Multimed Tools Appl (2008) 39:293–327
DOI 10.1007/s11042-007-0161-4

Semantic representation of multimedia content: Knowledge representation and semantic indexing

Phivos Mylonas · Thanos Athanasiadis · Manolis Wallace · Yannis Avrithis · Stefanos Kollias

Published online: 4 September 2007

Abstract In this paper we present a framework for unified, personalized access to heterogeneous multimedia content in distributed repositories. Focusing on semantic analysis of multimedia documents, metadata, user queries and user profiles, it contributes to the bridging of the gap between the semantic nature of user queries and raw multimedia documents. The proposed approach utilizes as input visual content analysis results, as well as analyzes and exploits associated textual annotation, in order to extract the underlying semantics, construct a semantic index and classify documents to topics, based on a unified knowledge and semantics representation model. It may then accept user queries, and, carrying out semantic interpretation and expansion, retrieve documents from the index and rank them according to user preferences, similarly to text retrieval. All processes are based on a novel semantic processing methodology, employing fuzzy algebra and principles of taxonomic knowledge representation. The first part of this work presented in this paper deals with data and knowledge models, manipulation of multimedia content annotations and semantic indexing, while the second part will continue on the use of the extracted semantic information for personalized retrieval.

P. Mylonas (✉) · T. Athanasiadis · Y. Avrithis · S. Kollias
School of Electrical and Computer Engineering, National Technical University of Athens,
9 Iroon Polytechniou Str., 157 73 Zographou Campus, Athens, Greece
e-mail: fmylonas@image.ntua.gr

T. Athanasiadis
e-mail: thanos@image.ntua.gr

Y. Avrithis
e-mail: iavr@image.ntua.gr

S. Kollias
e-mail: stefanos@cs.ntua.gr

M. Wallace
Department of Computer Science, University of Indianapolis, Athens Campus, 9 Ipitou Str., 105 57,
Athens, Greece
e-mail: wallace@uindy.gr

Keywords Multimedia content semantics extraction · Semantic indexing · Semantic classification

1 Introduction

Over the last decade, multimedia content indexing and retrieval has been influenced by the important progress in numerous fields, such as digital content production, archiving, multimedia signal processing and analysis, computer vision, artificial intelligence and information retrieval [9]. One major obstacle, though, multimedia retrieval systems still need to overcome in order to gain widespread acceptance, is the *semantic gap* [50, 70, 92]; the latter forms an existing problem and in this approach we provide a partial contribution towards its solution. This refers to the extraction of the semantic content of multimedia documents, the interpretation of user information needs and requests, as well as to the matching between the two. This hindrance becomes even harder when attempting to access vast amounts of multimedia information encoded, represented and described in different formats and levels of detail.

Although this gap has been acknowledged for a long time, multimedia analysis approaches are still divided into two main categories; the low-level multimedia analysis methods and tools on the one hand (e.g. [51, 58, 59, 62]) and the high-level semantic annotation methods and tools on the other hand (e.g. [11, 39, 80, 83]). It was only recently, that state-of-the-art multimedia analysis systems have started using semantic knowledge technologies, as the latter are defined by notions like the Semantic Web [17, 88] and ontologies [34, 76]. The advantages of using Semantic Web technologies for the creation, manipulation and post-processing of multimedia metadata is depicted in numerous activities [77], trying to provide "semantics to semantics".

Digital video is the most demanding and complex data structure, due to its large amounts of spatiotemporal interrelations; video understanding and indexing is a key step towards more efficient manipulation of visual media, presuming semantic information extraction. As it is extensively shown in the literature [29, 45, 79], it is true that multimedia standards, such as MPEG-7 [67] and MPEG-21 [24, 53], seek to consolidate and render effectively the infrastructure for the delivery and management of multimedia content and do provide important functionalities when dealing with aspects like the description of objects and associated metadata [71]. For instance, the Multimedia Description Scheme tools [14] specified by the MPEG-7 standard for describing multimedia content, include, among others, tools that represent the structure and semantics of multimedia data [10, 12, 15]. However, the important process of extraction of semantic descriptions from the content with the corresponding metadata, lies out of the scope of this standard, motivating heavy research efforts in the direction of automatic annotation of multimedia content [6, 13, 22, 93].

The need for machine-understandable representation and manipulation of the semantics associated with the MPEG-7 Descriptor Schemes and Descriptors, led to the development of ontologies for specific parts of MPEG-7 [33, 43, 44, 81]. In the approach proposed by Hunter [44], trials and tribulations of building such an ontology are presented, as well as its exploitation and reusability by other communities on the Semantic Web. In [81] on the other hand, the semantic part of MPEG-7 is translated into an ontology that serves as the core one for the attachment of domain specific ontologies, in order to achieve MPEG-7 compliant domain specific annotations, hence the initial conceptualization of the domain

specific ontologies needs to be "mapped" to the MPEG-7 modeling rationale. Furthermore, the most detailed approach towards the automatic mapping of the entire MPEG-7 standard to OWL [90] is presented in [33], based on a generic XML Schema [91] to OWL mapping. Finally, MPEG-7's visual part has been modeled in an RDFS-based ontology in [69], whose goal is to enable machines to generate and understand visual descriptions which can be used later for multimedia reasoning.

A number of research directions in the area of multimedia content management are briefly presented here and further analyzed in Section 2, each one with its own challenges. Starting with a bottom–up approach, detection of low-level multimedia objects to concepts forms a promising, standalone research area, which, however, does not usually take into account the semantics of the content or its context. Knowledge modeling for multimedia understanding (e.g. in terms of mediators) and semantic multimedia indexing and retrieval constitute two other very interesting research fields. Research efforts presented in both of them are very close to the herein proposed approach, but they seem to lack on scalability (e.g. in terms of the number of supported concepts) and support of uncertainty or contextual information. On top of that, document classification and topic identification are providing high-level access to multimedia content. However, most classification techniques do not follow a semantic interpretation and are statistical in principle, whereas the latest achievements in topic identification do not utilize any kind of knowledge or context.

In this work, our effort focuses on an integrated framework, offering transparent, unified and unsupervised access to heterogeneous multimedia content in distributed repositories. It acts complementary to the current state-of-the-art as it tackles most of the aforementioned challenges. Focusing on semantic analysis of multimedia documents, metadata, user queries and user profiles, it contributes towards bridging the gap between the semantic nature of user queries and raw multimedia documents, serving as a mediator between users and remote multimedia repositories. Its core contribution relies on the fact that it provides a unified way to represent high-level video semantics, deriving both from the multimedia content and the associated textual annotations. It utilizes video content analysis results, analyzes and extracts the underlying semantics from the text and thus satisfies the purely semantic needs of users.

More specifically, high-level concepts and relations between them are utilized to represent the detected objects and events from the low-level processing output. The main idea introduced here relies on the integrated handling of concepts in multimedia content. The accompanying metadata is then thoroughly exploited towards semantic interpretation of the multimedia content. Both the metadata and the content are mapped to a semantic index, which is constructed based on a unified, fuzzy knowledge model. This index is constantly updated and is used to connect concepts to multimedia documents. We focus mainly on the semantic handling of metadata, in terms of establishing a link between the content's textual annotation and the concepts. The latter, together with the mapping of low-level analysis results to concepts, form the semantic index. Finally, multimedia documents are classified to topics (e.g. sports, politics, etc.) through fuzzy clustering of the index at the concept level. User queries are then analyzed and processed to retrieve multimedia documents from the index, by carrying out semantic interpretation and expansion. In order to demonstrate the efficiency of the proposed framework we have developed two test scenarios: (i) a minimal step-by-step scenario involving five synthetic multimedia documents and (ii) a twofold scenario, introducing aggregated classification results over a real-life repository of multimedia documents, containing both multimedia programs and news items classified in 13 ground truth topics. Their classification results are also compared to an implementation of the pLSA algorithm [40] on the same data set.

The structure of this paper is as follows: Section 2 presents the current state-of-the-art in the field of multimedia document analysis and Section 3 provides an overview of our proposed framework, focusing on its structure and data models. Section 4 describes the underlying knowledge representation, including an innovative definition of taxonomic context. Continuing, Section 5 discusses semantic annotation procedure (based on concept detection from the content and on semantic interpretation of the accompanying metadata) to construct the semantic index, followed by semantic classification of multimedia documents through fuzzy clustering of the index. Section 6 presents semantic classification results and evaluation and in Section 7 partial conclusions are drawn, since this paper presents the first part of our complete work. In the second part of this work, "Semantic Representation of Multimedia Content: Retrieval and Personalization", we shall deal with semantic retrieval including user query interpretation and expansion, as well as with document ranking and user preference extraction.

2 Related work

In the context of both modeling knowledge for multimedia understanding and assigning low-level multimedia objects to concepts, Petridis et al. [61] describe a knowledge representation suitable for semantic annotation of multimedia content, whereas Bertini et al. [19, 20] propose the use of pictorially enriched ontologies within a specific domain, that include visual concepts together with linguistic keywords. Simou et al. propose an ontology infrastructure suitable for multimedia reasoning in [68], whereas in [41] Hollink et al. attempt to specify the necessary requirements a visual ontology for video annotation must fulfill and propose the use of a WordNet/MPEG-7 ontology combination towards that scope. Athanasiadis et al. [8] focus on the use of a multimedia ontology infrastructure for analysis and semantic annotation of multimedia content. Hoogs et al. [42] couple a classical image analysis objects and events recognition approach with WordNet's semantics, taking advantage of its hierarchical relationships structure, focusing on the information produced by the visual analysis task and resulting in an automated annotation of multimedia content. Contextual information is used only from the transcribed commentary, which improves the annotation accuracy, but is still insufficient and constrains the semantic search. Hauptmann [35] proposed the design of an automatically detectable concept ontology that could be utilized for annotation of broadcast video, but still it is not clear which concepts are suitable for inclusion in such an ontology. Of course, description of multimedia documents amounts to consider both structure and conceptual aspects (i.e. the content), as depicted in [79], whereas active W3C efforts, like [89], provide continuously research results towards standardization in the multimedia annotation on the Semantic Web field.

All of the above initiatives have produced interesting results towards satisfying the need for a semantic multimedia metadata framework that will facilitate multimedia applications development. Nevertheless, none of the above initiatives results in a unified treatment of the multimedia semantics integration process, which remains an open research problem; thus, we feel they are all complementary to the scope of the work presented in this paper. It is becoming apparent that integration of diverse, heterogeneous and distributed pre-existing-multimedia content will only be feasible if the current research activities are active in the direction of knowledge acquisition and modeling, capturing knowledge from raw information and multimedia content in distributed repositories to turn poorly structured information into machine-processable knowledge [50, 56]. A multimedia mediator system is designed in [21] to provide a well-structured and controlled gateway to multimedia

systems, focusing on schemas for semi-structured multimedia items and object-oriented concepts, while [4] focuses on security requirements of such mediated information systems. Altenschmidt et al. [3] enforces correct interpretations of queries by imposing constraints on the mappings between the target schema and the source schemas. On the other hand, [82] supports interaction schemes such as query by example, answer enlargement/reduction, query relaxation/restriction and adaptation facilities.

A lot of research efforts have also been spent in the field of semantic multimedia indexing and retrieval [5, 38, 55, 75]. One of the first integrated attempts was the Informedia project and its offsprings [36, 37], which combined speech, image, natural language understanding and image processing to automatically index video for intelligent search and retrieval. Papadopoulos et al. [60] propose a knowledge-assisted multimedia analysis technique based on context and spatial optimization. MARVEL multimedia analysis engine [72], on the other hand, utilizes multimodal machine learning techniques to organize semantic concepts using ontologies, exploiting semantic relationships in the process, thus being very close to the herein proposed architecture. Multimedia indexing is achieved through fusion of low-level visual feature descriptors, semantic concept models and clear text metadata. Snoek et al. [73, 74] propose in MediaMill a semantic pathfinder architecture for generic indexing of video sequences, obtaining promising results in the high-level feature extraction task of NIST TRECVID 2006 benchmark [57]. In particular, they extract semantic concepts from news video based on the exploration of different paths through three consecutive analysis steps: content analysis, style analysis, and context analysis focusing on a per concept basis. A major difference of our approach in comparison to [73] is the fact that the core of the latter system is built by an unprecedented lexicon of a limited amount of semantic concepts and a query-by-concept principle is followed. Finally, Dorai et al. [30] claim that interpretation of multimedia content must be performed from the aspect of its maker; automatic understanding and indexing of video is possible, based on the intended meaning and perceived impact on content consumers of a variety of visual and aural elements present.

