KNOWLEDGE REPRESENTATION FOR SEMANTIC MULTIMEDIA CONTENT ANALYSIS AND REASONING ⋆

K. PETRIDIS, I. KOMPATSIARIS AND M. G. STRINTZIS
Informatics and Telematics Institute
1st Km Thermi-Panorama Rd
GR-57001 Thermi-Thessaloniki, Greece
E-mail: kosmas, ikom, strintzi@iti.gr

S. BLOEHDORN, S. HANDSCHUH AND S. STAAB
University of Karlsruhe
Institute AIFB
D-76128 Karlsruhe, Germany
E-mail: sbl, sha, sst@aifb.uni-karlsruhe.de

N. SIMOU, V. TZOUVARAS AND Y. AVRITHIS
National Technical University of Athens
School of Electrical and Computer Engineering
9 Iroon Polytechniou Str., 157 73 Zographou, Athens, Greece
E-mail: nsimou, tzouvaras, iavr@image.ntua.gr

In this paper, a knowledge representation infrastructure for semantic multimedia content analysis and reasoning is presented. This is one of the major objectives of the aceMedia Integrated Project where ontologies are being extended and enriched to include low-level audiovisual features, descriptors and behavioural models in order to support automatic content annotation. More specifically, the developed infrastructure consists of the *core ontology* based on extensions of the DOLCE core ontology and the multimedia-specific infrastructure components. These are, the *Visual Descriptors Ontology*, which is based on an RDFS representation of the MPEG-7 Visual Descriptors and the *Multimedia Structure Ontology*, based on the MPEG-7 MDS. Furthermore, the developed *Visual Descriptor Extraction* tool is presented, which will support the initialization of domain ontologies with multimedia features.

⋆ This work was supported by the European Commission under contract FP6-001765 aceMedia (URL: http://www.acemedia.org).

1 Introduction

Video understanding and semantic information extraction 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, provide important functionalities such as manipulation and transmission of objects and metadata, their automatic extraction of metadata, and most importantly at a semantic level, is out of the scope of the standards and is left to the content developer.

In well-structured specific applications (e.g. sports and news broadcasting) domain-specific features that facilitate the modelling of higher level semantics can be extracted [1]. A priori knowledge representation models are used as a knowledge base that assists semantic-based classification and clustering [2]. In [3], 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 [4], fuzzy ontological relations and context-aware fuzzy hierarchical clustering are employed to interpret multimedia content for the purpose of automatic thematic categorization of multimedia documents. In [5] MPEG-7 compliant low-level descriptors are automatically mapped to appropriate intermediate-level descriptors forming a simple vocabulary termed object ontology. In [6] the problem of bridging the gap between low-level representation and high-level semantics is formulated as a probabilistic pattern recognition problem. Finally, in [7] and [8], hybrid methods extending the query-by-example strategy are developed.

Due to the limitations of the state of the art multimedia analysis systems [9], it is acknowledged that in order to achieve semantic analysis and knowledge mining from multimedia content, ontologies [10] are essential to express the key entities and relationships describing multimedia in a formal machine-processable representation [11]. Ontology modelling and ontology-based metadata creation currently address mainly textual resources [12] or simple annotation of photographs [13]. aceMedia investigates ontology modelling in terms of both methodology and expressiveness in order to address the additional requirements of multimedia resources [14]. The project will advance the state of the art by applying ontology-based discourse structure and analysis to multimedia resources.

In this paper, the aceMedia knowledge representation infrastructure for semantic multimedia content analysis and reasoning is presented. Explicit knowledge representation thereby aims among others at supporting audio-visual content analysis and object/event recognition, the creation of knowledge beyond object and scene recognition through reasoning processes and at enabling user-friendly and intelligent search and retrieval. The presented knowledge infrastructure consists of several parts, namely the *core ontology* as a basis for all components of the knowledge infrastructure, multimedia-specific conceptualizations in form of the *visual descriptor ontology* and the *multimedia structure ontology* and a user-friendly *visual descriptor extraction* tool that allows the initialization of domain ontologies with visual features.

It should be emphasized that this work, including both the developed ontologies and the visual descriptor extraction tool, are presented as a multimedia knowledge infrastructure that, in the framework of aceMedia, will support a wide range of applications, including knowledge-assisted analysis, context analysis, reasoning, semantic search and personalization. At the current stage of development, no testbed is available for benchmarking or experimental validation of the proposed representation approach. A first kind of such validation will be carried out as part of the knowledge-assisted analysis development, where visual descriptors of extracted objects in the multimedia content will be matched against prototype instances in the domain ontologies to assist the recognition process. Analysis will be context-sensitive, referring to alignment of heterogeneous ontologies [15] as well as visual scene context [16].

