fig. 3, the resulting system output is a segmentation mask outlining the semantic description of the scene. The different colors assigned to the generated atom-regions corresponding to the object classes defined in the domain ontology.

4 Conclusion

In this paper, an integrated infrastructure for semantic multimedia content annotation and analysis was presented. This framework comprises ontologies for the description of low-level visual features and for linking these descriptions to concepts in domain ontologies based on a prototype approach. The generation of the visual descriptors and the linking with the domain concepts is embedded in a user-friendly tool, which hides analysis-specific details from the user. Thus, the definition of appropriate visual descriptors can be accomplished by domain experts, without the need to have a deeper understanding of ontologies or low-level multimedia representations.

Finally, despite the early stage of multimedia analysis experiments, first results based on the ontologies presented in this work are promising and show that it is possible to apply the same analysis algorithms to process different kinds of images or video, by simply employing different domain ontologies. Apart from visual descriptions and relations, future focus will concentrate on the reasoning process and the creation of rules in order to detect more complex events. The examination of the interactive process between ontology evolution and use of ontologies for content analysis will also be the target of our future work, in the direction of handling the semantic gap in multimedia content interpretation.

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.

References

1. O. Mich R. Brunelli and C.M. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78–112, 1999.
2. A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12).
3. Siegfried Handschuh and Steffen Staab, editors. *Annotation for the Semantic Web*. IOS Press, 2003.
4. Siegfried Handschuh and Steffen Staab. Cream - creating metadata for the semantic web. *Computer Networks*, 42:579–598, AUG 2003. Elsevier.
5. J. Wielemaker A.Th. Schreiber, B. Dubbeldam and B.J. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, May/June 2001.
6. J. Hunter, J. Drennan, and S. Little. Realizing the hydrogen economy through semantic web technologies. *IEEE Intelligent Systems Journal - Special Issue on eScience*, 19:40–47, 2004.
7. R. Troncy. Integrating Structure and Semantics into Audio-Visual Documents. In *Proceedings of the 2nd International Semantic Web Conference (ISWC 2003)*, October 2003.
8. A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Proceedings of the 13th International Conference on Knowledge Acquisition, Modeling and Management, EKAW 2002*, volume 2473 of *Lecture Notes in Computer Science*, Siguenza, Spain, 2002.
9. ISO/IEC 15938-3 FCD Information Technology - Multimedia Content Description Interface - Part 3: Visual, March 2001, Singapore.
10. ISO/IEC 15938-5 FCD Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Scemes, March 2001, Singapore.
11. S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, S. Handschuh, I. Kompatsiaris, S. Staab, and M.G. Strintzis. Semantic Annotation of Images and Videos for Multimedia Analysis. In *Proceedings of the 2nd European Semantic Web Conference (ESWC 2005)*, May 2005.