On the other hand, research efforts dealing with document classification have matured and provided interesting results over the last decade. An exhaustive review of a detailed variety of categorization models may be found in [66]. New statistical models for classification of structured multimedia documents are presented, as the one in [28]. In the same context, of great interest is the field of unsupervised document clustering, where textual documents are clustered into groups of similar content according to a predefined similarity criterion, which in most cases is depicted by the widely accepted cosine coefficient of the vector space model. Complete linkage, single linkage or even group average hierarchical clustering algorithms are primarily used in document clustering [78]. The most computationally efficient method is the single link one and therefore has been extensively used in the literature; however, the complete link method is considered to be more effective, although it demands a higher computational cost [87]. MacLeod proposed the utilization of neural models for clustering in [49], as an alternative to document vector similarity models.

In terms of topic identification, fast algorithms have been introduced and utilized for browsing of large amounts of multimedia content [26]. Latent Semantic Analysis (LSA) [18, 27, 48] uses Singular Value Decomposition (SVD) to map documents and terms from their standard vector space representation to a lower dimensional latent space, thus revealing semantic relations between the entities of interest. An unsupervised generalization of LSA, probabilistic-LSA (pLSA) [40], which builds upon a statistical foundation, represents documents in a semantic concept space and extracts concepts automatically.

Furthermore, pattern recognition and machine learning techniques have also been applied to document classification, such as the fuzzy c-means algorithm [16] in the case of supervised multimedia documents classification. Finally, projection techniques [65] and k-means clustering [63] are proposed to speed up the distance calculations of clustering and their effectiveness is examined in [23]; however, document clustering results are very dependent on the original multimedia content dataset.

3 Overview of the proposed framework

3.1 Mediator structure

The proposed framework is depicted in Fig. 1. A single *user interface* provides a unified access to each individual repository, while the *multimedia repository interfaces* are responsible for the communication between the central unit and each multimedia repository.

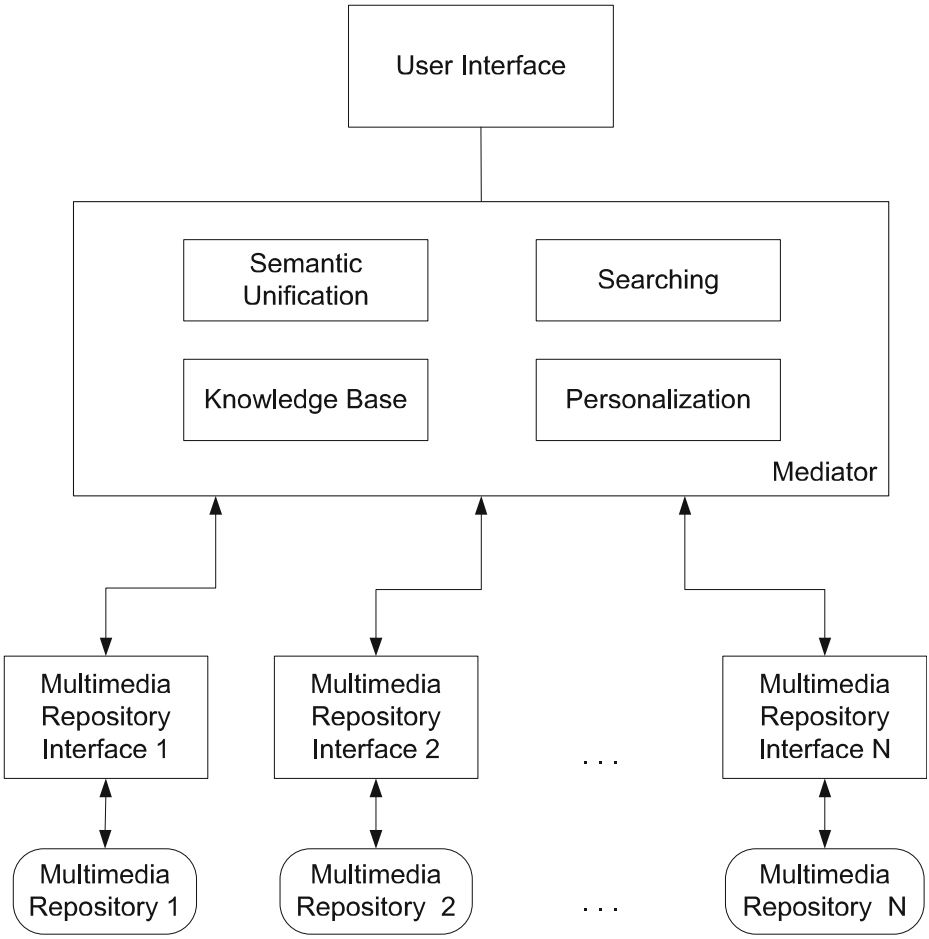

Fig. 1 Structure of the proposed framework

The central mediator consists of four parts: *knowledge base, semantic unification, searching* and *personalization*. The knowledge base consists of the knowledge model, the semantic index and the user profiles. The semantic unification unit deals with generating and updating the semantic index, while the personalization unit handles updating of user profiles. The searching unit analyzes the user query, carries out matching with the index and returns the retrieved documents to the user. The description of each unit's functionality follows the distinction in two main operation modes. In *query mode*, user requests are processed and the respective responses are assembled and presented. In *update mode*, the semantic index and user profiles are updated in an offline process to reflect content updates and usage, respectively.

3.2 Data models

To achieve unified access to heterogeneous multimedia documents, a uniform data model is of crucial importance. The proposed data model consists of three elements: the knowledge model, the semantic index and the user profiles, discussed as follows.

The *knowledge model* contains all semantic information used. It supports structured storage of concepts and relations that experts have defined manually for a limited set of specific multimedia document categories. Three information types are introduced in the model, namely *concepts* (e.g. objects, events, topics, agents, semantic places and times), *relations* linking concepts (e.g. "part of" or "specialization of") and a *quasi-taxonomic relation*, i.e. a taxonomic knowledge representation to interpret the meaning of a multimedia document, composed of several elementary relations, also referred to as *taxonomy.*

The *semantic index* contains sets of document locators for each concept, identifying which concepts have been associated to each available multimedia document and supporting unified access to the repositories. Document locators associated to index concepts may link to complete multimedia documents, objects or other video decomposition units.

User profiles contain all user information required for personalization, decomposed into usage history and semantic user preferences. The latter may be divided in two (possibly overlapping) categories, namely *preferences for topics* and *interests* (i.e., preferences for all other concepts). Semantic preferences are mined through the analysis of usage history records and are accompanied by weights indicating the intensity of each preference. The description of "negative" intensity is also supported, to model the user's dislikes.

4 Knowledge representation

The proposed knowledge model is based on a set of concepts and relations between them, which form the basic elements towards semantic interpretation. This knowledge representation allows the establishment of a detailed content description of all different multimedia documents in a unified way and is specifically constructed to match the specific textual annotations of the multimedia documents at hand. Due to the fact that relations among real-life concepts are always a matter of degree, and are, therefore, best modeled using fuzzy relations, the approach followed herein is based on a formal methodology and mathematical notation founded on fuzzy relational algebra [47]. Its basic principles are summarized in the following.

4.1 Mathematical notation

Given a universe U, a crisp set S on U is described by a membership function μ_S: $U \rightarrow \{0,1\}$, whereas a *fuzzy set* F on S is described by a membership function μ_F:$S \rightarrow [0,1]$. We may describe the fuzzy set F using the sum notation [52]: $F = \sum_i s_i/w_i = \{s_1/w_1, s_2/w_2, \ldots, s_n/w_n\}$, where $i \in N_n$, $n = |S|$ is the cardinality of S, $w_i = \mu_F(s_i)$ or, more simply, $w_i = F(s_i)$ and $s_i \in S$. The *height* of the fuzzy set F is defined as $h(F) = \max_i(F(s_i)), i \in N_n$. A *normal* fuzzy set is defined as a fuzzy set having a height equal to 1, whereas cp is an involutive fuzzy complement, i.e. $cp(cp(a)) = a$, for each $a \in [0,1]$ [47]. The product of a fuzzy set F with a number $y \in [0,1]$ is defined as $[F \cdot y]_{(x)} = F(x) \cdot y,\ \forall x \in S,\ y \in [0,1]$. The *support* of a fuzzy set F within U is the crisp set that contains all the elements of U that have non-zero membership grades in F.

A *fuzzy binary relation* on S is a function R:$S^2 \rightarrow [0,1]$. Its *inverse* relation is defined as $R^{-1}(x,y) = R(y,x)$. The *intersection*, *union* and sup-t *composition* of two fuzzy relations R_1 and R_2 defined on the same set S, are defined as:

$$(R_1 \cap R_2)(x,y) = t(R_1(x,y), R_2(x,y)) \tag{1}$$

$$(R_1 \cup R_2)(x,y) = u(R_1(x,y), R_2(x,y)) \tag{2}$$

$$(R_1 \circ R_2)(x,y) = \sup_{z \in S} t(R_1(x,z), R_2(z,y)) \tag{3}$$

respectively, where t and u are a t-norm and a t co-norm, respectively. The *standard* t-norm and t-conorm are the *min* and *max* functions, respectively. At this point one may think this is a similar approach to the ones presented in [31] and [32], where aggregation functions are utilized to deal with fuzziness in multimedia data; however, our work differs significantly in the fact that the aforementioned works do not deal with the semantic level of multimedia content and focus on performing efficient similarity search and classification in high dimensional data. An *Archimedean* t-norm[1] also satisfies the properties of continuity and subidempotency, i.e. $t(a,a) < a$, $\forall a \in (0,1)$. The identity relation R_I is the identity element of the sup-t composition: $R \circ R_I = R_I \circ R = R, \forall R$. The properties of *reflexivity*, *symmetry* and sup-t *transitivity* are defined as follows: R is called reflexive iff $R_I \subseteq R$; symmetric iff $R = R^{-1}$; antisymmetric iff $R \cap R^{-1} \subseteq R_I$; and sup-$t$ transitive (or simply transitive) iff $R \circ R \subseteq R$. In the above definitions, operations between relations are defined as in the case of fuzzy sets. For example, $\subseteq$ between two relations A and B is defined as:

$$A \subseteq B \Leftrightarrow A(x,y) \leq B(x,y)\ \forall x,y \tag{4}$$

A *transitive closure* of a relation is the smallest transitive relation that contains the original relation. In general, the closure of a relation is the smallest extension of the relation that has a certain specific property, such as reflexivity, symmetry or transitivity, whereas the sup-t transitive closure $Tr^t(R)$ of a relation R is formally given by $Tr^t(R) = \bigcup_{i=1}^{\infty} R^{(i)}$, where $R^{(i)} = R \circ R^{(i-1)}$ and $R^{(1)} = R$. It is proved that if R is reflexive, then its transitive closure is

[1] A t-norm is called Archimedean, if 0 and 1 are its only idempotents. An idempotent is something that - given a binary operation like the one presented herein - when multiplied by (or for a function, composed with) itself, gives itself as a result.

given by $Tr^t(R) = R^{(n-1)}$, where $n = |S|$ [47]. A fuzzy *ordering* relation is a fuzzy binary relation that is antisymmetric and transitive. A *partial ordering* is, additionally, reflexive. A fuzzy partial ordering relation R defines, for each element $s \in S$, the fuzzy set of its *ancestors* (dominating class) and its *descendants* (dominated class). We will use the notation $R(s)$ for the dominated class of s.

4.2 Fuzzy taxonomic relations

Retrieval systems based on lexical terms typically suffer from the problematic mapping of terms to concepts [64]. As more than one term may be associated to the same concept, and more than one concept may be associated to the same term, the processing of query and index information is not trivial. In order to overcome such problems, one should work directly with concepts, rather than terms. In the sequel, we shall denote by $S = (s_1, s_2, \ldots, s_n)$ the set of concepts that are known. A knowledge representation model may consist of the definitions of these concepts, together with their lexical descriptions, i.e. their corresponding terms, as well as a set of relations amongst the concepts. The objective is to construct a model in which the context determines the intended meaning of each term, and a term used in different context may have different meanings. An initial formal definition of such a model may be given as:

$$O = \{S, \{R_i\}\}, \quad i = 1 \ldots n \tag{5}$$

$$R_i : S \times S \to \{0,\ 1\}, \quad i = 1 \ldots n \tag{6}$$

where O is the knowledge model and R_i the i-th relation amongst the concepts. Although almost any type of relation may be included to construct the knowledge representation, the two main categories used are *taxonomic* (i.e. *ordering*) and *compatibility* (i.e. *symmetric*) relations. As proven in [1], compatibility relations fail to assist in the determination of the context of a query or a document; the use of ordering relations is necessary for such tasks. Thus, a main challenge is the meaningful exploitation of information contained in taxonomic relations.