The remainder of the paper is organized as follows: in Section 2, an overview over the knowledge infrastructure, the utilized knowledge representation languages and the core ontology is given. Section 3 describes the multimedia related ontologies and structures, i.e. the visual descriptor ontology and the multimedia structure ontology for ACE metadata, while in Section 4 the visual descriptor extraction tool is presented. We conclude with final remarks and an outlook in Section 5.

2 Knowledge Infrastructure

The challenge in building a knowledge infrastructure for multimedia analysis and annotation arises from two main factors. On one hand, when reasoning with multimedia data on a large scale, reasoners have to deal with large numbers of instantiations of the concepts and properties defined in the ontologies. On the other hand, multimedia data comes in two separate though intertwined layers: The *multimedia layer* deals with the semantics of properties and phenomena related to the representation of content within the media-data itself, e.g. its spatio-temporal structure or visual features for analysis. The *content layer* deals with the semantics of the actual content contained in the media data as it is perceived by the human media consumer. The knowledge infrastructure should model the multimedia layer data so as to support extraction and inferencing of content layer data.

Ontology Architecture This hybrid structure of the multimedia analysis task must be necessarily reflected in the ontology architecture. In this paper, we

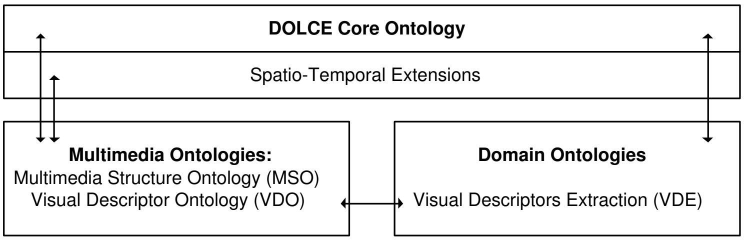

Fig. 1. Ontology Structure Overview

propose an ontology architecture that integrates these two layers. Fig. 1 summarizes the developed knowledge infrastructure consisting of:

- The *Multimedia Ontologies*, which model the multimedia layer data, especially the visualizations in still images and videos in terms of low-level features and media structure descriptions. Structure and semantics have been carefully modelled to be largely consistent with the existing MPEG-7 multimedia description standard. The corresponding structures are the *Visual Descriptors Ontology* (VDO), which is based on an RDF representation of the MPEG-7 Visual Descriptors and the *Multimedia Structure Ontology* (MSO), based on the MPEG-7 MDS.
- The *Domain Ontologies* that provide the necessary conceptualizations of the content layer, for the specific aceMedia application domains. For the initialization of the domain ontologies with prototype instances that link to representative multimedia features, the *Visual Descriptor Extraction* (VDE) tool has been developed.
- The *Core Ontology* that provides modelling of primitives at the root of the concept hierarchy and can be exploited by both types of ontologies. It is also meant to provide the necessary bridging between the other ontologies within the architecture.

Knowledge Representation Formalisms Several knowledge representation languages have been developed during the last years as ontology languages in the context of the Semantic Web, each with varying characteristics in terms of their expressiveness, ease of use and computational complexity. The aceMedia knowledge infrastructure is focused around two current languages:

Resource Description Framework Schema (RDFS) is a simple modelling language on top of the Resource Description Framework (RDF) formalism[1], both being developed by the W3C. RDFS offers primitives for defining knowledge models that are close to frame-based approaches. RDFS introduces classes, subsumption relationships on both classes and properties and global domain and range restrictions for properties as modelling primitives. The resulting formalism allows to model ontologies as a taxonomic structure of concepts with attributes and relations to other concepts as properties attached to each concept.

Web Ontology Language (OWL), a language inspired by description logics and also developed by the W3C, is designed to be used by applications that need increased expressive power compared to that supported by RDFS, by providing additional vocabulary along with formal semantics. OWL itself has three sub-languages with varying expressivity.

