# M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features*

Kosmas Petridis[1], Dionysios Anastasopoulos[1], Carsten Saathoff[2], Norman Timmermann[2], Yiannis Kompatsiaris[1], and Steffen Staab[2]

[1] Informatics and Telematics Institute, GR-57001 Thermi-Thessaloniki, Greece
{kosmas, anastas, ikom}@iti.gr
[2] University of Koblenz, D-56016 Koblenz, Germany
{saathoff, normania, staab}@uni-koblenz.de

**Abstract.** Annotations of multimedia documents typically have been pursued in two different directions. Either previous approaches have focused on low level descriptors, such as *dominant color*, or they have focused on the content dimension and corresponding annotations, such as *person* or *vehicle*. In this paper, we present a software environment to bridge between the two directions. *M-OntoMat-Annotizer* allows for linking low level MPEG-7 visual descriptions to conventional Semantic Web ontologies and annotations. We use *M-OntoMat-Annotizer* in order to construct ontologies that include prototypical instances of high-level domain concepts together with a formal specification of corresponding visual descriptors. Thus, we formalize the interrelationship of high- and low-level multimedia concept descriptions allowing for new kinds of multimedia content analysis, reasoning and retrieval.

## 1 Introduction

Representation 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-4 and MPEG-7 [1], provide important functionalities for the 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 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 current multimedia analysis methods and tools on the one hand and semantic annotation methods and tools on the other hand. State-of-the-art multimedia analysis systems are severely limiting themselves by resorting mostly to visual descriptions at a very low level, e.g.

* 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.

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part III, LNAI 4253, pp. 633–640, 2006.

the *dominant color* of a picture. This may be observed even though the need for semantic descriptions that help to bridge the so called *semantic gap* has been acknowledged for a long time [2, 3]. At the same time, the semantic annotation community has only recently started to work into the direction of semantic annotation in the multimedia domain and still remains a long way to go. Work in semantic annotation currently addresses mainly textual resources [4] or simple annotation of photographs [5, 6].

Acknowledging both the relevance of low-level visual descriptions as well as a formal, uniform machine-processable representation [7], we here try to bridge the chasm by providing a semantic annotation framework and corresponding tool, *M-OntoMat Annotizer*, for elicitating and representing knowledge both about the *content domain* and the *visual characteristics* of multimedia data itself. In the framework we propose, this link between the MPEG-7 visual descriptors and domain concepts is made explicit by means of a conceptualization based on a *prototyping* approach. The core idea of our approach lies in a way to associate concepts with instances that are deemed to be prototypical by their annotators with regard to their visual characteristics. To establish this semantic link we have implemented our framwork in a user-friendly annotation tool, *M-OntoMat-Annotizer*, extending our previous framework for semantic annotations of text [4]. The tool has been built in order to allow content providers to annotate visual descriptors without prior expertise in semantic web technologies or multimedia analysis. More specifically, it allows to extract MPEG-7 visual descriptors 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 [8]. The prototype approach specifically provides an OWL-DL friendly way of linking classes to concrete visual characteristics.

The existence of such a knowledge base may be exploited in a variety of ways. In particular, we envision its exploitation in two modes:

*(1) Direct exploitation:* In this mode, an application uses the knowledge base directly. For instance, during the semantic annotation process one may gather information like *the blue cotton cloth 4711 in image 12 has a rippled texture described by values 12346546.* Such kind of semantic knowledge may be used later, e.g. for combined retrieval by semantics and similarity in an internet shop. Obviously, such kind of knowledge is expensive to be acquired manually, even when resorting to a user friendly tool. Thus, this kind of knowledge may only be provided for valuable data, such as images or videos of commercial products or of items from museum archives.

*(2) Indirect exploitation:* In this mode, the a-priori knowledge base 'only' serves as a data set provided to prepare an automatic multimedia analysis tool. For instance, consider the provider of a sports portal offering powerful access to his database on tennis, soccer, etc. He uses semantic annotation of multimedia images or videos in order to prepare an analysis system. For instance, he uses *M-OntoMat-Annotizer* in order to describe the shape and the texture of tennis balls,

rackets, nets, or courts and he feeds these descriptions into an analysis system. The system uses the descriptions in order to learn how to tag and relate segments of images and video keyframes with domain ontology concepts. A customer at the portal may then ask the system what it could derive about the images and the videos, e.g. he could ask for all the scenes in which a ball touches a line in a tennis court.

