M-OntoMat-Annotizer: Linking Ontologies with Multimedia Low-Level Features for Automatic Image Annotation

Carsten Saathoff, Norman Timmermann, Steffen Staab
ISWeb
University of Koblenz, D-56016 Koblenz
{saathoff,normania,staab}@uni-koblenz.de

Kosmas Petridis, Dionysios Anastasopoulos, Yiannis Kompatsiaris
Informatics and Telematics Institute
GR-57001 Thermi-Thessaloniki, Greece
{kosmas,anastas,ikom}@iti.gr

1. INTRODUCTION

Understanding and semantic annotation of multimedia content have been identified as important steps towards more efficient manipulation and retrieval of visual media. Although new multimedia standards, such as 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 is out of the scope of these standards and is still left to the content manager. However, since the generation of annotations manually is tedious and often expensive, methods to automatically annotate images with semantic descriptions are under active research. In the aceMedia IST FP6 project[1], ontologies and Semantic Web technologies are employed in order to achieve a knowledge-based semantic analysis of multimedia content. This allows for more generic algorithms not limited to specific objects, but capable to handle a diverse number of concepts depending on the provided domain knowledge.

In this paper we present M-OntoMat-Annotizer, a tool covering the step of knowledge acquisition for automatic annotation of multimedia content. The tool allows to extract MPEG-7 visual descriptors [1] from both images and videos and to store these descriptors as so-called visual prototypes of ontology classes. The prototypes are stored as RDF instances using a RDF version of the MPEG-7 visual descriptors. The prototype approach specifically provides an OWL-DL friendly way of linking classes to concrete visual characteristics. In the following we first present the overall knowledge-assisted analysis framework, and then continue with M-OntoMat-Annotizer. We conclude with a short outlook on future work.

2. KNOWLEDGE ASSISTED ANALYSIS

Analysis of multimedia content is under active research for several years now, and progress has been made in several application domains, such as person detection or face recognition. However, all these approaches have in common that they are limited to very specific problems and lack the applicability for other domains. In [2], we propose an architecture for automatic annotation of multimedia content that is independent of specific algorithms, but uses ontologies enriched with low-level features to label regions in images with semantic concepts.

In order to handle the semantic gap in multimedia content interpretation, aceMedia proposed and implemented a comprehensive ontology infrastructure. An important part of this infrastructure is the Visual Descriptor Ontology (VDO), developed to link ontology concepts to low-level visual descriptors. It is based on MPEG-7 but modeled in RDFS, which allows for the direct integration with other RDF data used throughout the project. The descriptors are represented as so called *prototypes*, which are instances of the domain concepts linked to specific visual descriptors. The additional super-concept *Prototype* assures that prototypical instances can later be distinguished from the "real" metadata. By using the prototype approach to represent the visual features of concepts, we avoid direct linking of concepts to instances, and the ontologies are kept OWL DL compatible. Details about the aceMedia Knowledge Infrastructure and the VDO in particular can be found in [3].

We will shortly outline the analysis procedure for still images. Initially the image is segmented into a number of regions. For each region the MPEG-7 visual descriptors are extracted and then compared to the prototype instances stored in the active domain ontology. Using this approach, for each domain concept a distance to the descriptors of the region can be computed. This allows to decide which concept provides the best match for the specific region. Finally, the region is labeled with the concept providing the smallest distance. Apparently, the algorithm is domain independent, since it uses a generic distance computation which only relies on the visual descriptors. The concepts that can be detected, and especially the definition of the concepts, are completely defined in the ontologies and the extracted visual prototypes, so that switching the algorithm to another domain could be easily achieved by providing a different domain ontology and according prototypes.

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

Demos and Posters of the *3rd European Semantic Web Conference (ESWC 2006)*, Budva, Montenegro, 11th - 14th June, 2006

3. M-ONTOMAT-ANNOTIZER

In order to exploit the ontology infrastructure mentioned above and enrich the domain ontologies with multimedia descriptors, M-OntoMat-Annotizer (M stands for Multimedia) [3] was implemented. 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 supporting the initialization of RDF 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 OntoMat-Annotizer components.

