Knowledge Assisted Analysis & Categorization for Semantic Video Retrieval[1]

Manolis Wallace, Thanos Athanasiadis and Yannis Avrithis

Image, Video and Multimedia Systems Laboratory
School of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou St., 157 73 Zographou, Greece
{wallace,thanos,iavr}@image.ntua.gr

Abstract. In this paper we discuss the use of knowledge for the analysis and semantic retrieval of video. We follow a fuzzy relational approach to knowledge representation, based on which we define and extract the context of either a multimedia document or a user query. During indexing, the context of the document is utilized for the detection of objects and for automatic thematic categorization. During retrieval, the context of the query is used to clarify the exact meaning of the query terms and to meaningfully guide the process of query expansion and index matching. Indexing and retrieval tools have been implemented to demonstrate the proposed techniques and results are presented using video from audiovisual archives.

1 Introduction

The advances in multimedia databases and data networks along with the success of standardization efforts of MPEG-4 [1] and MPEG-7 [2] have driven audiovisual archives towards the conversion of their manually indexed material to digital, network accessible resources, including video, audio and still images. By the end of last decade the question was not on whether digital archives are technically and economically viable, rather on how they would be *efficient* and *informative* [3]. In this framework, different scientific fields, such as database management, image/video analysis, computational intelligence and the semantic web, have observed a close cooperation [4].

Access, indexing and retrieval of image and video content have been dealt either with content-based, or metadata-based techniques. In the former case, image and video content is analyzed, visual descriptors are extracted and content-based indexes are generated, to be used in query by example scenarios [5]. It has been made clear however, that query by visual example is not able to satisfy multiple search usage requirements [6]. In the latter case, various types of metadata, mainly textual, are typically attached to the original data and used to match against user queries. Although this makes textual (e.g., keyword) search possible, to which users are more

1 This work has been partially funded by the EU IST-1999-20502 project FAETHON.

accustomed, the main disadvantage of this approach is the lack of semantic interpretation of the queries that may be posed [7].

The proposed technique achieves semantic handling of archive content using an encyclopedia which contains definitions of abstract semantic classes. The creation of the encyclopedia relies both on human experts and existing ontologies. During document analysis and indexing, semantic entities of the multimedia document descriptions are linked to the abstract ones of the encyclopedia. During retrieval, the supplied keywords of the user query are translated into the semantic entities and the documents whose descriptions have been linked to the requested semantic entities are retrieved. Fuzzy relations are used for knowledge representation, and fuzzy algebraic techniques for query analysis, video document indexing and matching between the two. After providing the general structure of the proposed video retrieval approach, the paper presents the proposed knowledge representation and continues by discussing the application of this knowledge in analysis and retrieval. Indicative results are provided on indexing and retrieval of content from audiovisual archives.

2 Overview

Three main entities participate in the process of video retrieval are: the user, the actual video content and the knowledge that is available to the system, as depicted in Fig. 1. Since each one is expressed in a fundamentally different way, the main effort is to produce techniques that can achieve a uniform representation of all three, so that information provided from each one may be combined and matched. In this work, uniform representation is attempted at a semantic level.

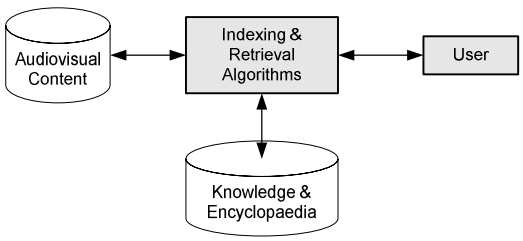

Fig. 1. The proposed framework for knowledge-assisted video indexing and retrieval.