In addition, for a knowledge model to be highly descriptive, it must contain a large number of distinct and diverse relations among concepts. Available information will then be scattered among them, making each one of them inadequate to describe a context in a meaningful way. The relations need to be combined to provide a view of the knowledge that suffices for context definition and estimation. Fuzzy relations have been proposed to handle such issues when modeling real-life information [2]. In particular, several commonly encountered relations, that can be modeled as *fuzzy ordering relations*, can be combined for the generation of a meaningful, fuzzy, quasi-taxonomic relation. A new knowledge model $O_{\mathcal{F}}$ is constructed in this case:

$$O_{\mathcal{F}} = \{S, \{r_i\}\}, \quad i = 1 \ldots n \tag{7}$$

$$r_i = \mathcal{F}(R_i) : S \times S \to [0,\ 1], \quad i = 1 \ldots n \tag{8}$$

where $\mathcal{F}$ denotes the fuzzification of the R_i relations. Based on the relations r_i we construct the following combined relation:

$$T = Tr^t\left(\underset{i}{\cup} r_i^{p_i}\right), \quad p_i \in \{-1,\ 1\}, \quad i = 1 \ldots n \tag{9}$$

where the value of p_i is determined by the semantics of each relation used in the construction of T (e.g. order of arguments a, b in Table 1), since some relations may need to be inversed before being used in the construction of T. The transitive closure in (9) is required in order for T to be taxonomic, as the union of transitive relations is not necessarily transitive. For the purpose of analyzing multimedia document descriptions, relation T has been generated with the use of a set of fuzzy taxonomic relations, whose semantics are defined in MPEG-7 [46] and summarized in Table 1.

In this case, T becomes [84]:

$$T = Tr^t\left(Sp \cup P^{-1} \cup Ins \cup \mathrm{Pr}^{-1} \cup Pat \cup L \cup Ex\right) \tag{10}$$

Based on the semantics of the participating relations, it is easy to see that T is ideal for the determination of the topics that a concept may be related to, as well as for the estimation of the common meaning, i.e. the context, of a set of concepts. All relations used for the generation of T are partial ordering relations. Still, there is no evidence that their union is also antisymmetric, a property also required for it to be taxonomic. Quite the contrary, T may vary from being a partial ordering to being an equivalence relation. This is important as true semantic relations also fit in this range-total symmetry as well as total antisymmetry often have to be abandoned when modeling real life. Still, the semantics of the used relations, as well as our experiments, indicate that T is very close to antisymmetric. Therefore, we classify it as quasi-ordering or *quasi-taxonomic*.

Another benefit of this approach is that conceptual taxonomies and relations in the knowledge model are modeled as weighted parent–child pairs and can be represented as square matrices of dimension equal to the size of known concepts. Although this representation alone does not provide for optimal exploitation of storage and computing resources, we have implemented a compact sparse representation model for the taxonomies and designed an incremental transitive closure algorithm (ITC) that terminates extremely faster than the best known approach to transitive closure of weighted binary relations [86]. The algorithm terminates in below-linear time, making it possible to edit the taxonomies in a trial-and-error methodology, which greatly facilitates the process of knowledge refinement by a human expert.

Table 1 Fuzzy taxonomic relations used for generation of T

Name	Symbol	Meaning	Example	
			a	*b*
Part	$P(a, b)$	b is a part of a	Human body	Hand
Specialization	$Sp(a, b)$	b is a specialization of a	Vehicle	Car
Example	$Ex(a, b)$	b is an example of a	Player	Jordan
Instrument	$Ins(a, b)$	b is an instrument of a	Music	Drums
Location	$L(a, b)$	b is the location of a	Concert	Stage
Patient	$Pat(a, b)$	b is a patient of a	Course	Student
Property	$Pr(a, b)$	b is a property of a	Jordan	Star

4.3 Taxonomic context model

When using a taxonomic knowledge representation to interpret the meaning of a multimedia document, it is the context of a term that provides its truly intended meaning. In other words, the true source of information is the co-occurrence of certain concepts and not each one independently. Thus, the common meaning of terms should be used in order to best determine the concepts to which they should be mapped. In the following we shall refer to this as their *context*, keeping in mind that it constitutes only one possible expression for the notion of context [54].

The fact that relation T is (almost) an ordering relation allows us to use it in order to define, extract and use the context of a set of concepts. Relying on the semantics of relation T, we define the context $K(s)$ of a single concept $s \in S$ as the set of its antecedents in relation T.

More formally, $K(s) = T_{\leq}(s)$, following the standard superset/subset notation from fuzzy relational algebra. Assuming that a set of concepts A is crisp, i.e. that all considered concepts belong to the set with degree equal to 1, the context of the set, which is again a set of concepts, can be defined simply as the set of their common antecedents, as formally depicted in (11) and represented in Fig. 2 for concepts *ball*, *referee* and *basket*:

$$K(A) = \cap K(s_i), \quad s_i \in A \tag{11}$$

Obviously, as more concepts are considered, the context becomes narrower, i.e. it contains less concepts and to smaller degrees. Letting $A,B \subseteq S$ and $A \supset B \rightarrow K(A) \subseteq K(B)$ and considering Fig. 3, this would imply that given $A = \{s_3, s_4, s_5\}$ and $B = \{s_3, s_4\}$, then $K(A) = \{s_1\}$ and $K(B) = \{s_1, s_2\}$.

When the definition of context is extended to the case of fuzzy sets of concepts, this property must still hold. Moreover, we demand that the following properties hold as well, because of the nature of fuzzy sets:

- $A(s) = 0 \Rightarrow K(A) = K(A - \{s\})$, i.e., no context narrowing
- $A(s) = 1 \Rightarrow K(A) \subseteq K(s)$, i.e., full narrowing of context
- $K(A)$ decreases monotonically with respect to $A(s)$

Let $\mathcal{K}(s)$ be the "considered" context of s, i.e. the concept's context when taking its degree of participation to the set into account. According to the above properties, when A is a normal fuzzy set, $\mathcal{K}(s)$ should be low when the degrees of taxonomy are low and the degree of participation $A(s)$ is high or in other words when the context of the crisp entity $K(s)$ is low. This is modeled as $cp(\mathcal{K}(s)) \doteq cp(K(s)) \cap (S \cdot A(s))$, where $S \cdot A(s)$ is the product of set S with membership degree $A(s)$, as defined in subsection 4.1, and the $\doteq$ sign designates an equality that comes from definition. By applying de Morgan's law, we obtain:

$$\mathcal{K}(s) \doteq K(s) \cup cp(S \cdot A(s)) \tag{12}$$

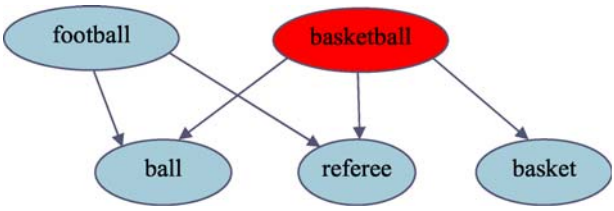

Fig. 2 Concept *basketball* is the only common antecedent of all three concepts *ball*, *referee* and *basket* in relation T, i.e. *basketball* is the context of *ball*, *referee* and *basket*

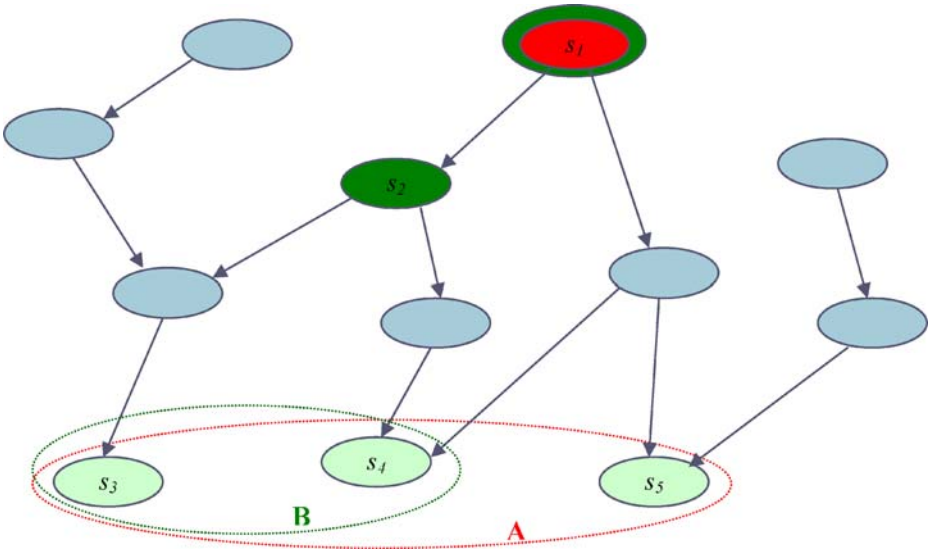

Fig. 3 As more concepts are considered, the context contains less concepts and to lower degrees: considering only the first two leaves from the left (s_3, s_4), the context contains two concepts (s_1, s_2), whereas considering all the leaves (s_3, s_4, s_5) narrows the context to just one common ascendant (s_1)

Then the set's context is easily calculated as:

$$K(A) = \cap_i \mathcal{K}(s_i),\ s_i \in A \tag{13}$$

To further enlighten this, let us extend the concept illustration presented previously in the crisp case (Fig. 2), by considering fuzziness and relation T in the estimation of the context of the set of concepts (Fig. 4), resulting in the following numerical example. Letting A be a normal fuzzy set $A = \text{ball}/1.0 + \text{referee}/0.6 + \text{basket}/0.8$ defined on the crisp set $S = \{\text{ball, referee, basket}\}$. The quantities $cp(S \cdot A(ball))$, $cp(S \cdot A(referee))$ and $cp(S \cdot A(basket))$ are calculated as:

$$cp(S \cdot A(ball)) = cp(ball/1.0 + referee/1.0 + basket/1.0) = ball/0 + referee/0 + basket/0$$
$$cp(S \cdot A(basket)) = cp(ball/0.6 + referee/0.6 + basket/0.6) = ball/0.4 + referee/0.4 + basket/0.4$$
$$cp(S \cdot A(referee)) = cp(ball/0.8 + referee/0.8 + basket/0.8) = ball/0.2 + referee/0.2 + basket/0.2$$

According to Eq. (12), we have:

$$\mathcal{K}(ball) = football/0.8 + basketball/0.8 \tag{14}$$

$$\mathcal{K}(basket) = football/0.75 + basketball/0.75 + ball/0.4 + referee/0.4 + basket/0.4 \tag{15}$$

$$\mathcal{K}(referee) = basketball/0.75 + ball/0.2 + referee/0.2 + basket/0.2 \tag{16}$$

As a result *K(A)* is calculated from (13) as:

$$\begin{aligned} \boldsymbol{K}(\boldsymbol{A}) &= \mathcal{K}(ball) \cap \mathcal{K}(referee) \cap \mathcal{K}(basket) = \\ &= football/0 + basketball/0.75 + ball/0 + referee/0 + basket/0 \\ &= \boldsymbol{basketball/0.75} \end{aligned} \tag{17}$$

Schematically, this is shown in the following Fig. 4.

Considering the semantics of the *T* relation and the process of context determination, it is easy to realize that when the concepts in a set are highly related to a common meaning, the context will have high degrees of membership for the concepts that represent this common meaning. Therefore, the height of the context $h(K(A))$ may be used as a measure of the semantic correlation of concepts in set *A*. We refer to this measure as *intensity* of the context. The intensity of the context also represents the degree of relevance of the concepts in the set.

5 Semantic indexing

As mentioned in the Introduction, the transparent, unified and unsupervised access to heterogeneous multimedia is achieved through the integrated handling of concepts and relations that represent detected objects and events from the low-level processing output, as well as from the accompanying metadata. Low-level analysis results and metadata are both mapped to concepts and construct a semantic index. This process is based on the fuzzy knowledge model described in Section 4. Additionally, multimedia documents are classified to topics (e.g. sports, politics, etc.) through fuzzy clustering of concepts associated to a document according to their common meaning. In this way, the semantic index is further enhanced with more abstract concepts that are difficult to be detected by low-level processing techniques alone.

