A MPEG-Based Architecture for Generic Distributed Multimedia Scenarios

Anna Carreras[1], Ruben Tous[1], Eva Rodríguez[1], Jaime Delgado[1], Giovanni Cordara[2], Gianluca Francini[2], Diego Gibellino[2]

[1] Universitat Politècnica de Catalunya
Departament d'Arquitectura de Computadors
Campus Nord, Mòdul D6, Jordi Girona 1-3
E-08034 Barcelona, Spain
{annac, rtous, evar, jaime.delgado}@ac.upc.edu

[2] Telecom Italia Lab, Via Reiss Romoli
274 -10148 Torino, Italy
{giovanni.cordara, gianluca.francini, diego.gibellino}@telecomitalia.it

Journal of Digital Information Management

ABSTRACT: *This paper presents a standards-based architecture for a complex and generic distributed multimedia scenario, which combines content search and retrieval, DRM, and context-based content adaptation together. It is an innovative and totally generic approach trying to narrow the semantic-gap by integrating a flexible language for multimedia search based on MPEG Query Format (MPQF) standard with the application of video analysis algorithms for the automatic extraction of lowlevel features and with the use of contextual information.*

Categories and Subject Descriptors
H.5.1 [Multimedia Information Systems]: H.2.4 [Systems]; Query Processing: H.3.3 [Information Search and Retrieval]

General Terms : Mobile Multimedia Services, Ubiquitous computing, Wireless data transmission, Context aware mobile system

Keywords: Context-based adaptation, Digital Rights Management-Rights Expression Language, Features extraction, MPEG Query Format.

Received: 1 July 2009; Revised 16 August 2009; Accepted 24 August 2009

1. Introduction

A substantial amount of work has been done in the area of Universal Multimedia Access (UMA) and the most recent works focus on the maximisation of user experience. Nevertheless, these approaches are usually application-specific and it is easy to identify serious limitations in terms of interoperability and extensibility. The majority of the research activities reported in this area focus mainly on content adaptation [1], where the use of contextual information and metadata is essential to achieve efficient and useful adaptations that enrich the user experience. Furthermore, the rising tide of available content has created the need of new tools for guiding users to be able to search and find what is of their interest. In both (content adaptation and search) applications, similar problems need to be addressed:

- The lack of context and metadata textual descriptors.
- The semantic gap, i.e. the gap between the low-level features (LLFs) that can be automatically extracted from digital contents and the interpretation that a user would have of the same content.
- The use of standards.

The affordability of consumer electronic devices such as MP3 and recordable players, digital cameras, etc., allow users to become content producers as well as consumers. This evolution creates new and interesting challenges: the abovementioned issues are tightly related to the industry exploitation of new search and retrieval solutions, able to address the specific requirements of the emerging social networks featuring tons of user generated content (UGC). The combination of techniques taking advantage of automatically extracted features, textual metadata and context information can represent the key for the success of such services, offering simple and seamless management of personal digital media repositories in the home network or on the big Internet, as well as Premium content catalogue browsing and search features.

Furthermore, a multipurpose framework dealing with heterogeneous contents needs to take into account the enforcement of Digital Rights Management (DRM) technologies during content access and consumption. Such a feature represents a key issue for business models design, ensuring a transparent and correct usage of the content throughout each stage of the value chain (from the content creator, through the service provider, and to the end user).

Following the background information about the different lines of research activities integrated in the theme of the proposed work (Section 2), this paper will address all the identified challenges by presenting a standards-based architecture for a complex and generic distributed multimedia scenario, which combines search and retrieval, DRM and context-based content adaptation (Section 3). Finally, before the conclusions and the future work, an application scenario based on Social Networks is described in order to better evaluate our proposal (Section 4).

2. Background

2.1. Multimedia Search and Retrieval

In this section we will first analyse the requirements of today's multimedia search and retrieval services, and present the MPEG Query Format (MPQF) as the most suitable solution as it satisfies those requirements. Finally, we will also refer to different search and retrievals algorithms based on video processing techniques.

2.1.1. Unified Querying Languages and Interfaces

The first thing to take into account when defining a search and retrieval service is that user information needs can be expressed in many different ways. On the one hand, when search preferences can be expressed in terms of precise conditions as those in a relational algebra expression, clearly determining which objects of collection to select, it is known as data retrieval (DR). In this case, a single erroneous object among a thousand retrieved objects means a total failure. In the context of multimedia search and retrieval, DR refers to queries expressed in terms of metadata and also in terms of low-level features. On the other hand, there are user information needs which cannot be easily formalized. Information retrieval (IR) aims to retrieve information which might be relevant to the user, given a query written from the user's point of view. In the context of multimedia search and retrieval, IR refers to text keywords and query-byexample (QBE) for instance.

Querying today's digital contents can imply the combination of data retrieval-like conditions referred to a well-defined data model and also information retrieval-like conditions.