Our long term objectives are dedicated to the indirect exploitation of semantic multimedia annotation as presented in the second paragraph, which is an ongoing comprehensive and complex endeavor, providing a flexible infrastructure for further multimedia content analysis and reasoning, object recognition, metadata generation, indexing and retrieval. In the context of this paper, we only sketch the main steps of our approach.

The remainder of the paper is organized as follows: after briefly studying related work in section 2, we present in section 3 a description of the knowledge-assisted analysis process, which exploits the developed infrastructure and annotation framework. The annotation process needed for initializing the knowledge base with prototype instances of domain concepts in question, including a description of the actual implementation of the *M-OntoMat-Annotizer* tool are given in section 4. We conclude with a summary of our work and future directions in section 5.

## 2 Related Work

In the *multimedia analysis* area, knowledge about multimedia content domains, as for example reported in [9], is a promising approach by which higher level semantics can be incorporated into techniques that capture the semantics through automatic parsing of multimedia content.

Such techniques are turning to knowledge management approaches, including Semantic Web technologies to solve this problem [10]. In [11], semantic entities, in the context of the MPEG-7 standard, are used for knowledge-assisted video analysis and object detection, thus allowing for semantic level indexing. In [12] a framework for learning intermediate level visual descriptions of objects organized in an ontology is presented that aid the system to detect domain objects.

In [13], a-priori knowledge representation models are used as a knowledge base that assists semantic-based classification and clustering. MPEG-7 compliant low-level descriptors are automatically mapped to appropriate intermediate-level descriptors forming a simple vocabulary termed object ontology. Additionally, an object ontology is introduced to facilitate the mapping of low-level to high-level features and allow the definition of relationships between pieces of multimedia information. This ontology paradigm is coupled with a relevance feedback mechanism to allow for precision in retrieving the desired content.

Work in *semantic annotation* [14] has so far mainly focused on textual resources [4] or simple annotation of photographs [5, 6]. A presentation of an earlier version of *M-OntoMat-Annotizer* can be found in [15].

## 3 Knowledge Assisted Analysis

In order to handle the semantic gap in multimedia content interpretation, the aceMedia IST-FP6 Integrated project[1] 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 superconcept *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 [15].

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.

## 4 M-OntoMat-Annotizer

In order to exploit the ontology infrastructure mentioned above and enrich the domain ontologies with multimedia descriptors, we have developed the M-OntoMat-Annotizer (M stands for Multimedia), extending the CREAM (CREAting Metadata for the Semantic Web) framework [4] and its reference implementation, OntoMat-Annotizer[2], with the Visual Descriptor Extraction (VDE) tool, in order to allow low-level feature annotation. Figure 1 shows the integrated architecture the modules of which are explained in the following in more detail.

**Core OntoMat-Annotizer.** OntoMat-Annotizer supports two core applications: (i) it is used as an annotation tool for web pages and (ii) it acts as the basis of an ontology engineering environment. Also, by providing a flexible plug-in interface it offers the possibility to implement new components and extend the core functionality of OntoMat-Annotizer.

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

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

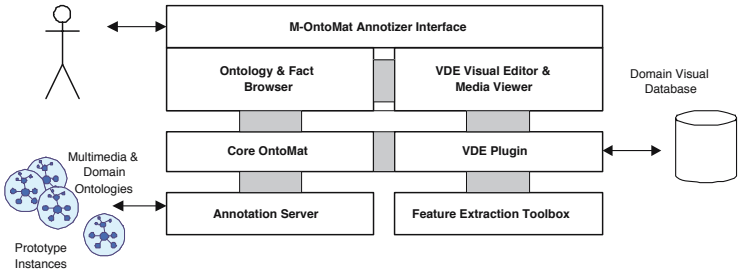


**Fig. 1.** M-OntoMat-Annotizer and VDE plug-in design architecture

**Annotation Server.** The annotation server acts in the background and stores the entities of the knowledge base, maintains their mutual references and is responsible for maintaining the overall integrity of the stored entities.