5.1 Indexing framework

The proposed framework for semantic indexing is shown in Fig. 5. The semantic index that links concepts to multimedia documents is constructed in two steps, depicted as *Concept Detection* and *Classification*. First, multimedia documents are mapped to concepts along with a degree of confidence, where this mapping includes both visual content analysis results and semantic interpretation of lexical terms from the available

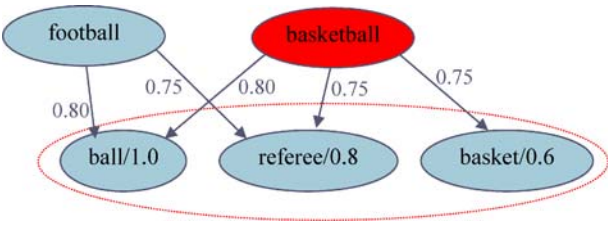

Fig. 4 Similar to the crisp example, concept *basketball* is the only common antecedent of all three concepts *ball*, *referee* and *basket* in this part of relation *T*, i.e. *basketball* is the context of *ball*, *referee* and *basket* with a degree of 0.75

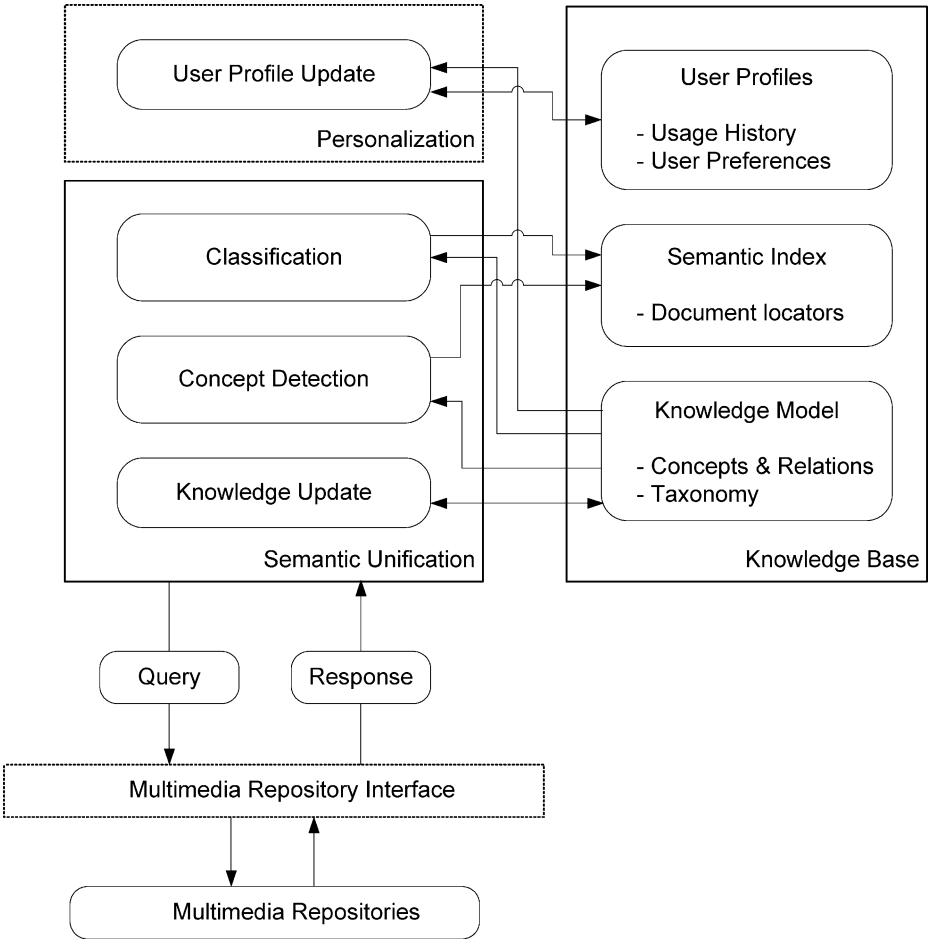

Fig. 5 A schematic diagram for semantic indexing in the context of the proposed framework

textual annotation. On a second step, semantic indexing (associated concepts) of each document is further analyzed in order to estimate the degree to which the given document is related to each one of the predefined topics.

The process of low-level analysis and content-based extraction of concepts is described in [7]. On the other hand, handling of textual annotation, entered manually by an expert and accompanying multimedia documents, is semantically interpreted, i.e., mapped to concepts, and the result is again stored in the semantic index. During this process, concepts are identified through matching with their definitions in the knowledge model similar to the process followed for user query interpretation [85]. Links between detected concepts and the corresponding multimedia document are then added to the index; confidence values are also added depending on the location of each concept in the description and the degree of concept's matching.

5.2 Semantic classification

The indexing procedure described so far refers to the construction of the semantic index, i.e. an association between multimedia documents and concepts, obtained through analysis of either the raw video content or the associated textual annotation. In a further analysis process, each document is analyzed to detect associated topics. This is achieved by clustering concepts associated to a document according to their common meaning. The set to be clustered for document d is its support S_d defined in subsection 4.1:

$$S_d = \{s \in S : I(s,d) > 0\} \tag{18}$$

where I represents the semantic index and is a fuzzy relation between documents d and concepts s, i.e. $I(s, d)$ represents the degree of membership of concept s in document d. Letting G_d be the set of clusters detected in d, each cluster $c \in G_d$ is a crisp set of concepts, $c \subseteq S_d$. However, this alone is not sufficient for our approach, as we need to support documents belonging to multiple distinct topics by different degrees and at the same time retain the robustness and efficiency of the hierarchical clustering approach. Thus, without any loss of functionality or increase of computational cost, we replace the crisp clusters c with fuzzy normalized clusters c^n. Then, aggregating the context of each cluster c^n, we identify the fuzzy set W_d of topics related to document d. The sections below provide details on the initial concept clustering, the cluster fuzzification process, as well as the final topic classification of a document.

5.2.1 Concept clustering

Most clustering methods found in the literature belong to either of two general categories, *partitioning* and *hierarchical* [78]. In contrast to their hierarchical counterparts, partitioning methods require the number of clusters as input. Since the number of topics that may be encountered in a multimedia document is not known beforehand—although the overall number of possible topics Y is available—the latter are inapplicable [52]. The same applies to the use of fuzzy c–means [16], a supervised clustering method which allows one concept to belong to two or more clusters. However, it also requires the number of concept clusters as input, i.e. it uses a hard termination criterion on the amount of clusters. In general, hierarchical methods are divided into agglomerative and divisive. The former are more widely studied and applied, as well as more robust. Their general structure, adjusted for the needs of the problem at hand, is as follows:

1. When considering document d, turn each concept $s \in S_d$ into a singleton, i.e. into a cluster c of its own.
2. For each pair of clusters c_1, c_2 calculate a compatibility indicator $CI(c_1,c_2)$. The CI is also referred to as cluster similarity, or distance metric.
3. Merge the pair of clusters that have the best CI. Depending on whether this is a similarity or a distance metric, the best indicator could be selected using the *max* or the *min* operator, respectively.
4. Continue at step 2, until the termination criterion is satisfied. The termination criterion most commonly used is the definition of a threshold for the value of the best compatibility indicator.

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* for CI, and the identification of

the optimal *terminating step*, i.e. the definition of a meaningful termination criterion. In this work, the intensity of the common context $h(K(c_1 \cup c_2))$ is used as a distance metric for two clusters c_1,c_2 quantifying their semantic correlation. The process terminates when the concepts are clustered into sets that correspond to distinct topics, identified by the fact that their common context has low intensity. Therefore, the termination criterion is a threshold on the selected compatibility metric. The output is a set of clusters G_d, where each cluster $c \in G_d$ is a crisp set of concepts, $c \subseteq S_d$.

5.2.2 Cluster fuzzification

This clustering method determines successfully the count of distinct clusters that exist in S_d, but is inferior to partitioning approaches in the following senses: (i) it only creates crisp clusters, i.e. it does not allow for degrees of membership in the output, and (ii) it only creates partitions, i.e. it does not allow for overlapping among the detected clusters. However, in real-life a concept may be related to a topic to a degree other than 1 or 0, and may also be related to more than one distinct topic. In order to overcome such problems, fuzzification of the clusters is carried out.

In particular, we construct a fuzzy classifier, i.e. a function C_c: $S \rightarrow [0,1]$ that measures the degree of correlation of a concept s with cluster c. Apparently, a concept s should be considered correlated with cluster c, if it is related to the common meaning of the concepts in c. Therefore, the quantity $C_1(c,s) = h(K(c \cup \{s\}))$, forms an appropriate measure of correlation. Of course, not all clusters are equally compact; we may measure cluster compactness using the similarity among the concepts it contains, i.e. using the intensity of the cluster's context. Therefore, the above correlation measure needs to be adjusted to the characteristics of the cluster in question:

$$C_c(s) = \frac{C_1(c,s)}{h(K(c))} = \frac{h(K(c \cup \{s\}))}{h(K(c))} \tag{19}$$

It is easy to see that this measure obviously has the following properties:

- $C_c(s) = 1$, if the semantics of s imply it should belong to c. For example: $C_c(s) = 1$, $\forall s \in c$.
- $C_c(s) = 0$, if the semantics of s imply it should not belong to c.
- $C_c(s) \in (0,1)$, if s is neither totally related, nor totally unrelated to c.

Using this classifier, we may expand the detected crisp clusters to include more concepts. Cluster c is replaced by the *fuzzy* cluster $c^f \supseteq c$: $c^f = \sum_{s \in S_d} s/C_c(s)$, using the sum notation for fuzzy sets.

The process of fuzzy hierarchical clustering has been based on the crisp set S_d, thus ignoring fuzziness in the semantic index. In order to incorporate this information when calculating the clusters that describe a document's content, we adjust the degrees of membership for them as follows:

$$c^i(s) = t\left(c^f(s), I(s,d)\right), \ \forall s \in S_d \tag{20}$$

where t is a t–norm. The semantic nature of this operation demands that t is an Archimedean norm. Each one of the resulting clusters corresponds to one of the distinct topics of the document. In order to determine the topics that are related to a cluster c^i, two

things need to be considered: the scalar cardinality of cluster $|c^i|$ and its context. Since context has been defined only for normal fuzzy sets, we need to *normalize* the cluster as follows:

$$c^n(s) = \frac{c^i(s)}{h(c^i(s))}, \forall s \in S_d \tag{21}$$

5.2.3 Fuzzy topic classification

The first step in order to identify the fuzzy set W_d of topics related to document d, is the calculation of $W(c)$, i.e. the set of topics related to each cluster c. We first estimate W^*, which is derived from the normalized cluster c^n and denotes the output of the process in case of neglecting cluster cardinalities. In general, concepts that are not contained in the context of c^n cannot be considered as being related to the topic of the cluster. Therefore $W(c) \subseteq W^*(c^n) = w(K(c^n))$, where w is a weak modifier. Modifiers (also met in the literature as *linguistic hedges* [47]) are used in this work to adjust mathematically computed values so as to match their semantically anticipated counterparts.

If the concepts that index document d are all clustered in a unique cluster c^i, then $W_d = W^*(c^n)$ is a meaningful approach. On the other hand, when more than one cluster is detected, then it is imperative that cluster cardinalities are considered as well. Clusters of extremely low cardinality probably only contain misleading concepts, and therefore need to be ignored in the estimation of W_d. On the contrary, clusters of high cardinality almost certainly correspond to the distinct topics that d is related to, and need to be considered in the estimation of W_d. The notion of "high cardinality" is modeled with the use of a "large" fuzzy number L(·), which forms a function from the set of real positive numbers to the [0,1] interval, quantifying the notion of "large" or "high". The topics that are related to each cluster are computed, after adjusting membership degrees according to scalar cardinalities, as follows: $W(c) = W^*(c^n) \cdot L(|c^i|)$.

The set of topics that correspond to a document is the set of topics that belong to any of the detected clusters of concepts that index the given document: $W_d = \bigcup_{c \in G_d} W(c)$ where $\cup$ is a fuzzy co-norm and G_d is the set of clusters that have been detected in d. It is easy to see that $W_d(s)$ will be high if a cluster c^i, whose context contains s, is detected in d, and additionally, if the cardinality of c^i is high and the degree of membership of s in the context of the cluster is also high (i.e., if the topic is related to the cluster and the cluster does not consist of misleading concepts). Finally, in order to validate the results of fuzzy classification, i.e. assure that the set of topics W_d that correspond to a document d are derived from the set of all possible topics Y, we compute the quantity $Z_d = W_d \cap Y$.

6 Experimental results

In order to demonstrate the efficiency of the proposed framework and methodology followed herein, we have developed two test scenarios. We first present a minimal scenario with five synthetic documents, and then introduce aggregated classification results over a real-life repository of 484 multimedia documents classified in 13 topics in a twofold approach.