The encyclopedia contains definitions of both simple objects and abstract object classes. Simple objects can be automatically detected in a video document by matching either their visual and audio descriptors with the corresponding ones in the encyclopedia, stored using the structures of Fig 2., or the metadata descriptions with the textual descriptions of the objects. Visual descriptors are extracted using algorithms described in [8], and linked to object/event definitions in the encyclopedia. Abstract classes and concepts cannot be automatically detected in the video documents; they have to be inferred from the simple objects that are identified, thus semantically enriching the indexing of the documents. The user query is issued in a textual form. It is

then automatically mapped to semantic entities found in the encyclopedia, and expanded so as to include entities that are not mentioned but implied. The query in its semantic form is then matched to the semantic indexing of the document, thus providing the system response. This response, being constructed in a semantic way, is much more intuitive than the one provided by existing video retrieval systems. In the following sections we briefly describe the knowledge representation utilized in this work, and explain how it is applied in offline and online tasks.

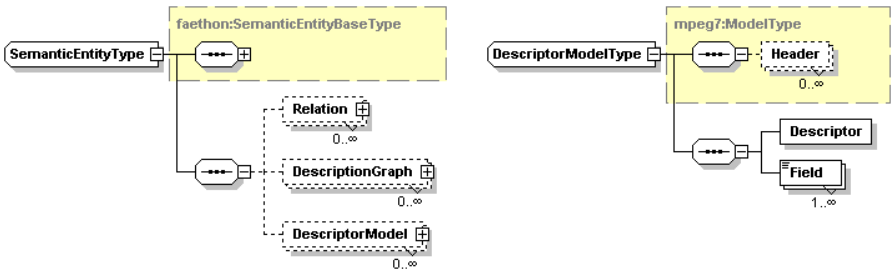

Fig. 2. The structures used to represent the visual and audio characteristics of semantic entities.

3 Knowledge Representation and Context Detection

3.1 Fuzzy relational knowledge representation

Although any type of relation may be contained in an ontology, the two main categories are taxonomic (i.e. ordering) and compatibility (i.e. symmetric) relations. Compatibility relations have traditionally been exploited by information retrieval systems for tasks such as query expansion. They are ideal for the description of similarities, but fail to assist in the determination of context; the use of ordering relations is necessary for such tasks. Thus, a challenge of intelligent information retrieval is the meaningful exploitation of information contained in taxonomic relations of an ontology.

The specialization relation *Sp* is a fuzzy partial ordering on the set of semantic entities. $Sp(a,b) > 0$ means that the meaning of *a* includes the meaning of *b*. The context relation *Ct* is also a fuzzy partial ordering on the set of semantic entities. $Ct(a,b) > 0$ means that *b* provides the context for *a* or, in other words, that *b* is the thematic category that $\underline{a}$ belongs to. Other relations considered in the following have similar interpretations. Their names and corresponding notations are given in Table 1.

Fuzziness of the aforementioned relations has the following meaning: high values of $Sp(a,b)$ imply that the meaning of *b* approaches the meaning of *a*, while as $Sp(a,b)$ decreases, the meaning of *b* becomes narrower than the meaning of *a*. A similar meaning is given to fuzziness of other semantic relations as well. The knowledge contained in these relations is combined into a single quasi-taxonomic relation as follows [9]:

$$T = (Sp \cup Ct^{-1} \cup Ins \cup P \cup Pat \cup Loc \cup Ag)^{n-1} \quad (1)$$

where n is the count of entities in the semantic encyclopedia.

Table 1. The fuzzy semantic relations used for the determination of the context

Symbol	Name	Symbol	Name
Sp	Specialization	*Pat*	Patient
Ct	Context	*Loc*	Location
Ins	Instrument	*Ag*	Agent
P	Part		

3.2 The Notion of Context

In the processes of video content and user query analysis we utilize the common meaning of semantic entities. Let $A = \{s_1, s_2, \ldots, s_n\}$, denote a set of semantic entities, and $S \supseteq A$ be the global set of semantic entities. The common meaning of, and more generally, whatever is common among the elements of A is their context $K(A)$. Assuming that A is a crisp set, i.e. that no fuzzy degrees of membership are contained, the context of the group, which is again a set of semantic entities, can be defined simply as the set of the common descendants of the members of the set.

$$K(A) = \bigcap_{s_i \in A} T(s_i) \tag{2}$$

In the fuzzy case and assuming that fuzzy set A is normal, we extend the above definition as follows, where c is an Archimedean fuzzy complement [10]:

$$K(A) = \bigcap_{s_i \in A} K(s_i) \tag{3}$$

$$K(s_i) = T(s_i) \cup c(A(s_i)) \tag{4}$$

4 Analysis and Indexing

4.1 Fuzzy Hierarchical Clustering of Semantic Entities

The detection of the topics that are related to a document d requires that the set of semantic entities that are related to it are clustered, according to their common meaning. Since the number of topics that exist in a document is not known beforehand, partitioning methods are inapplicable for this task [11] and a hierarchical clustering algorithm needs to be applied [12].