**Domain Visual Database.** As easy content access is crucial for annotation and content analysis processes, a visual database containing content related to the domain examined and analyzed is always necessary.

**Feature Extraction Toolbox.** The actual extraction of the visual descriptors is performed using a feature extraction toolbox, namely the *aceToolbox*, a content pre-processing and feature extraction toolbox developed inside aceMedia project. The aceToolbox saves the extracted MPEG-7 Descriptors in XML format.

**VDE Visual Editor and Media Viewer.** 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 Figure 2, seamlessly integrates with the common OntoMat interfaces. Usually, the user needs to extract the features (multimedia descriptors) of a specific object inside the image/frame. For this reason, the VDE application lets the user draw a region of interest in the image/frame and apply the multimedia descriptors extraction procedure 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. By selecting a specific concept in the OntoMat ontology browser and selecting a region of interest the user can extract and link concepts with appropriate prototype instances by means of the underlying functionalities of the VDE plugin.

**VDE Plug-in.** The *Visual Descriptor Extraction* (VDE) tool is implemented as a plug-in to OntoMat-Annotizer and is the core component for extending its capabilities and supporting the initialization of ontologies with low-level multimedia features. The VDE plugin manages the overall low-level feature extraction and linking process by communicating with the other components.

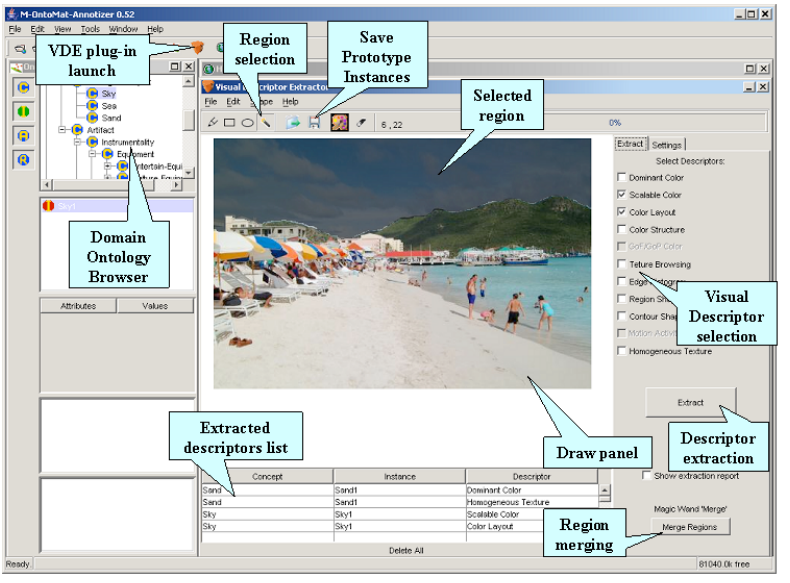


**Fig. 2.** The VDE plugin into M-OntoMat-Annotizer user interface

### 4.1 VDE Extraction and Annotation Process

Triggered by the users extraction command, the VDE plugin *extracts* the requested MPEG-7 Descriptors through calls to the Feature Extraction Toolbox. The VDE plug-in supports the *transformation* of the extracted XML multimedia resources into instances of the visual descriptors defined in the VDO, by means of an XSL transformation specification that creates a corresponding descriptor instance for each extraction, which is handed to the knowledge base on the annotation server:

```
<!-- Values of "ScalableColorDescriptor" instance "VDE_INST_0001"-->
<vdo:ScalableColorDescriptor
  rdf:ID="http://www.acemedia.org/ontologies/VDO#VDE_INST_0001">
    <vdo:coefficients> 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 </vdo:coefficients>
    <vdo:numberOfBitPlanesDiscarded> 6 </vdo:numberOfBitPlanesDiscarded>
    <vdo:numberOfCoefficients> 0 </vdo:numberOfCoefficients>
</vdo:ScalableColorDescriptor>
```