Table 2 Concept names and mnemonics; topics are shown in boldface

Concept	Mnemonic	Concept	Mnemonic
War	**war**	Performer	prf
Tank	tnk	Speak	spk
Missile	msl	**Theater**	**thr**
Explosion	exp	Sitting person	spr
Launch of missile	lms	Screen	scr
Fighter airplane	far	Curtain	crn
Army or police uniform	unf	Seat	sit
F16	f16	Tier	tir
Shoot	sht	**Football**	**fbl**
River	riv	Lawn	lwn
Arts	**art**	Goal	gol
Cinema	**cnm**	Football player	fpl
Scene	scn	Goalkeeper	glk

6.1 A simple scenario

The first scenario comprises five synthetic documents (d_1,...,d_5), a set of concepts S and a small taxonomic relation T defined on S. The set of concepts together with their mnemonics is presented in Table 2, relation T in Table 3 and the classification results in Table 4. Relation elements that are implied by transitivity are omitted for the sake of clarity; *sup-product* is assumed for transitivity and the t-norm used for the transitive closure of relation T is Yager's t-norm with parameter 3. Additionally, the co-norm used in Eq. (12) is the bounded sum, while in (16), the t-norm used is the product. In the implementation of Section's 5.2.3 steps, $w(a) = \sqrt{a}$ is the weak modifier used, whereas the standard co-norm *max* is utilized for final topic extraction. Finally, the threshold used for the termination criterion of the clustering algorithm is 0.3 and the "large" fuzzy number L(·) is defined as the triangular fuzzy number (1.3, 3, ∞)[2] [47].

Document d_1 has been constructed assuming that it contains a shot from a theater hall; the play is war-related. Objects and events are assumed to have been detected with a limited degree of certainty, as is typically expected from the process of extraction of concepts directly from raw media. Furthermore, detected concepts are not always related to the overall topic of the document. For instance, a *tank* may appear in a shot from a theater as part of the play, but this information cannot aid in the process of semantic classification and consequently is ignored by the topic extraction algorithm. The same applies to *speak* as well. The semantic indexing of document d_1 is as follows:

$$d_1 = prf/0.9 + \text{spr}/0.9 + \text{spk}/0.6 + \text{sit}/0.7 + \text{crn}/0.8 + \text{scn}/0.9 + \text{tnk}/0.7 \tag{22}$$

Document d_2 contains a shot from a cinema hall. The film is again war-related. The semantic indexing of d_2 is represented as:

$$d_2 = \text{spr}/0.9 + \text{spk}/0.8 + \text{sit}/0.9 + \text{scr}/1 + \text{tnk}/0.4 \tag{23}$$

[2] Let a, b, $c \in R$, $a < b < c$. The fuzzy number $u{:}R \rightarrow [0,1]$ denoted by (a, b, c) and defined by u(x)=0 if $x \leq a$ or $x \geq c$, $u(x) = \frac{x-a}{b-a}$, if $x \in [a, b]$ and $u(x) = \frac{c-x}{c-b}$ if $x \in [b, c]$ is called a triangular fuzzy number.

Table 3 The taxonomic relation T; zero and implied by reflexivity elements are omitted

s_1	s_2	T	s_1	s_2	T	s_1	s_2	T
war	unf	0.90	war	lms	0.70	war	exp	0.60
war	far	0.80	fbl	lwn	0.90	fbl	gol	0.80
war	tnk	0.80	cnm	scr	0.90	fbl	sit	0.60
war	msl	0.80	cnm	spr	0.80	cnm	sit	0.60
thr	scn	0.90	fbl	spr	0.60	fbl	sht	0.90
thr	prf	0.90	thr	sit	0.60	fbl	tir	0.80
thr	spr	0.80	thr	crn	0.70	fbl	fpl	0.90
far	f16	1.00	art	thr	0.80	art	cnm	0.80
fpl	glk	1.00						

The concept clustering process results into 3 crisp clusters:

$$G_{d_2} = \{c_1, c_2, c_3\} = \{(spr, scr, sit), spk, tnk\} \tag{24}$$

Due to the simplicity of the content of document d_2 and the small amount of its detected concepts, using the context-based classifier introduced in 5.2.2 does not expand the detected crisp clusters to include other concepts. This was already expected by observing the structure of the T relation in Fig. 6, since the semantics of all concepts in d_2 imply either a full or a absolutely absent relation to c. We further adjust the degrees of membership for them according to Eq. (16) and using the product t-norm as follows:

$$c_1^i(s) = spr/0.9 + scr/1.0 + sit/0.9 \tag{25}$$

$$c_2^i(s) = spk/0.8 \tag{26}$$

$$c_3^i(s) = tnk/0.4 \tag{27}$$

Each one of the above clusters corresponds to one of the distinct topics associated with d_2 and in order to determine them we must consider both the scalar cardinality of each cluster, as well as its context. For each cluster of document d_2 we have:

$$h\left(c_1^i(s)\right) = 1.0 \text{ and } \left|c_1^i(s)\right| = 3 \tag{28}$$

Table 4 Document classification results. Values below 0.1 are omitted

Topic	Document				
	d_1	d_2	d_3	d_4	d_5
War					0.77
Arts	0.84	0.77			0.85
Cinema		0.76			0.86
Theater	0.89				0.33
Football			0.84	0.37	0.76

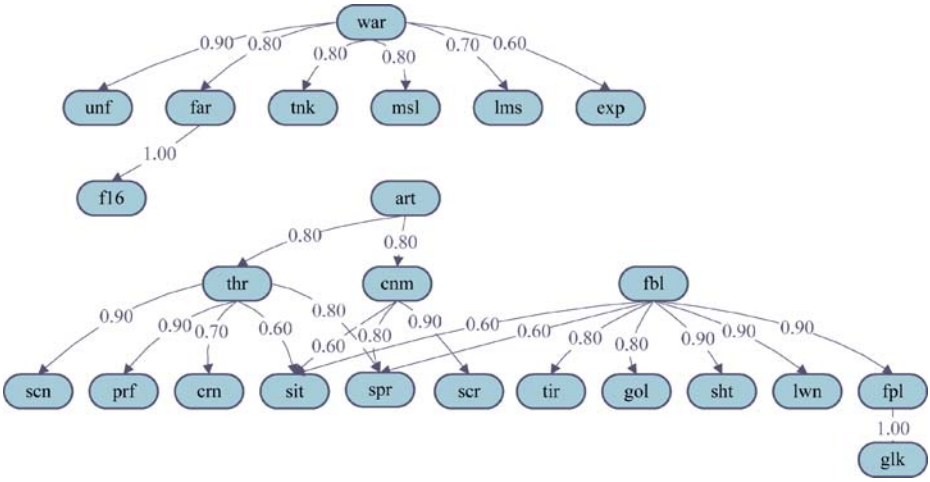

Fig. 6 Example of *T* relation construction

$$h\left(c_2^i(s)\right) = 0.8 \text{ and } \left|c_2^i(s)\right| = 1 \tag{29}$$

$$h\left(c_3^i(s)\right) = 0.4 \text{ and } \left|c_3^i(s)\right| = 1 \tag{30}$$

Consequently, normalization of the above clusters results into:

$$c_1^n(s) = spr/0.9 + scr/1.0 + sit/0.9 \tag{31}$$

$$c_2^n(s) = spk/1.0 \tag{32}$$

$$c_3^n(s) = tnk/1.0 \tag{33}$$

Their context is calculated from Eq. (13) as:

$$K\left(c_1^n(s)\right) = cnm/0.6 + art/0.58 \tag{34}$$

$$K\left(c_2^n(s)\right) = \text{Ø} \tag{35}$$

$$K\left(c_3^n(s)\right) = war/0.8 \tag{36}$$

Applying the weak modifier $w(a) = \sqrt{a}$, we obtain:

$$W^*\left(c_1^n\right) = w\left(K\left(c_1^n\right)\right) = \sqrt{K\left(c_1^n\right)} = cnm/0.77 + art/0.76 \tag{37}$$

$$W^*\left(c_2^n\right) = w\left(K\left(c_2^n\right)\right) = \text{Ø} \tag{38}$$

$$W^*\left(c_3^n\right) = w\left(K\left(c_3^n\right)\right) = \sqrt{K\left(c_3^n\right)} = war/0.89 \tag{39}$$

As described in Section 5.2.3, clusters c_2^n and c_3^n are of extremely low cardinality, thus containing misleading concepts regarding the topics of document d_2. After adjusting the membership degrees of the clusters according to their scalar cardinalities using the triangular fuzzy number (1.3, 3, ∞), both clusters are ignored in the estimation of W_{d_2}. Finally, the set of topics of document d_2 is given by:

$$W_{d_2} = \bigcup_{c \in G_d} W(c) = W(c_1) = W^*\left(c_1^n\right) = cnm/0.77 + art/0.76 \tag{40}$$

Although some concepts are common between documents d_1 and d_2, and they are related to both topics *theater* and *cinema*, the algorithm correctly detects that their overall topic is different. This is accomplished by considering that *screen* alters the context and thus the overall meaning.

Documents d_3 and d_4 are both related to football. Their difference is the certainty with which concepts have been detected in them. As can be seen, the algorithm successfully incorporates uncertainty of the input in its result:

$$\begin{aligned} d_3 = {} & spr/0.8 + \mathrm{unf}/0.9 + \mathrm{lwn}/0.6 + \mathrm{gol}/0.9 \\ & + \mathrm{tir}/0.7 + \mathrm{spk}/0.9 + \mathrm{glk}/0.6 + \mathrm{sht}/0.5 \end{aligned} \tag{41}$$

$$\begin{aligned} d_4 = {} & \mathrm{spr}/0.2 + \mathrm{unf}/0.3 + \mathrm{lwn}/0.4 + \mathrm{gol}/0.3 \\ & + \mathrm{tir}/0.4 + \mathrm{spk}/0.2 + \mathrm{glk}/0.3 + \mathrm{sht}/0.4 \end{aligned} \tag{42}$$

Finally, document d_5 is a sequence of shots from a news broadcast. Due to the diversity of stories presented in it, the concepts that are detected and included in the semantic index are quite unrelated to each other. In order to clarify the process of cluster fuzzification, let us analyze further the specific steps for document d_5. The semantic indexing of the document is given by:

$$\begin{aligned} d_5 = {} & \mathrm{spr}/0.9 + \mathrm{unf}/0.8 + \mathrm{lwn}/0.5 + \mathrm{gol}/0.9 + \mathrm{tir}/0.7 \\ & + \mathrm{spk}/0.9 + \mathrm{glk}/0.8 + \mathrm{sht}/0.5 + \mathrm{prf}/0.7 + \mathrm{sit}/0.9 + \mathrm{crn}/0.7 \\ & + \mathrm{scn}/0.8 + \mathrm{tnk}/0.9 + \mathrm{msl}/0.8 + \mathrm{exp}/0.9 + \mathrm{riv}/1.0 \end{aligned} \tag{43}$$

Considering the fuzziness of the index as described in Section 5.2.2, we compute the following five fuzzy clusters c^i of concepts for document d_5:

$$c_1^i = spk/0.9 \tag{44}$$

$$c_2^i = riv/1.0 \tag{45}$$

$$c_3^i = \mathrm{spr}/0.9 + \mathrm{prf}/0.7 + \mathrm{sit}/0.77 + \mathrm{crn}/0.7 + \mathrm{scn}/0.8 \tag{46}$$

$$c_4^i = \mathrm{spr}/0.9 + \mathrm{lwn}/0.5 + \mathrm{gol}/0.9 + \mathrm{tir}/0.7 + \mathrm{glk}/0.8 + \mathrm{sht}/0.5 + \mathrm{sit}/0.9 \tag{47}$$

$$c_5^i = \mathrm{unf}/0.8 + \mathrm{tnk}/0.9 + \mathrm{msl}/0.8 + \mathrm{exp}/0.9 \tag{48}$$

Concepts such as *seat* and *sitting person* are assigned to more than one cluster, as they are related to more than one of the contexts that are detected in the document. Moreover, the first two clusters c_1^i and c_2^i are ignored in the process of identifying the fuzzy set W_{d_5} of topics related to document d_5, because of their low cardinality. Based on the methodology presented in Section 5.2.3, considering the context of each of the remaining clusters c_3^i, c_4^i and c_5^i, and acknowledging the fact that the latter has been defined only for normal fuzzy sets, we identify the topics related to d_4, as described in the last column of Table 4. We observe that the algorithm successfully identifies the existence of more than one distinct topic in the document.