At the current state of the aceMedia project, the knowledge infrastructure is set up using RDFS, Rules and Axioms complementing the RDFS models are provided using F-Logic [17]. This approach is expected to be complemented by using an appropriate sub-language of OWL at a later stage. This decision also reflects that a full usage of the increased expressiveness of OWL requires specialized and more advanced inference engines, especially when dealing with large numbers of instances as in aceMedia. Existing inference engines are not able to consider all features provided by the different OWL dialects or do not allow to do this efficiently. On the other hand, reasoning on the purely conceptual level as supported by most OWL reasoners is of less interest for multimedia analysis.

Core Ontology The role of core ontologies 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. Core ontologies are typically conceptualizations that contain specifications of domain independent concepts and relations based on formal principles derived from philosophy, mathematics, linguistics and psychology.

The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [18] was explicitly designed as such an ontology. It is minimal in that it includes only the most reusable and widely applicable upper-level categories, rigorous in terms of axiomatization and extensively researched and documented. The RDFS version of DOLCE currently contains about 79 high level concepts and 81 high level properties among them.

[1] RDF itself is not a Knowledge Representation system but tries to improve data interoperability on the Web. This is achieved by specializing the XML data model through a graph-based data model similar to the semantic networks formalism.

In the context of aceMedia, DOLCE has been chosen as core ontology for several reasons. DOLCE is rich and consistent in its internal structure and axiomatization. The design of DOLCE builds on insights from many scientific disciplines, especially from philosophy, mathematics and psychology. DOLCE has been carefully designed to be reusable and exhibits an inherent modular nature. In fact, being part of the WonderWeb Foundational Ontology Library, DOLCE is intended to be mapped to other foundational ontologies (possibly more suitable for certain applications), and is supposed to be extended with many modules covering different domains, tasks and/or lexical resources. DOLCE contains explicit conceptualizations relevant for aceMedia by including the concept of qualities that can be perceived and spatio-temporal concepts descriptions.

Although the DOLCE core ontology provides means for representing spatio-temporal qualities, reasoning with such descriptions requires the coding of additional relations that describe the relationship between space and/or time regions. Based on concepts taken from the 'Region Connecting Calculus' [19], Allen's interval calculus [20] and directional models [21] [22], we have carefully extended DOLCE to accomodate the corresponding directional and topological relationships in the spatial and temporal domains. Directional spatial relations describe how visual segments are placed and relate to each other in 2D or 3D space (e.g., left and above). Topological spatial relations describe how the spatial boundaries of the segments relate (e.g., touches and overlaps). In a similar way, temporal segment relations are used to represent temporal relationships among segments or events; the normative binary temporal relations correspond to Allen's temporal interval relations.

3 Multimedia Ontology and Structure

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)[23] 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 are a development tool which duplicate information in the XML schemas, led to the decision of excluding them from the final specification. Although the lack of 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 a data model and a definition of the semantics of MPEG-7 terms and 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. To simplify the process, we used a core subset of the MPEG-7 specification together with a top-down approach to generate the ontology. Hence, our approach was to firstly determine the basic multimedia entities (classes) and their hierarchies from the Multimedia Description Scheme basic entities [24]. Next, the structural hierarchies were determined from the Segment Description Schemes and the different multimedia and generic properties associated with the multimedia entities in order to construct a Multimedia Structure Ontology (MSO). Finally, a Visual Descriptors Ontology (VDO) has been constructed, which contains the representations of the MPEG-7 visual descriptors.

Multimedia Structure Ontology The top-level multimedia content entities of the MSO are described in MDS FCD [24]. Within MPEG-7, multimedia content is classified into five types: Image, Video, Audio, Audiovisual and Multimedia. Each of these types has its own segment subclasses. MPEG-7 provides a number of tools for describing the structure of multimedia content in time and space. The Segment DS (Section 11 of [24]) describes a spatial and/or temporal fragment of multimedia content. A number of specialized subclasses are derived from the generic Segment DS. These subclasses describe the specific types of multimedia segments, such as video segments, moving regions, still regions and mosaics, which result from spatial, temporal and spatiotemporal segmentation of the different multimedia content types. Multimedia resources can be segmented or decomposed into sub-segments through four types of decomposition:

– Spatial Decomposition - e.g., spatial regions within an image;
– Temporal Decomposition - e.g., temporal video segments within a video;
– Spatiotemporal Decomposition - e.g., moving regions within a video;
– Media Source Decomposition - e.g., the different tracks within an audio file or the different media objects

Visual Descriptor Ontology The VDO, contains a set of visual descriptors to be used for knowledge-assisted analysis of multimedia content. 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). The VDO is divided into two main sections. The first contains the descriptors that can be extracted using the content preprocessing toolbox described in the following section and will form the core of the VDO. The second contains useful additional descriptors, whose extraction may be considered in next versions of the content preprocessing toolbox. The core set of descriptors is a subset of the descriptors standardized by MPEG-7 and were selected as the most appropriate (automatically extractable, useful for automatic analysis) among the others for the creation of the VDO. The entire VDO follows closely the specification of the MPEG-7 Visual Part [25], but several modifications were carried out in order to adapt to the datatype representations available in RDFS.