The two key points in hierarchical clustering are the identification of the clusters to merge at each step, i.e. the definition of a meaningful metric $d(c_1,c_2)$ for a pair of clusters c_1, c_2 , and the identification of the optimal terminating step, i.e. the definition

of a meaningful termination criterion. From a semantic point of view, two entities are related to the extent that they refer to the same concepts or belong to the same abstract class. Thus, we utilize as a metric for the comparison of clusters c_1 and c_2 the intensity of their common context:

$$d(c_1, c_2) = h(K(c_1 \cup c_2)) \quad (5)$$

The termination criterion is a threshold on the selected compatibility metric. This clustering method, being a hierarchical one, successfully determines the count of distinct clusters that exist in document *d*, but suffers from a major disadvantage; it only creates crisp partitions, not allowing for overlapping of clusters or fuzzy membership degrees. Using the semantic entities contained in each cluster *c* we design a fuzzy classifier that is able to generate fuzzy partitions as follows:

$$C_c(s) = \frac{h(K(c \cup s))}{h(K(c))} \quad (6)$$

4.2 Thematic Categorization and Detection of Events and Objects

For each cluster of semantic entities detected in document *d*, the context of the cluster describes the topics that are related to the entities of the cluster. This information can be utilized both for thematic categorization of the document and for detection of other objects, through simple inference. Thematic categories are semantic entities that have been selected as having a special meaning. The context of the fuzzy clusters is first calculated, using the inverse of relation *T* as the base quasi-taxonomy. The thematic categories that are found to belong to the context of the cluster are extracted as the thematic categorization of the cluster. Information obtained from clusters with small cardinality is ignored, as possibly erroneous.

$$TC(d) = \mathbf{U}\left[TC(c_i)L(|c_i|)\right] \quad (7)$$

$$TC(c_i) = w(K(c_i) \cap S_{TC}) \quad (8)$$

where *w* is a weak modifier and S_{TC} is the set of thematic categories. On the other hand, the presence of simple events and objects is typically detected to a smaller extent. We infer that a simple semantic entity is present when one of its special cases has been detected:

$$O(d) = \mathbf{U}O(c_i) \quad (9)$$

$$O(c_i) = \left\{ s_j / \min(h(T(s_j) \cap K(c_i)), h(T(s_l), \{s_j\})) \right\}, s_l \in c_i \quad (10)$$

Both results are used to enrich the initial indexing of the document. Thus, if *I*(*d*) is the fuzzy set of entities to which the document is linked, the set is extended as follows:

$$I'(d) = I(d) \cup TC(d) \cup O(d) \quad (11)$$

5 Video Retrieval

5.1 Query Analysis

5.1.1 Context-Sensitive Query Interpretation

Ideally, a user query consists of keywords that correspond to one semantic entity. In some cases though, this is not true and some words can be matched to more than one semantic entity. It is left to the system to the make the decision, based on the context of the query, which semantic entity was implied by the user. However, the detection of the query context cannot be performed before the query interpretation is completed, which in turn needs the result of the query context mining. Therefore both tasks must be done simultaneously. Let the textual query contain the terms t_i, i=1,2,… Let also τ_i be the textual description of semantic entities s_{ij}, j=1,2,...,M_i. Then there exist $N_Q = \prod_i M_i$ distinct combinations of semantic entities that may be used for the representation of the user query. Out of the candidate queries q_k, k = 1,2,...,N_Q, the one that has the most intense context is selected:

$$q = q_i \in \{q_1,...,q_{N_Q}\} : h(q_i) \geq h(q_j) \forall q_j \in \{q_1,...,q_{N_Q}\} \quad (12)$$

5.1.2 Context-Sensitive Query Expansion

Query expansion enriches the query in order to increase the probability of a match between the query and the document index. The presence of several semantic entities in the query created during the query interpretation defines a context, which is used to direct the expansion process.