6.2 Classification performance evaluation

To illustrate further the performance of our methodology, we carried out an evaluation experiment over a real-life repository of a set Q of multimedia documents, derived from the multimedia repositories of the Hellenic Broadcast Corporation (ERT), Film Archive Austria (FAA), Film Archive Greece (FAG) and Austrian Broadcasting Corporation (ORF). All documents are dominated by large diversity in terms of content, making the overall unification effort challenging. The material sums up to about 80 hours of news programs and documentaries distributed across 484 multimedia documents at hand, whereas their duration ranges from 55″ to 35′.28″. Each multimedia document is manually annotated with human understandable keywords, includes a number of multimedia programs and news items and each item contains one to a few decades of concepts, resulting in an overall of approximately 30,000 concepts within the entire set of documents' annotations. Due to the differences in size and contents of the video programs available, we decided to follow a twofold approach on the available multimedia documents; on the one hand we conducted experiments with the original set of multimedia programs and on the other hand we carried out the same series of experiments with the extended set of news items. Since a number of news items constitutes a multimedia program, they are significantly smaller in duration (ranging from 1′.28″ to 5′.12″), but they are still multimedia documents and at the same time their amount is larger, i.e. a total of distinct 1,976 news items are available. In this approach we handle both multimedia programs and news items in a unified way as multimedia documents; this resulted in two sets of experiments, as depicted in the following. For the reason of simplicity and scalability, $|Y|=N=13$ indicative content topics were selected amongst the concepts, namely: *sports, politics, religion, news, leaders, military, art, health, traveling, happening, education, protests* and *history*. All documents were manually classified in advance, to construct the ground truth (GT) for the evaluation of our classification approach. Due to the subjectivity introduced by the manual process, classification of the ground truth was crisp in principal; however, one document—either multimedia program or news item—could belong to multiple topics, due to the possible thematic parts it may contain, resulting into an artificial enlargement of the original programs' and items' data set from 484 to $Q=653$ multimedia programs and from 1,976 to $Q=2{,}733$ news items:

$$Q = \bigcup_{i=1}^{N} Q_i, \quad Q_i \cap Q_j \neq \emptyset, \quad i,j \in \{1,\ldots,N\} \tag{49}$$

where Q_i is the set of multimedia documents associated actually with topic i.

This crisp GT generation approach results into a pessimistic evaluation of our methodology, illustrated in Table 7 and Table 8, however, its implementation is a straightforward task to follow by the annotators and fully represents a real-life situation.

A series of experiments was carried out to measure effectiveness and performance of classification by obtaining *specificity* (*sp*), *sensitivity* (*sn*) and *effectiveness* (*e*) measurements. Let the number of documents related to a topic correctly recognized as belonging to the specific topic, i.e. multimedia documents correctly classified, be denoted by *true positives TP* and the number of documents incorrectly recognized not to belong to this topic, i.e. multimedia documents incorrectly classified, be denoted by *false negatives FN*. Similarly, let the number of documents actually not related to each specific topic under consideration correctly and incorrectly classified be denoted by *true negatives TN* and *false positives FP*, respectively. Then,

$$sp = specificity = \frac{TN}{TN + FP} \tag{50}$$

$$sn = sensitivity = \frac{TP}{TP + FN} \tag{51}$$

$$e = effectiveness = \frac{1}{a(1/sp) + (1-a)(1/sn)} \tag{52}$$

where parameter a influences the estimation of effectiveness e, allowing different weighting of specificity and sensitivity, i.e. a low value of a favors sensitivity, whereas a high value favors specificity. The aim of any experiment is to maximize both *sp* and *sn* values. However, it has been experimentally and theoretically established that an increase in one value, leads to the decreasing of the other. Table 5 and Table 6 illustrate two *fuzzy confusion matrices* containing information about actual and detected topics of multimedia documents (i.e. multimedia programs and news items, respectively), where:

$$D_{ij} = \sum_{d \in Q_i} W_d(j), \quad i,j \in \{1, \ldots, N\} \tag{53}$$

and $W_d(j)$ is the degree to which document d is classified in topic j.

In this case values of TP_t, FP_t, TN_t, FN_t for each topic $t \in \{1,...N\}$ are defined as:

$$TP_t = D_{tt}, \quad FP_t = \sum_{\substack{k=1 \\ k \neq t}}^{N} D_{tk}, \quad TN_t = \sum_{\substack{k=1 \\ k \neq t}}^{N} D_{kk} \quad \text{and} \quad FN_t = \sum_{\substack{k=1 \\ k \neq t}}^{N} D_{kt} \tag{54}$$

For instance:

$$TP_{sports} = 45$$
$$FP_{sports} = 1.04 + 0 + 1.44 + 2.10 + 0 + 0.13 + 0.66 + 0.60 + 1.26 + 0.32 + 0 + 0.14 = 7.69$$
$$TN_{sports} = 35.10 + 22.36 + \ldots + 43.00 = 372.91$$
$$FN_{sports} = 0.72 + 0 + 0.76 + 0 + 0 + 0 + 0 + 0.68 + 0.72 + 0 + 0 + 0.54 = 3.42$$

According to the above definitions, using a value of 0.5 for a, i.e. $e(a = 0.5) = \frac{2(sp)(sn)}{sp+sn}$, and given the fuzzy confusion matrices presented in Table 5 and Table 6, we measured specificity, sensitivity and effectiveness for each one of the 13 topics against the ground truth, as depicted in Table 7 and Table 8, for multimedia programs and news items, respectively. As expected observing both Tables, there is no significant variance of the specificity, sensitivity or effectiveness indices between the different topics. For instance,

Table 5 Fuzzy confusion matrix D corresponding to semantic classification of multimedia programs

Actual topics	Detected topics												
	Sports	Politics	Religion	News	Leaders	Military	Art	Health	Travelling	Happening	Education	Protests	History
Sports	45.00	1.04	0.00	1.44	2.10	0.00	0.13	0.66	0.60	1.26	0.32	0.00	0.14
Politics	0.72	35.10	1.05	1.17	1.35	0.90	1.92	0.48	1.16	2.59	2.97	1.15	2.75
Religion	0.00	0.72	22.36	0.12	2.31	0.90	2.17	0.51	0.00	0.00	1.76	0.44	1.45
News	0.76	0.80	0.00	49.28	0.60	2.45	1.32	1.35	0.58	0.57	1.44	0.88	4.80
Leaders	0.00	0.21	1.95	0.26	33.44	1.86	0.00	0.48	0.45	0.11	0.24	0.60	2.97
Military	0.00	1.14	0.64	0.60	0.28	29.88	0.00	0.00	0.00	0.00	0.00	1.68	2.45
Art	0.00	0.00	0.15	0.00	1.40	0.00	26.40	0.00	1.53	2.80	0.90	0.00	1.02
Health	0.00	0.00	0.00	0.35	0.00	0.00	0.00	22.14	1.26	1.28	1.10	0.00	1.45
Travelling	0.68	0.00	0.16	0.84	0.00	0.00	0.00	0.00	25.52	0.00	0.00	0.00	0.84
Happening	0.72	0.70	0.27	0.16	1.35	0.19	2.80	1.20	1.92	33.93	1.12	1.04	0.42
Education	0.00	0.44	0.52	0.58	0.19	0.00	1.25	1.08	0.60	0.32	25.20	0.35	0.64
Protests	0.00	0.33	0.72	1.50	1.20	0.54	0.00	0.42	0.60	0.54	1.30	26.66	0.87
History	0.54	0.78	0.70	1.20	0.62	0.46	0.00	0.32	1.60	0.96	1.20	1.08	43.00
Total:	48.42	41.26	28.52	57.50	44.84	37.18	35.99	28.64	35.82	44.36	37.55	33.88	62.80

Table 6 Fuzzy confusion matrix D corresponding to semantic classification of news items

Actual topics	Detected topics												
	Sports	Politics	Religion	News	Leaders	Military	Art	Health	Travelling	Happening	Education	Protests	History
Sports	189.00	2.45	0.00	7.91	8.79	0.00	0.41	4.14	0.75	3.96	0.50	0.00	2.20
Politics	0.75	141.30	10.55	26.94	45.22	16.96	1.51	0.75	2.73	3.49	9.33	12.28	19.63
Religion	0.00	3.39	90.30	1.51	5.18	0.00	0.97	0.00	0.00	0.00	1.38	1.38	10.93
News	17.90	47.73	0.00	197.12	7.54	7.69	7.60	2.83	9.11	4.77	16.96	4.84	6.28
Leaders	0.00	4.62	3.67	0.82	140.80	7.79	0.00	0.00	0.94	0.35	1.88	5.02	15.54
Military	0.00	7.16	0.50	9.42	2.64	119.52	0.00	0.00	0.00	0.00	0.00	5.93	24.18
Art	0.00	0.00	2.36	1.26	0.63	0.00	100.00	0.00	0.53	21.35	5.65	0.00	1.07
Health	0.00	0.00	0.00	2.20	0.00	0.00	0.00	91.84	3.96	1.51	4.84	0.00	1.82
Travelling	0.00	0.00	0.00	4.62	0.00	0.00	0.00	0.00	110.88	0.00	0.00	0.00	0.66
Happening	8.29	1.76	0.85	1.51	0.85	0.00	15.07	0.00	0.00	137.46	1.00	8.16	0.88
Education	0.00	3.45	0.00	5.46	1.19	0.00	2.36	5.65	1.88	0.50	109.20	15.39	3.01
Protests	0.00	19.69	1.51	14.13	3.77	1.70	0.00	0.00	0.00	2.83	8.16	103.20	5.46
History	0.00	11.02	8.79	5.65	10.71	8.67	0.00	0.50	1.00	0.00	5.02	6.78	176.30
Total:	215.94	242.57	118.53	278.54	227.31	162.32	127.91	105.72	131.79	176.21	163.93	162.99	267.96

Table 7 Classification evaluation results for the proposed semantic classification and pLSA for 653 multimedia programs

Topics	GT	TP	FP	TN	FN	Semantic			PLSA		
						sp	sn	e	sp	sn	e
Sports	60	45.00	7.69	372.91	3.42	0.98	0.93	0.95	0.95	0.91	0.93
Politics	59	35.10	18.21	382.81	6.16	0.95	0.85	0.90	0.93	0.83	0.87
Religion	37	22.36	10.38	395.55	6.16	0.97	0.78	0.87	0.89	0.72	0.80
News	82	49.28	15.55	368.63	8.22	0.96	0.86	0.91	0.94	0.85	0.89
Leaders	51	33.44	9.13	384.47	11.40	0.98	0.75	0.85	0.88	0.69	0.77
Military	46	29.88	6.79	388.03	7.30	0.98	0.80	0.88	0.89	0.74	0.81
Art	41	26.40	7.80	391.51	9.59	0.98	0.73	0.84	0.91	0.67	0.77
Health	32	22.14	5.44	395.77	6.50	0.99	0.77	0.87	0.94	0.71	0.81
Travelling	35	25.52	2.52	392.39	10.30	0.99	0.71	0.83	0.90	0.66	0.76
Happening	54	33.93	11.89	383.98	10.43	0.97	0.76	0.86	0.94	0.70	0.80
Education	41	25.20	5.97	392.71	12.35	0.99	0.67	0.80	0.96	0.62	0.75
Protests	44	26.66	8.02	391.25	7.22	0.98	0.79	0.87	0.88	0.72	0.79
History	71	43.00	9.46	374.91	19.80	0.98	0.68	0.80	0.95	0.68	0.79
				macro-average		0.97	0.78	0.87	0.92	0.74	0.82
				micro-average		0.99	0.77		0.91	0.71	

Table 8 Classification evaluation results for the proposed semantic classification and pLSA for 2,733 news items

Topics	GT	TP	FP	TN	FN	Semantic			pLSA		
						sp	sn	e	sp	sn	e
Sports	247	189.00	31.12	1517.92	26.94	0.98	0.88	0.92	0.96	0.86	0.91
Politics	301	141.30	150.12	1565.62	101.27	0.91	0.58	0.71	0.90	0.57	0.70
Religion	135	90.30	24.74	1616.62	28.23	0.98	0.76	0.86	0.90	0.70	0.79
News	355	197.12	133.23	1509.80	81.42	0.92	0.71	0.80	0.92	0.71	0.80
Leaders	212	140.80	40.63	1566.12	86.51	0.97	0.62	0.76	0.88	0.57	0.69
Military	196	119.52	49.83	1587.40	42.80	0.97	0.74	0.84	0.85	0.68	0.76
Art	160	100.00	32.84	1606.92	27.91	0.98	0.78	0.87	0.93	0.72	0.81
Health	136	91.84	14.32	1615.08	13.88	0.99	0.87	0.93	0.94	0.80	0.86
Travelling	146	110.88	5.28	1596.04	20.91	0.99	0.84	0.91	0.90	0.77	0.83
Happening	204	137.46	38.37	1569.46	38.75	0.98	0.78	0.87	0.94	0.74	0.83
Education	181	109.20	38.90	1597.72	54.73	0.98	0.67	0.79	0.94	0.61	0.74
Protests	182	103.20	57.24	1603.72	59.79	0.97	0.63	0.76	0.91	0.58	0.71
History	278	176.30	58.15	1530.62	91.66	0.96	0.66	0.78	0.94	0.64	0.77
				macro-average		0.96	0.72	0.82	0.92	0.68	0.78
				micro-average		0.99	0.72		0.92	0.66	

effectiveness factor considering all 13 topics, ranges from 0.80 to 0.95 (Table 7), whereas we notice in general specificity is higher than sensitivity. Furthermore, all topics present sensitivity values close to specificity, denoting an overall satisfactory performance of the proposed framework over the entire data set, considering the variety of topics. Comparing the average specificity and sensitivity values for multimedia programs in Table 7 and news items in Table 8, we see that in both cases specificity is rather high, which means that multimedia documents are not assigned incorrect additional topics, while in the case of the 653 multimedia programs, the sensitivity value outperforms the equivalent value of the 2,773 news items. This is justified by the fact that multimedia programs have larger duration (and thus are indexed by more concepts in the semantic index) and are related to multiple topics, something which validates the choice of a fuzzy semantic index and fuzzy clustering approach; without it, classification of a document to multiple topics would not have been possible.