4 Visual Descriptor Extraction Tool

The extraction of a set of prototypical low-level visual descriptors for different domain concepts and their integration into the ontology structure poses two issues.

Firstly, on the conceptual side, it must be clear how domain concepts can be linked with actual instance data without having to cope with meta-modelling. The approach we have adopted is pragmatical easily extendible and conceptually clean. We propose to enrich the knowledge base with instances of domain concepts that serve as *prototypes* for these concepts. Each of these is linked to the appropriate visual descriptor instances.

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 corresponding statements distinguish prototype instances from ordinary instances of the domain concepts that are added to the knowledge base, e.g. after the successful application of the analysis algorithms. The prototype instances are linked to the regions for which the prototypical descriptors have been computed by means of the relation depicted-by. 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 that link to the core ontology and are implemented in RDF:

```
<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>
```

```
<rdf:Property
  rdf:ID="http://www.acemedia.org/ontologies/VDO-EXT#HAS-DESCRIPTOR ">
    <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>
```

Secondly, tool support is needed to create these statements from actual data in a user friendly-way. Therefore, the purpose of the needed tool is to provide a friendly graphical user interface allowing management of multimedia content (images and videos), the extraction of visual descriptor instances and linking these with domain ontology prototype instances.

Currently, there are many existing tools for ontology-based annotation. Most of them are document-centric, therefore they only support textual annotation of web pages and ontologies and there is no solution available so far, for the annotation of multimedia resources with low-level visual features. Thus, there is a need for the development of a tool that bridges and integrates the domain concepts and the low-level multimedia content description features.

In the aceMedia project, it has been decided to extend the CREAM (CREAting Metadata for the Semantic Web) framework [26], in order to develop the multimedia ontology annotation tool mentioned above. OntoMat-Annotizer is the reference implementation of CREAM. It is Java-based and developed in the research field of ontology-based knowledge representation. There now exist two applications for the OntoMat-Annotizer framework: (i) it is used as an annotation tool for web pages and (ii) it acts as the basis of an ontology engineering environment. Moreover, OntoMat-Annotizer, provides a flexible plug-in interface for further application extensions called OntoPlugin and offers the possibility to implement new components and extend the core functionality of OntoMat-Annotizer.

The tool is implemented as a plug-in to OntoMat-Annotizer, extends its capabilities and supports initialization of ontologies with low-level multimedia features. This plug-in is called Visual Descriptor Extractor (VDE). Fig. 2 demonstrates the design architecture that has been followed, in order to achieve such an implementation.

The VDE supports loading and processing of visual content (images and videos). 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 and apply the multimedia descriptors extraction procedure only to the specific selected region.