More formally, we replace each semantic entity s_i with a set of semantic entities $X(s_i)$; we will refer to this set as the expanded semantic entity. In a context-sensitive query expansion, the degree of significance, x_{ij}, of the entity s_j in the expanded semantic entity $X(s_j)$ is not only proportional to the weight w_i, but depends on the degree of the relation $T(s_i,s_j)$ as well. We define the measure of relevance as:

$$h_j = \max\left(\frac{h(T(s_j) \cap K(q))}{h(K(q))}, c(h(K(q)))\right) \quad (13)$$

The fuzzy complement c in this relation is Yager's complement with a parameter of 0.5. Considering now the initial entity's importance in the query and the degree to which the initial and the candidate entity are related, we have

$$x_{ij} = h_j q(s_i) T(s_i, s_j) \quad (14)$$

5.2 Index Matching

When querying, we treat $X(s_i)$ considering a union operation, i.e. documents that match any entity contained in $X(s_i)$ are selected. If I is the semantic index in the form of a fuzzy relation from semantic entities to documents, then the set of documents R_i that match extended entity $X(s_i)$ is calculated as

$$R_i = X(s_i) \circ I \quad (15)$$

On the other hand, results from distinct entities are typically treated using an intersection operator, i.e. only documents that match all of the entities of the query are selected. This is quite limiting; it is more intuitive to return the documents that match all of the terms in the query first, followed by those documents that match less of the query terms. We achieve such an intuitive results utilizing an ordered weighted average operator in order to produce the overall result. Thus, the overall result is calculated as

$$R = OWA(R_1, R_2, ...) \quad (16)$$

6 Results

The methodologies presented herein have been developed and applied to a set of 175 documents from the audiovisual archive of the Hellenic Broadcasting Corporation (ERT) i.e. videos of total duration of 20 hours. The a/v documents were already extensively annotated according to the MPEG-7 standard. In this section we provide some indicative results of the application on both the indexing and retrieval processes.

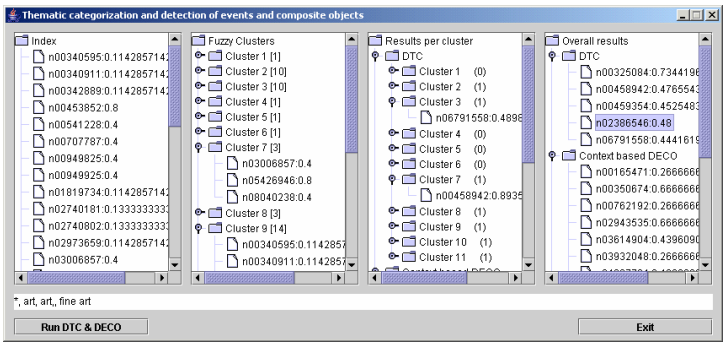

Fig. 3. Results of multimedia document analysis, object detection and thematic categorization.

In Fig. 3 we present an implementation of the document analysis methodologies described in section 4. In the first column, the IDs of the objects detected in the video are presented. In the semantic encyclopaedia each one of these IDs is related to a textual description, a set of keywords, and possibly a set of audiovisual descriptors. For example, ID n02386546 is related to keywords “art” and “fine art”, as can be seen at the lower part of the application. In the second column the entities have been clustered applying the fuzzy hierarchical clustering algorithm. In column 3, thematic categorization information is extracted from each cluster and some simple objects are detected. Finally, in column 4 results are summarized for all documents. Not all thematic categories detected in distinct clusters participate in the overall result; findings that correspond to clusters of small cardinality have been ignored, as they are possibly misleading. On the contrary, all information for detected objects is kept.