As already discussed in this paper, we propose a multimedia document indexing and classification methodology utilizing a context and taxonomic knowledge model based on relation *T*. At this point we provide and compare its results against the use of a document indexing and classification technique, which does not utilize knowledge or context in the process. A well-known, suitable technique that attempts to extract implicit semantics without the use of knowledge is one based on Latent Semantic Analysis (LSA); the latter uses no humanly constructed dictionaries, knowledge bases, semantic networks, grammars, syntactic parsers or morphologies and takes as input only raw text, as clearly stated in [48]. In this case, we consider documents and concepts obtained by text analysis as the algorithm's input and without utilizing knowledge information, we attempt to identify topics, by associating them to unobserved classes. This problem is very well tackled by a probabilistic view on LSA, namely Probabilistic Latent Semantic Analysis (pLSA) [40], which has a sound statistical foundation and utilizes a latent variable model for unobserved classes. Starting from documents and concepts, pLSA can identify implicit unobserved class variables and has been partially used for text categorization in [25]. We implemented the pLSA model herein, using concepts (deriving either from textual annotation or video analysis results), multimedia documents and topics instead of terms, documents and latent classes, respectively, in order to perform classification of multimedia documents into the 13 topics. pLSA results are shown in the last three columns of Table 7 and Table 8.

Macro-averaging and micro-averaging observed in both tables are two conventional methods to average performance across topics. Macro-averaged scores are averaged values over the number of topics, whereas micro-averaged scores are averaged values over the number of all counted documents. As a first comparative observation, the proposed approach provides better results when dealing with more precise topics, such as health, traveling or religion, than when tackling general topic categories, such as sports, politics or news and this is expected, considering the fact that the notion of the former is easier to identify in the taxonomic knowledge. Furthermore, in both cases of multimedia programs and news items (Table 7 and Table 8), results obtained from the proposed semantic classification method outperform the ones obtained from the application of the pLSA algorithm and improvement varies up to 14%. This is also anticipated, since in our approach we utilize context and a taxonomic knowledge model that defines explicitly the relations between concepts and topics. The latter is impossible to tackle within the pLSA approach, which even without the use of knowledge does provide promising results. Consequently, the main observation of the results denotes that a statistical approach, as the one followed by pLSA, cannot surpass a knowledge-driven approach, as the one proposed herein.

7 Conclusions and future work

The overall core contribution of this work has been the provision of uniform access to heterogeneous multimedia repositories. This is accomplished by mapping all multimedia content and metadata to a semantic index used to serve user queries, based on a common taxonomic knowledge model. A key aspect in these developments has been the exploitation of semantic metadata. Moreover, a semantic classification of multimedia documents to associated topics is performed following a novel fuzzy hierarchical clustering technique. The latter is performed by clustering semantic concepts associated to each document according to their taxonomic context and supports identification of multiple distinct topics per document. Classification results are presented over two experimental scenarios on multimedia repositories tackling both synthetic and real-life data and results are very promising.

The methodologies presented in this paper can be exploited towards the development of more intelligent, efficient and personalized content access systems. However, in order to further verify their efficiency when faced with real-life data, we have implemented and tested them thoroughly in the framework of a multimedia retrieval, personalization and filtering application presented in the second part of this work to follow, "Semantic Representation of Multimedia Content: Retrieval and Personalization". Another interesting perspective presented in this sequel is personalized content retrieval based on usage history.

References

1. Akrivas G, Stamou G, Kollias S (2004) Semantic association of multimedia document descriptions through fuzzy relational algebra and fuzzy reasoning. IEEE Trans Syst Man Cybern Part A, 34:(2), March
2. Akrivas G, Wallace M, Andreou G, Stamou G, Kollias S (2002) "Context-Sensitive Semantic Query Expansion", Proceedings of the IEEE international conference on artificial intelligence systems (ICAIS), Divnomorskoe, Russia, September 2002
3. Altenschmidt C, Biskup J (2002) Explicit representation of constrained schema mappings for mediated data integration. In: Bhalla S (ed) Databases in networked information systems, pp 103–132
4. Altenschmidt C, Biskup J, Flegel U, Karabulut Y (2003) Secure mediation: requirements, design, and architecture. J Comput Secur 11(3):365–398, March
5. Amir A et al (2003) IBM research TRECVID-2003 video retrieval system. Proceedings of NIST TRECVID workshop, Gaithersburg, MD, USA, November 2003
6. Argillander J, Iyengar G, Nock H (2005) Semantic annotation of multimedia using maximum entropy models. Proceedings of IEEE international conference on acoustics, speech, and signal processing, (ICASSP '05), March 2005
7. Athanasiadis Th, Avrithis Y (2004) Adding semantics to audiovisual content. Proceedings of the international conference for image and video retrieval (CIVR '04), Dublin, Ireland, July 2004
8. Athanasiadis Th, Tzouvaras V, Petridis K, Precioso F, Avrithis Y, Kompatsiaris Y (2005) Using a multimedia ontology infrastructure for semantic annotation of multimedia content. Proceedings of the 5th international workshop on knowledge markup and semantic annotation (SemAnnot '05). Galway, Ireland, November 2005
9. Baeza-Yates RA, Ribeiro-Neto BA (1999) Modern information retrieval. ACM Press/Addison-Wesley
10. Benitez AB, Chang S-F (2003) Extraction, description and application of multimedia using MPEG-7. Proceedings of the 37th Asilomar conference on signals, systems and computers. Pacific Grove, California, USA, November 2003
11. Benitez AB, Chang S-F (2003) Image classification using multimedia knowledge networks. Proceedings of the IEEE international conference on image processing (ICIP'03). Barcelona, Spain 2003

12. Benitez AB et al (2000) Object-based multimedia content description schemes and applications for MPEG-7. Image Communication Journal 16:235–269 (invited paper on a special issue on MPEG-7)
13. Benitez AB, Chang S-F, Smith JR (2001) "IMKA: a multimedia organization system combining perceptual and semantic knowledge". Proceedings of the 9th ACM multimedia, Ottawa, Canada 2001
14. Benitez AB, Zhong D, Chang S, Smith J (2001) MPEG-7 MDS content description tools and applications. Proceedings of the international conference on computer analysis of images and patterns (CAIP), Warsaw, Poland
15. Benitez AB et al (2002) Semantics of multimedia in MPEG-7. Proceedings of the IEEE international conference on image processing, vol. 1, pp 137–140
16. Benkhalifa M, Bensaid A, Mouradi A (1999) Text categorization using the semi-supervised fuzzy c-means algorithm". Proceedings of the 18th international conference of the North American Fuzzy Information Processing Society-NAFIPS, pp 561–565
17. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Sci Am 28(5):34–43
18. Berry MW, Dumais ST, O'Brien GW (1995) Using linear algebra for intelligent information retrieval. SIAM Rev 37(4):177–196
19. Bertini M, Cucchiara R, Del Bimbo A, Torniai C (2005) Video annotation with pictorially enriched ontologies. Proceedings of the IEEE international conference on multimedia and expo, Amsterdam, The Netherlands, July 2005
20. Bertini M, Del Bimbo A, Torniai C (2005) Automatic video annotation using ontologies extended with visual information. Proceedings of the 13th annual ACM international conference on Multimedia, Singapore, November 2005
21. Biskup J, Freitag J, Karabulut Y, Sprick B (1997) A mediator for multimedia systems. Proceedings of the 3rd international workshop on multimedia information systems, Como, Italy, September 1997
22. Bloehdorn S et al (2005) Semantic annotation of images and videos for multimedia analysis. Lecture notes in computer science—The semantic web: research and applications, vol. 3532, Springer, pp 592–607
23. Burgin R (1995) The retrieval effectiveness of five clustering algorithms as a function of indexing exhaustivity. J Am Soc Inf Sci 46(8):562–572
24. Burnett I et al (2003) MPEG-21 goals and achievements. IEEE Multimedia 10(4):60–70
25. Cai L, Hofmann T (2003) Text categorization by boosting automatically extracted concepts. Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval, Toronto, Canada, July/August 2003, pp 182–189
26. Cutting D, Karger DR, Pedersen JO, Tukey JW (1992) Scatter/Gather: a cluster-based approach to browsing large document collections. Proceedings of the ACM/SIGIR, pp 318–329
27. Deerwester SC, Dumais ST, Landauer TK, Furnas GW, Harshman RA (1990) Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci 41(6):391–407
28. Denoyer L, Gallinari P, Vittaut J-N, Brunesseaux S (2003) Structured multimedia document classification. Proceedings of the ACM DOCENG conference, Grenoble, France
29. Doerr M, Hunter J, Lagoze C (2003) Towards a core ontology for information integration. J Digit Inf 4(1), April
30. Dorai C, Venkatesh S (2001) Computational media aesthetics: finding meaning beautiful. IEEE Multimed 8(4):10–12
31. Fagin R, Kumar R, Sivakumar D (2003) Efficient similarity search and classification via rank aggregation. Proceedings of the 2003 ACM SIGMOD international conference on management of data, San Diego, California, USA, June 2003, pp 301–312
32. Fagin R, Lotem A, Naor M (2003) Optimal aggregation algorithms for middleware. J Comput Syst Sci 66:614–656
33. García R, Celma O (2005) Semantic integration and retrieval of multimedia metadata. Proceedings of the 5th international workshop on knowledge markup and semantic annotation (SemAnnot), Galway, Ireland, November 2005
34. Gruber TR (1993) A translation approach to portable ontology specification. Knowl Acquis 5:199–220
35. Hauptmann AG (2004) Towards a large scale concept ontology for broadcast video. Proceedings of the 3rd international conference on image and video retrieval (CIVR'04), Dublin, Ireland, July 2004
36. Hauptmann AG (2005) Lessons for the future from a decade of informedia video analysis research. Lect Notes Comput Sci 3568:1–10
37. Hauptmann AG, Yan R, Ng TD, Lin W, Jin R, Derthick M, Christel M, Chen M, Baron R (2002) Video classification and retrieval with the informedia digital video library system. Proceedings of the text and retrieval conference (TREC02), Gaithersburg, MD, USA, November 2002
38. Hauptmann AG et al (2003) Informedia at TRECVID 2003: analyzing and searching broadcast news video. Proceedings of the NIST TRECVID workshop, Gaithersburg, MD, USA, November 2003