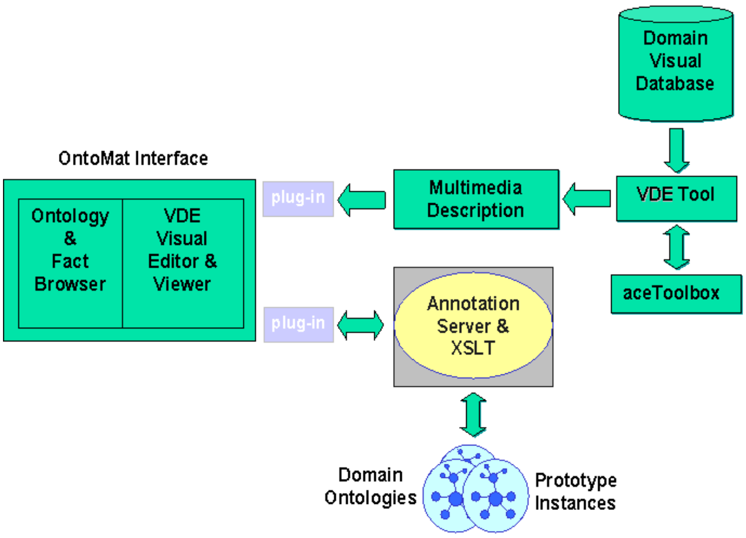

Fig. 2. OntoMat-Annotizer and VDE plug-in design architecture

The VDE supports the extraction of the core MPEG-7 Descriptors included in the VDO. For this purpose, it uses the aceToolbox, a content pre-processing toolbox developed inside aceMedia project, which is responsible for the low-level analysis and MPEG-7 feature extraction. A sample XML output of the MPEG-7 Contour Shape Descriptor follows:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001schema/Mpeg7-2001.xsd">
    <DescriptionUnit xsi:type="DescriptorCollectionType">
        <Descriptor xsi:type="ContourShapeType">
            <GlobalCurvature>9  7 </GlobalCurvature>
            <PrototypeCurvature>3  9 </PrototypeCurvature>
            <HighestPeakY>31</HighestPeakY>
            <Peak peakX="16" peakY="6"/>
            <Peak peakX="20" peakY="1"/>
            <Peak peakX="10" peakY="5"/>
            <Peak peakX="6"  peakY="6"/>
            <Peak peakX="49" peakY="6"/>
            <Peak peakX="5"  peakY="7"/>
            <Peak peakX="55" peakY="7"/>
            <Peak peakX="45" peakY="6"/>
            <Peak peakX="1"  peakY="7"/>
        </Descriptor>
    </DescriptionUnit>
</Mpeg7>
```

As described above, the VDE is based on the OntoPlugin interface and has been implemented as a plug-in in the OntoMat-Annotizer application. The VDE plug-in supports the transformation of the extracted multimedia resources (MPEG-7 Descriptors in XML format) into instances of the visual descriptors defined in the VDO by means of an XSL transformation specification. The selection of the appropriate domain concept is performed by the user himself and guides the tool in producing the appropriate domain concept prototype instances and linking these with the newly created instances of visual descriptors.

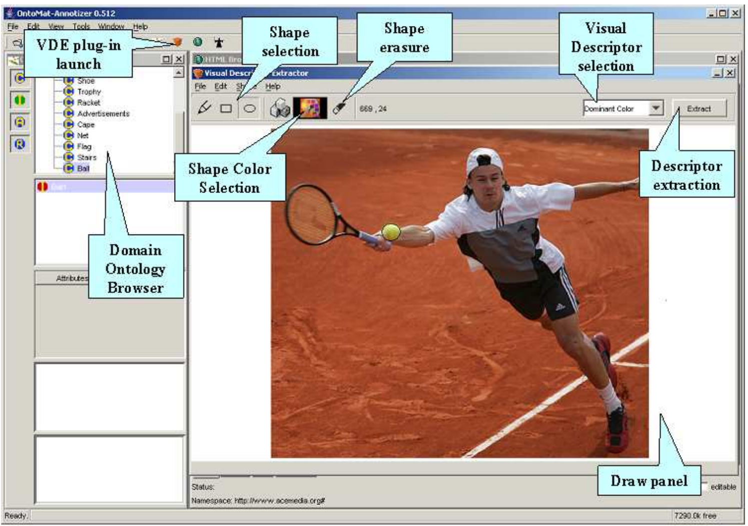

Fig. 3. The VDE plugin into OntoMat-Annotizer tool

The graphical interface of the integrated VDE tool into OntoMat-Annotizer is presented in Fig. 3. After loading the domain ontology in the ontology browser, the user is able to select a specific concept and create a prototype instance of this concept. Then, by drawing the region of interest inside the loaded image and extracting a visual descriptor, the specific prototype instance is automatically linked with the extracted visual descriptor values. This procedure is succeeded in the VDE by implementing the ontology linking approach described at the beginning of this section.

5 Conclusions

In this paper, the aceMedia knowledge representation infrastructure for semantic multimedia content analysis and reasoning was presented. In aceMedia, ontologies are extended and enriched through the Visual Descriptors Ontology and the Multimedia Structure Ontology to include low-level audiovisual descriptors in order to support automatic content annotation. Appropriate knowledge representation formalisms, a core ontology and a user-friendly tool that allows the initialization of domain ontologies with visual features have been developed to complete this infrastructure.

The presented knowledge representation will be used for audio-visual content analysis and object/event recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia content search and retrieval. Based on this infrastructure, aceMedia targets knowledge discovery and embedded self-adaptability to enable content to be self organising, self annotating, self associating; more readily searched (faster, more relevant results); and adaptable to user requirements (self reformatting). The next step in this direction is certainly the development of a testbed for experimental validation of the proposed representation approach, as part of knowledge-assisted content analysis.

References

1. W. Al-Khatib, Y.F. Day, A. Ghafoor, and P.B. Berra. Semantic modeling and knowledge representation in multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):64–80, Jan/Feb 1999.