SEMANTIC AND METADATA SEARCH - SemanticResponse

The expanded set of semantic entities has been matched with the following multimedia documents in the Faethon semantic index, with degree of relevance:

Id	Title	SourceArchive	Score
35	Flugzeugkatastrophe	**FAA**	0.9
1199	Επικαιρότητες Αυγούστου 1974	**ERT**	0.8
52	Die Vietnamkrise	**FAA**	0.78
1	Sensationell neuen Rettungsmethode	**FAA**	0.72
1514	Περιοδικό πιο	**ERT**	0.68
AVQ-A-004129-0038	Archeological excavations in Rome	**Alinari**	0.6
11	Ausbau und Elektrifizierung der Strecke Graz-Bruck	**FAA**	0.5
FCC-F-021960-0000	Exodus of the Belgian population	**Alinari**	0.45

Pages: << Previous | QueryInterpretation | QueryExpansion | SemanticResponse | PresentationResponse | ClassificationResponse | Next >>

Fig. 4. Ranked multimedia documents retrieved for a user query with the keyword “politics”.

Fig. 4 demonstrates the results of a user query corresponding to the keyword “politics”. Thematic categorization and detection of events and objects has been performed beforehand. The search process starts with query interpretation and expansion. Index matching is then performed and the list of the retrieved documents is ranked according to their degree of relevance to the semantic entity “politics”. As the data set was used for the adjustment of the knowledge (in example for the thematic categories extraction), the results during the evaluation process were remarkably good. As future work we intend to extend the data set and conduct further evaluations in a more generalized set of documents.

7 Conclusions

In this paper we have discussed the utilization of semantic knowledge for the analysis and retrieval of video. Specifically, we have followed a fuzzy relational approach to knowledge representation, based on which we have defined and extracted the context, i.e. the common overall meaning, of a set of semantic entities. The notion of context has then been applied in the understanding of both the user and the video content.

As far as the analysis of video is concerned, the context of simple objects detected in its content has been used to detect the topics to which the video is related. From those topics we perform thematic categorization and detect simple objects that were not identified in the video but their presence can be inferred. As far as the user is concerned, the context of the user query is used to clarify the exact meaning of the query terms and to meaningfully guide the process of query expansion. Results have been provided that are indicative of the proposed approach.

References

1. Koenen R.: "Overview of the MPEG-4 Standard", ISO/IEC JTC 1/SC 29/WG 11/N4668, March 2002
2. T. Sikora. The MPEG-7 Visual standard for content description - an overview. IEEE Trans. on Circuits and Systems for Video Technology, 11(6):696–702, June 2001
3. Y. Avrithis, G. Stamou, A. Delopoulos and S. Kollias: Intelligent Semantic Access to Audiovisual Content. In Proc. of 2nd Hellenic Conference on Artificial Intelligence (SETN'02), Thessaloniki, Greece, April 11-12, 2002
4. 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
5. A. Smeulders, M. Worring, and S. Santini: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 22(12), 2000
6. M. La Cascia, S. Sethi, and S. Sclaroff: Combining textual and visual cues for content-based image retrieval on the world wide web. IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'98), June 1998
7. G. Stamou, Y. Avrithis, S. Kollias, F. Marques and P. Salembier: Semantic Unification of Heterogenous Multimedia Archives. In Proc. of 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'03), London, UK, April 9-11, 2003
8. P. Tzouveli, G. Andreou, G. Tsechpenakis, Y. Avrithis and S. Kollias, "Intelligent Visual Descriptor Extraction from Video Sequences," in Proc. of 1st International Workshop on Adaptive Multimedia Retrieval (AMR '03), Hamburg, Germany, September 15-18, 2003
9. Wallace, M., Akrivas, G. and Stamou, G.: Automatic Thematic Categorization of Documents Using a Fuzzy Taxonomy and Fuzzy Hierarchical Clustering. In Proc. of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), St. Louis, MO, USA, May 2003
10. Klir G. and Bo Yuan, Fuzzy Sets and Fuzzy Logic, Theory and Applications, New Jersey, Prentice Hall, 1995
11. Miyamoto S.: Fuzzy sets in information retrieval and cluster analysis. Kluwer Academic publishers, 1990
12. S. Theodoridis and K. Koutroumbas: Pattern Recognition. Academic Press, 1998