Many modern multimedia databases (MMDBs) and various providers of multimedia search and retrieval services already offer advanced indexing and retrieval techniques for multimedia contents. However, their databases and service interfaces are proprietary, and therefore the solutions differ and do not interoperate.

Our proposed search and retrieval service is based on MPEG Query Format (MPQF) standard in order to guarantee the interoperability needed to ease the access to repositories by users and applications, and to allow the deployment of distributed search and aggregation services.

2.1.2. MPEG Query Format Overview

The MPEG Query Format (MPQF) is Part 12 of ISO/IEC 15938-12, "Information Technology -Multimedia Content Description Interface", better known as MPEG-7. The standardization process in this area started in 2006, and MPQF became an ISO/IEC final standard after the 85th MPEG meeting in July 2008.

MPQF is an XML-based query language that defines the format of queries and replies to be interchanged between clients and servers in a distributed multimedia information search-and-retrieval context. The two main benefits of standardising such kind of a language are 1) interoperability between parties (e.g., content providers, aggregators and user agents) and 2) platform independence (developers can write their applications involving multimedia queries independently of the database used, which fosters software reusability and maintainability). The major advantage of having MPEG rather than industry forums leading this initiative is that MPEG specifies international, open standards targeting all possible application domains, which are not conditioned by partial interests or restrictions.

2.1.3. Multimedia Search and Retrieval Techniques

Video data can be indexed based on its audiovisual content (such as colour, speech, motion, shape, and intensity), and semantic content in the form of text annotations. Because machine understanding of the video data is still an unsolved research problem, text annotations are often used to describe the content of video data according to the annotator's understanding and the purpose of that video data.

As far as indexing and retrieval techniques for the visual content are concerned, content-based solutions propose a set of methods based on low-level features such, for example, colours and textures. Several frameworks dealing with the automatic extraction of low level features have been proposed [3]; their main disadvantage, however, relies in the impossibility for such systems to process complex queries to express high level semantic concepts, like, for example "find a video of my sister on a beach at the sunset". Some recent technologies allow the indexing of content based on high level concepts through specific algorithms, but the categorization is limited to few concepts due to the implicit constraints imposed by those algorithms. In TRECVID 2007, the search task consisted of finding shots in a test collection satisfying queries expressed by *topics* – a kind of complex high level features. Examples of such topics are: "waterfront with water and buildings" and "street protest or parade". This TRECVID contest regarded 24 topics. The recall and precision values are still quite modest, as compared to other information retrieval scenarios.

Some opportunities to improve those systems can be offered by the combination of signal and symbolic characterizations in order to diminish the semantic gap and support more general queries: this approach allows to take into account low level and high level concepts and to enable different query paradigms (search by similarity, search by analogy, etc.).

2.2. Concepts and Models for Context and Metadata

Even if the study of metadata and context has been carried out for many decades, nowadays there is still some confusion on defining and modelling context and metadata.

For example, when dealing with Information Services, the dynamic behaviour of some metadata descriptors is sometimes interpreted as context, as in [4]. In other works usually focused on mobile applications, such as [5], context metadata is defined as "information that describes the context in which a certain content item was created". And finally, advanced works trying to integrate context and content like [6] decided that "the term context refers to whatever is common among a set of elements".

Context and metadata are clearly associated to knowledge; "meta-data" is information about data, but what is context doesn't seem to be so clear. We agree with the probably most generic definition found in the literature provided by A. Dey [7]:

"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves."

Furthermore, while metadata is supported by really mature standardised schemas such as the one defined by MPEG-7, there is a clear need of defining a common schema of contextual information for generic multimedia scenarios. If not, it will be difficult to take the maximum advantage of it use because the advanced works in the area (such as the previously identified) won't be able to be extended and interoperable with the incoming complex multimedia scenarios of the future.

2.3. Context-Aware Content Adaptation

Context-aware content adaptation has become an important line of research, however it has always lacked of standardised models to represent and manage contextual information. Three main initiatives should be identified: the CC/PP (Composite Capability/Preference Profile) created by the W3C which defines an RDF-based framework for describing device capabilities and user preferences, the UAProf (User Agent Profile) of the Mobile Alliance Forum which provides an open vocabulary for WAP (Wireless Access Protocol) clients to communicate their capabilities to servers, and the Usage Environment Description (UED) tool

included in MPEG-21 Digital Item Adaptation (DIA) which consists of a complete set of context descriptors for a multimedia adaptation scenario. It includes user information, network and terminal capabilities as well as natural environment descriptors.

The first two are limited to specific applications, and represent a small subset of contextual information only. Without doubt, the most complete initiative trying to identify and represent the context for generic multimedia applications has been carried out by the MPEG community by means of the 7th part of its MPEG-21 standard. It is called MPEG-21 DIA (Digital Item Adaptation) and includes all kind of descriptors to facilitate context-based content adaptation.

2.4. Digital Rights Management (DRM) Initiatives

A DRM system provides intellectual property protection so that only authorized users can access and use protected digital assets according to the rights expressions, which govern these assets.