2. 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.
3. G. Tsechpenakis, G. Akrivas, G. Andreou, G. Stamou, and S.D. Kollias. Knowledge-Assisted Video Analysis and Object Detection. In *Proc. European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems (Eunite02)*, Algarve, Portugal, September 2002.
4. M. Wallace, G. Akrivas, P. Mylonas, Y. Avrithis, and S. Kollias. Using Context and Fuzzy Relations to Interpret Multimedia Content. In *Proc. 3rd International Workshop on Content-Based Multimedia Indexing, CBMI'03*, Rennes, France, September 22-24 2003.
5. V. Mezaris, I. Kompatsiaris, N. V. Boulgouris, and M. G. Strintzis. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Audio and Video Analysis for Multimedia Interactive Services*, 14(5):606–621, May 2004.
6. M. Ramesh Naphade, I.V. Kozintsev, and T.S. Huang. A factor graph framework for semantic video indexing. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(1):40–52, Jan. 2002.

7. W. Chen and S.-F. Chang. VISMap: an interactive image/video retrieval system using visualization and concept maps. In *Proc. IEEE Int. Conf. on Image Processing*, volume 3, pages 588–591, 2001.
8. S.S.M. Chan, L. Qing, Y. Wu, and Y. Zhuang. Accommodating hybrid retrieval in a comprehensive video database management system. *IEEE Trans. on Multimedia*, 4(2):146–159, June 2002.
9. O. Mich R. Brunelli and C.M. Modena. A survey on video indexing. *Journal of Visual Communications and Image Representation*, 10:78–112, 1999.
10. Steffen Staab and Rudi Studer. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer Verlag, Heidelberg, 2004.
11. 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.
12. H.-P. Schnurr M. Erdmann, A. Maedche and S. Staab. From Manual to Semi-automatic Semantic Annotation: About Ontology-based Text Annotation Tools. In P. Buitelaar and K. Hasida, editors, *Proc. COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, Luxembourg, August 5-6 2000.
13. J. Wielemaker A.Th. Schreiber, B. Dubbeldam and B.J. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, May/June 2001.
14. J. Hunter. Adding Multimedia to the Semantic Web: Building an MPEG-7 Ontology. In *Proc. The First Semantic Web Working Symposium, SWWS'01*, Stanford University, California, USA, July 2001.
15. P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. C-OWL: Contextualizing ontologies. In *Proc. 2nd International Semantic Web Conference, ISWC'03*, 2003.
16. A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context-based vision system for place and object recognition. In *Proc. Intl. Conf. on Computer Vision, ICCV'03*, 2003.
17. Michael Kifer and Georg Lausen. F-logic: a higher-order language for reasoning about objects, inheritance, and scheme. In *Proceedings of the 1989 ACM SIGMOD international conference on Management of data*, pages 134–146. ACM Press, 1989.
18. 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.
19. A. Cohn, B. Bennett, J. M. Gooday, and N. M. Gotts. *Representing and Reasoning with Qualitative Spatial Relations about Regions*, pages 97–134. Kluwer Academic Publishers, 1997.
20. J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(1):832–843, 1983.
21. Dimitris Papadias and Yannis Theodoridis. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science*, 11:111–138, 1997.

22. S. Skiadopoulos and M. Koubarakis. Composing cardinal direction relations. *Artificial Intelligence*, 152:143–171, 2004.
23. UML Resource Center, http://rational.com/uml/index.jsp.
24. ISO/IEC 15938-5 FCD Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Scemes, March 2001, Singapore.
25. ISO/IEC 15938-3 FCD Information Technology - Multimedia Content Description Interface - Part 3: Visual, March 2001, Singapore.
26. S. Handschuh and S. Staab. Authoring and Annotation of Web Pages in CREAM. In *Proceedings of the 11th International World Wide Web Conference, WWW 2002*, Honolulu, Hawaii, May 7-11 2002.