39. Henderson JM, Hollingworth A (1999) High level scene perception. Annu Rev Psychol 50:243–271
40. Hofmann T (1999) Probabilistic latent semantic indexing. Proceedings of the 22nd ACM-SIGIR international conference on research and development in information retrieval, pp 50–57
41. Hollink L, Worring M, Schreiber G (2005) Building a visual ontology for video retrieval. Proceedings of the ACM multimedia, Singapore, November 2005
42. Hoogs A, Rittscher J, Stein G, Schmiederer J (2003) Video content annotation using visual analysis and a large semantic knowledgebase. Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), Madison, Wisconsin, USA, June 2003
43. Hunter J (1999) A proposal for an MPEG-7 description definition language. MPEG-7 AHG test and evaluation meeting, Lancaster, February 1999
44. Hunter J (2001) Adding multimedia to the semantic web-building an MPEG-7 ontology. Proceedings of the international semantic web working symposium (SWWS), California, USA, July 30–August 1
45. Hunter J (2003) Enhancing the semantic interoperability of multimedia through a core ontology. IEEE Trans Circuits Syst Video Technol 13(1):49–58
46. ISO/IEC FDIS 15938-5, ISO/IEC JTC 1/SC 29 M 4242 (2001) Information technology multimedia content description interface Part 5: multimedia description schemes, pp 442–448, October 2001
47. Klir G, Bo Yuan (1995) Fuzzy sets and fuzzy logic, theory and applications. Prentice Hall, New Jersey
48. Landauer T, Foltz P, Laham D (1998) An introduction to latent semantic analysis. Discourse Process 25:259–284
49. MacLeod K (1990) An application specific neural model for document clustering. Proceedings of the 4th annual parallel processing symposium, vol. 1, pp 5–16
50. Mich O, Brunelli R, Modena CM (1999) A survey on video indexing. J Vis Commun Image Represent 10:78–112
51. Milanese R (1993) Detecting salient regions in an image: from biology to implementation. *PhD Thesis*, University of Geneva, Switzerland
52. Miyamoto S (1990) Fuzzy sets in information retrieval and cluster analysis. Kluwer Academic Publishers, Dordrecht/Boston/London
53. MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm
54. Mylonas Ph, Avrithis Y (2005) Context modeling for multimedia analysis and use. Proceedings of the 5th international and interdisciplinary conference on modeling and using context (CONTEXT '05), Paris, France 2005
55. Naphade M, Huang T (2001) A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Trans Multimedia 3(1):141–151
56. Naphade MR, Kozintsev IV, Huang TS (2002) A factor graph framework for semantic video indexing. IEEE Trans Circuits Syst Video Technol 12(1):40–52, January
57. NIST TRECVID (2006), http://www-nlpir.nist.gov/projects/trecvid/
58. Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comp Vis 42:145–175
59. Osberger W, Maeder AJ (1998) Automatic identification of perceptually important regions in an image. Proceedings of IEEE International Conference on Pattern Recognition
60. Papadopoulos G, Mylonas Ph, Mezaris V, Avrithis Y, Kompatsiaris I (2006) Knowledge-assisted image analysis based on context and spatial optimization. International Journal on Semantic Web and Information Systems 2(3):17–36
61. Petridis K et al (2006) Knowledge representation and semantic annotation of multimedia content. IEE Proc Vis Image Signal Process (special issue on knowledge-based digital media processing) 153(3):255–262, June 2006
62. Rapantzikos K, Avrithis Y, Kollias S (2005) On the use of spatiotemporal visual attention for video classification". Proceedings of international workshop on very low bitrate video coding (VLBV '05), Sardinia, Italy, September 2005
63. Sahami et al (1997) Real-time full-text clustering of networked documents. Proceedings of the National Conference on Artificial Intelligence, p 845
64. Salembier P, Smith JR (2001) MPEG-7 multimedia description schemes. IEEE Trans Circuits Syst Video Technol 11(6):748–759
65. Schutze et al (1997) Craig projections for efficient document clustering. SIGIR Forum (ACM Special Interest Group on Information Retrieval), pp 74–81
66. Sebastiani F (2002) Machine learning in automated text categorization. ACM Comput Surv 34(1):1–47
67. Sikora T (2001) The MPEG-7 Visual standard for content description—an overview. IEEE Trans Circuits Syst Video Technol (special issue on MPEG-7) 11(6):696–702

68. Simou N, Saathoff C, Dasiopoulou S, Spyrou E, Voisine N, Tzouvaras V, Kompatsiaris I, Avrithis Y, Staab S (2005) An ontology infrastructure for multimedia reasoning. International workshop VLBV05, Sardinia, Italy, September 2005
69. Simou N, Tzouvaras V, Avrithis Y, Stamou G, Kollias S (2005) A visual descriptor ontology for multimedia reasoning. Proceedings of the workshop on image analysis for multimedia interactive services (WIAMIS '05), Montreux, Switzerland, April 2005
70. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380
71. Smith JR (2004) Video indexing and retrieval using MPEG-7. In: B Furht, O Marques (eds) The handbook of image and video databases: design and applications. CRC Press
72. Smith JR (2006) "MARVEL: Multimedia Analysis and Retrieval System", http://www.research.ibm.com/marvel/details.html (November 6)
73. Snoek C et al (2005) MediaMill: exploring news video archives based on learned semantics. Proceedings of ACM Multimedia, Singapore, November 2005
74. Snoek C, Worring M, Geusebroek J-M, Koelma D, Seinstra F, Smeulders A (2006) The semantic pathfinder for generic news video indexing. Proceedings of the 2006 international conference on multimedia and expo (ICME), Toronto, Canada, July 2006
75. Snoek C, Worring M, Hauptmann A (2006) Learning rich semantics from news video archives by style analysis. ACM Transactions on Multimedia Computing, Communications and Applications, 2(2):91–108
76. Staab S, Studer R (2004) Handbook on ontologies. International handbooks on information systems. Springer-Verlag, Heidelberg, New York
77. Stamou G, Kollias S (eds) (2005) Multimedia content and the semantic web: methods, standards and tools. Wiley & Sons Ltd
78. Theodoridis S, Koutroumbas K (1998) Pattern recognition. Academic Press
79. Troncy R (2003) Integrating structure and semantics into audio-visual documents. Proceedings of the 2nd international semantic web conference (ISWC'03), LNCS 2870, Florida, USA, October 2003, pp 566–581
80. Tsechpenakis G, Akrivas G, Andreou G, Stamou G, Kollias S (2002) Knowledge-assisted video analysis and object detection. Proceedings of European symposium on intelligent technologies, hybrid systems and their implementation on smart adaptive systems (Eunite02), Albufeira, Portugal, September 2002
81. Tsinaraki C, Polydoros P, Christodoulakis S (2004) Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing. Proceedings of the 16th international conference on advanced information systems engineering (CAiSE 2004), Riga, Latvia, June 2004
82. Tzitzikas Y, Meghini C, Spyratos N (2004) Towards a generalized interaction scheme for information access. Foundations of information and knowledge systems: third international symposium (FoIKS 2004), Wilheminenburg Castle, Austria, February 17–20, 2004
83. Voisine N, Dasiopoulou S, Mezaris V, Spyrou E, Athanasiadis Th, Kompatsiaris I, Avrithis Y, Strintzis MG (2005) Knowledge-assisted video analysis using a genetic algorithm. Proceedings of the 6th international workshop on image analysis for multimedia interactive services (WIAMIS 2005), April 2005
84. Wallace M, Akrivas G, Mylonas Ph, Avrithis Y, Kollias S (2003) Using context and fuzzy relations to interpret multimedia content. Proceedings of the 3rd international workshop on content-based multimedia indexing (CBMI), IRISA, Rennes, France, September 2003
85. Wallace M, Avrithis Y, Stamou G, Kollias S (2005) Knowledge-based multimedia content indexing and retrieval. In: Stamou G, Kollias S (eds) Multimedia content and semantic web: methods, standards and tools. Wiley
86. Wallace M, Avrithis Y, Kollias S (2006) Computationally efficient sup-t transitive closure for sparse fuzzy binary relations. Fuzzy Sets Syst 157(3):341–372
87. Willett P (1988) Recent trends in hierarchic document clustering: a critical review. Inf Process Manag 24(5):577–597
88. W3C, Semantic Web, www.w3.org/2001/sw: (November 6, 2006).
89. W3C, SWBPD MM Task Force Description, http://www.w3.org/2001/sw/BestPractices/MM/image_annotation.html (November 6, 2006).
90. W3C, Web Ontology Language-OWL, http://www.w3.org/TR/owl-features/ (November 6, 2006).
91. W3C, XML Schema, http://www.w3.org/XML/Schema (November 6, 2006).
92. Zhao R, Grosky WI (2002) Narrowing the semantic gap-improved text-based web document retrieval using visual features. IEEE Trans Multimedia (special issue on multimedia databases) 4(2), June 2002
93. Zhong D, Chang S-F (1999) An integrated system for content-based video object segmentation and retrieval. IEEE Trans Circuits Syst Video Technol 9(8):1259–1268, December

Phivos Mylonas MSc (Computer Science), is currently a Researcher by the Image, Video and Multimedia Laboratory, School of Electrical and Computer Engineering, Department of Computer Science of the National Technical University of Athens, Greece. He obtained his Diploma in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 2001, his Master of Science in Advanced Information Systems from the National & Kapodestrian University of Athens (UoA) in 2003 and is currently pursuing his Ph.D. degree at the former University. His research interests lie in the areas of content-based information retrieval, visual context representation and analysis, knowledge-assisted multimedia analysis, issues related to personalization, user adaptation, user modeling and profiling, utilizing fuzzy ontological knowledge aspects. He has published 10 international journals and book chapters, he is the author of 23 papers in international conferences and workshops, a reviewer for Multimedia Tools and Applications and IEEE Transactions on Circuits and Systems for Video Technology journals and has been involved in the organization of 7 international conferences and workshops. He is an IEEE member since 1999, an ACM member since 2001, a member of the Technical Chamber of Greece since 2001 and a member of the Hellenic Association of Mechanical & Electrical Engineers since 2002.

Thanos Athanasiadis was born in Kavala, Greece, in 1980. He received the Diploma in electrical and computer engineering from the Department of Electrical Engineering, National Technical University of Athens (NTUA), Athens, Greece, in 2003. He is currently working toward the Ph.D. degree at the Image Video and Multimedia Laboratory, NTUA. His research interests include knowledge-assisted multimedia analysis, image segmentation, multimedia content description, as well as content-based multimedia indexing and retrieval.

Manolis Wallace was born in Athens, Greece, in 1977. He received the Diploma in electrical and computer engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2001, and the Ph. D. degree in Electrical and Computer Engineering from the Computer Science Division, NTUA, in 2005. He has been with the University of Indianapolis, Athens, Greece, since 2001, where he currently serves as an Assistant Professor. Since 2004, he has been the Acting Chair of the Department of Computer Science. His main research interests include handling of uncertainty, information systems, data mining, personalization, and applications of technology in education. He has published more than 40 papers in the above fields, ten of which in international journals. He is the Guest Editor of two journal special issues and a coauthor of a book on image processing. He is a reviewer of five IEEE Transactions and of five other journals. Dr. Wallace has participated in various conferences as a Session Organizer or Program Committee Member and he is the Organizing Committee Chair of AIAI 2006.

Yannis Avrithis was born in Athens, Greece in 1970. He received the Diploma degree in Electrical and Computer Engineering (ECE) from the National Technical University of Athens (NTUA) in 1993, the M.Sc. degree in Communications and Signal Processing (with Distinction) from the Department if Electrical and Electronic Engineering of Imperial College of Science, Technology and Medicine, University of London, in 1994, and the Ph.D. degree in ECE from NTUA in 2001. He is currently a senior researcher at the Image, Video and Multimedia Systems Laboratory of the ECE School of NTUA, conducting research in the area of semantic image and video analysis, coordinating R&D activities in national and European projects, and lecturing in NTUA. His research interests include spatiotemporal image / video segmentation and interpretation, knowledge-assisted multimedia analysis, content-based and semantic indexing and retrieval, video summarization, automatic and semi-automatic multimedia annotation, personalization, and multimedia databases. He has been involved in 13 European and 9 National R&D projects, and has published 23 articles in international journals, books and standards, and 50 in conferences and workshops in the above areas. He has contributed to the organiza-tion of 13 international conferences and workshops, and is a reviewer in 15 conferences and 13 scientific journals. He is an IEEE member, and a member of ACM, EURASIP and the Technical Chamber of Greece.

Stefanos Kollias was born in Athens, Greece, in 1956. He received the Diploma in electrical and computer engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 1979, the M.Sc. degree in communication engineering in 1980 from UMIST, Manchester, U.K., and the Ph.D. degree in signal processing from the Computer Science Division, NTUA. He has been with the Electrical Engineering Department, NTUA, since 1986, where he currently serves as a Professor. Since 1990, he has been the Director of the Image, Video, and Multimedia Systems Laboratory, NTUA. He has published more than 120 papers, 50 of which in international journals. He has been a member of the technical or advisory committee or invited speaker in 40 international conferences. He is a reviewer of ten IEEE Transactions and of ten other journals. Ten graduate students have completed their doctorate under his supervision, while another ten are currently performing their Ph.D. thesis. He and his team have been participating in 38 European and national projects.