The VDE Visual Editor and Media Viewer presents a graphical interface for loading and processing of visual content (images and videos), visual feature extraction and linking with domain ontology concepts. The interface, as shown in Fig. 1, seamlessly integrates with the common OntoMat-Annotizer ones. Usually, the user needs to extract the visual features (i.e. descriptors included in the VDO) of a specific object inside the image/frame. M-OntoMat-Annotizer lets the user draw a region of interest in the image/frame and apply the multimedia descriptor extraction only to the specific selected region. Alternatively, M-OntoMat-Annotizer also supports automatic segmentation of the image/frame; whenever a new image/frame is loaded it is automatically segmented into regions. The user can then select a desired region or even merge two or more regions and proceed with the extraction.

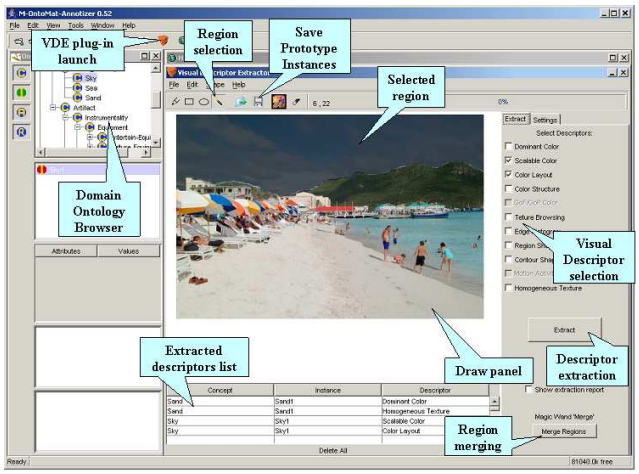

Figure 1: The M-OntoMat-Annotizer user interface

By specifying an instance of a concept in the ontology browser and selecting a region of interest, the user can extract and link appropriate visual descriptor instances with instances of domain concepts that serve as *prototypes* for these concepts. The created statements are added to the knowledge base and can be retrieved in a flexible way during multimedia content analysis. M-OntoMat-Annotizer saves the domain concept prototype instances together with the corresponding descriptors, in a separate RDFS file and leaves the original domain ontology unmodified.

M-OntoMat-Annotizer is publicly available as free software through the aceMedia web site since last May[3]. An updated version of the tool is expected to be published during summer 2006.

4. CONCLUSIONS

In this paper we presented M-OntoMat-Annotizer, a tool for enriching domain ontologies with MPEG-7 visual descriptors expressed in RDF. We also presented an approach to exploit the stored information for the automatic and domain independent annotation of images.

We currently plan further extensions of the tool. The main focus will be on the implementation of a high-level multimedia annotation tool based on M-OntoMat-Annotizer. Apparently this denotes the opposite direction of its current purpose, i.e. not annotating the ontologies with low-level features, but annotating the content with semantic metadata. Obviously, using the current plug-in, annotations could be made on a region level. Especially using the automatic segmentation capability of M-OntoMat-Annotizer, the detailed annotation would become less tedious. Furthermore, the generation of such annotations leads to the second planned extension: the extraction of spatial, topological and contextual knowledge from annotated content that can be used for multimedia reasoning and improve the automatic annotation significantly. Therefore, the tool can both be used by users to annotate their images for later retrieval or organization, but also as a means to generate a-priori knowledge useful for the knowledge-assisted analysis of multimedia content and multimedia reasoning.

Acknowledgments 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] ISO/IEC 15938-3 FCD Information Technology - Multimedia Content Description Interface - Part 3: Visual, March 2001, Singapore.

[2] T. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis and I. Kompatsiaris. Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content. In *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005) at the 4th International Semantic Web Conference (ISWC 2005)*, Nov. 2005.

[3] 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.

[4] Siegfried Handschuh and Steffen Staab. Cream - creating metadata for the semantic web. *Computer Networks*, 42:579–598, AUG 2003. Elsevier.

[2] `http://annotation.semanticweb.org/ontomat/`

[3] `http://www.acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html`