Then, the VDE automatically links the newly created visual descriptor instance with the selected domain concept prototype instance on the annotation server:

```
<!-- A prototype instance "Ball_Prototype_1" definition of tennis
concept "Ball". "Ball_Prototype_1" instance "hasDescriptor"
"VDE_INST_0001", which corresponds to the scalable color descriptor
instance above -->
<vdoext:Prototype
  rdf:ID="http://www.acemedia.org/ontologies/VDO-EXT#Ball_Prototype_1">
    <rdf:type rdf:resource="http://www.acemedia.org/ontologies/TENNIS#Ball"/>
    <vdoext:hasDescriptor
      rdf:resource="http://www.acemedia.org/ontologies/VDO#VDE_INST_0001"/>
</vdoext:Prototype>
```

As seen above, these prototype instances are not only instances of the domain concept in question but are also stated to be instances of a separate Prototype concept. The created statements are added to the knowledge base and can be retrieved in a flexible way during analysis. The necessary conceptualizations can be seen as extensions to the VDO (*VDO-EXT ontology*) that link to the core ontology and are implemented in RDF:

```
<!-- Definition of concept "Prototype" as a subclass of Dolce's
"Physical-Object" -->
<rdfs:Class
  rdf:ID="http://www.acemedia.org/ontologies/VDO-EXT#Prototype">
    <rdfs:subClassOf rdf:resource=
      "http://ontology.ip.rm.cnr.it/ontologies/DOLCE-Lite#Physical-Object"/>
</rdfs:Class>

<!-- Definition of relation "hasDescriptor" having Dolce's
"Physical-Object" as domain and VDO's "VisualDescriptor" as range-->
<rdf:Property
  rdf:ID="http://www.acemedia.org/ontologies/VDO-EXT#hasDescriptor ">
    <rdfs:domain rdf:resource=
      "http://ontology.ip.rm.cnr.it/ontologies/DOLCE-Lite#Physical-Object"/>
    <rdfs:range
      rdf:resource="http://www.acemedia.org/ontologies/VDO#VisualDescriptor"/>
</rdf:Property>
```

All the prototype instances can be saved in a RDFS file. The VDE tool saves the domain concept prototype instances together with the corresponding transformed descriptors, *separately* from the ontology file, thus leaving the original domain ontology unmodified.

M-OntoMat-Annotizer is publicly available as free software through the aceMedia web site since May 2005[3]. An updated version of the tool is expected to be published during summer 2006.

## 5 Conclusions

In this paper we presented M-OntoMat-Annotizer, a tool for enriching domain ontologies with MPEG-7 visual descriptors expressed in RDF. We currently plan further extensions of the tool. One direction is the implementation of a high-level multimedia annotation tool based on M-OntoMat-Annotizer. 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, another direction is 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.

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

## References


[1] S.-F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6):688–695, June 2001.

[2] O. Mich R. Brunelli and C.M. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78–112, 1999.

[3] 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).

[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] L. Hollink, A.Th. Schreiber, J. Wielemaker, and B. Wielinga. Semantic annotation of image collections. In *Proceedings of the K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation, Florida*, 2003.

[7] P. Wittenburg D. Thierry and H. Cunningham. The Automatic Generation of Formal Annotations in a Multimedia Indexing and Searching Environment. In *Proc. ACL/EACL Workshop on Human Language Technology and Knowledge Management*, Toulouse, France, 2001.

[8] 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.

[9] 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.

[10] A. Yoshitaka, S. Kishida, M. Hirakawa, and T. Ichikawa. Knowledge-assisted content-based retrieval for multimedia databases. *IEEE Multimedia*, 1(4):12–21, Winter 1994.

[11] R. Tansley, C. Bird, W. Hall, P. Lewis, and M. Weal. Automating the linking of content and concept. In *Proc. ACM Int. Multimedia Conf. and Exhibition (ACM MM-2000)*, Oct./Nov. 2000.

[12] Nicolas Maillot, Monique Thonnat, and Céline Hudelot. Ontology based object learning and recognition: Application to image retrieval. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004), 15-17 November 2004, Boca Raton, FL, USA*, pages 620–625, 2004.

[13] I. Kompatsiaris, V. Mezaris, and M. G. Strintzis. *Multimedia content indexing and retrieval using an object ontology.* Multimedia Content and Semantic Web - Methods, Standards and Tools, Editor G.Stamou, Wiley, New York, NY, 2004.

[14] Siegfried Handschuh and Steffen Staab, editors. *Annotation for the Semantic Web.* IOS Press, 2003.

[15] 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.