Nowadays, there are several commercial initiatives that specify a complete DRM system. Moreover, there are standard initiatives that specify the elements that form a part of a DRM system and the relationships between them. Among the standard initiatives, the most relevant is the MPEG-21 standard that defines a framework for dealing with different aspects of multimedia information management. This standard has normatively specified the different elements and formats needed to support the multimedia delivery chain. In the different parts of this standard, these elements are standardised by defining the syntax and semantics of their characteristics, such as the interfaces to these elements.

In a DRM system rights expressions are defined to be the terms that govern the use of the digital assets. They are presented to the different actors of the digital value chain in the form of licenses expressed according to a Rights Expression Language (REL). RELs specify the syntax and semantics of a language that will be used to express the permissions and restrictions of use of a digital content. Licenses created according to a specific REL are associated to digital assets and can be interpreted and enforced by a DRM system.

2.4.1. Adaptation Authorisation

Adaptation operations should only be performed if they do not violate any condition expressed in the licenses. MPEG-21 DIA specifies description formats for *permissions and conditions for multimedia conversions* that are useful to determine which changes (adaptations) are permitted on content in view and under what kind of conditions.

3. Architecture Description

The proposed search and retrieval architecture for a complex and generic distributed multimedia scenario is depicted in Fig. 1. The core of the architecture is a multimedia content search module based on MPQF, which is elaborated in Section 3.1. Furthermore, the content retrieval service is augmented with a context-based content adaptation service based on MPEG-21 DIA and a DRM service based on MPEG-21 REL, which are discussed in detail in Section 3.2 and Section 3.3 respectively.

The proposed architecture is flexible and extensible not only because standards have been used, but also because of its modular structure. It allows many different types of multimedia content providers ranging from a simple metadata repository to an advanced provider such as Multimedia Content Provider #1 in Fig. 1. Furthermore, the application of video analysis algorithms for the automatic extraction of low level features

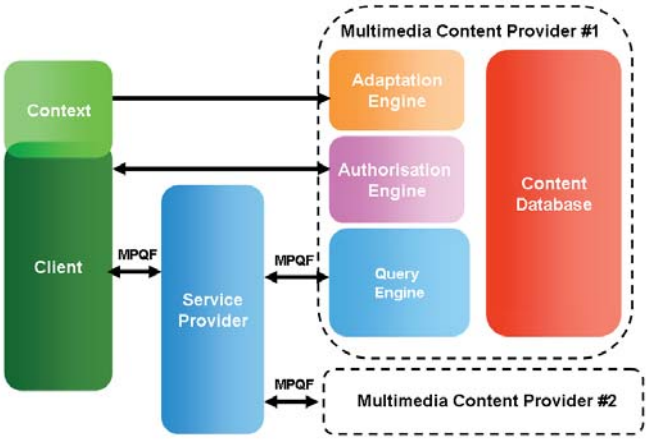

Figure 1. Proposed search and retrieval architecture for a complex and generic distributed multimedia scenario

could also be integrated in order to address the lack of metadata textual descriptors.

The users can transparently use the DRM engine and the adaptation engine to enrich the content retrieved from the content databases and the query engine. Furthermore these services can also be called by the user application, directly. Finally, the management of contextual information based on profiles (presented in Section 3.2) enriches content adaptation services, as well as content search and retrieval, whenever necessary; and therefore, will also be addressed. .

3.1. Multimedia Content Search and Retrieval Service

MPEG Query Format-Required search functionalities amongst the modules in the proposed architecture vary depending on their roles. On the one hand, service providers (e.g. content aggregators) need collecting metadata descriptions from content providers, and this is usually performed through a harvesting mechanism. Metadata harvesting consists on collecting the metadata descriptions of digital items (usually in XML format) from a set of digital content providers and storing them in a central server. Metadata is lighter than content, and therefore, it is feasible to store the necessary amount of metadata with the service provider, so that real-time access to information about distributed digital content becomes possible without the burden of performing a parallel real-time querying on the underlying target content databases. The search functionalities required for harvesting are very simple, because “harvesters” usually request information on updated records using a datestamp range. On the other hand, content “retailers”, which include service providers and also some content providers (generally medium or large scale providers), should be able to deploy value-added services offering fine-grained access to digital items, and advanced search and retrieval capabilities.

We have chosen the MPEG Query Format as the interface between parties in either of the two different situations described below. Although there exist mechanisms for metadata harvesting (e.g., Open Archives Initiative), MPQF can offer not only a similar functionality, but also a broad range of advanced multimedia search and retrieval capabilities. One of the key features of MPQF is that it is designed for expressing queries combining the expressive style of IR systems with the expressive style of XML DR systems (e.g., XQuery), embracing a broad range of ways of expressing user information needs. Regarding IR-like criteria, MPQF offers a broad range of possibilities that include, but are not limited to query-by-example-description, query-bykeywords, query-by-example-media, query-by-feature-range, query-by-

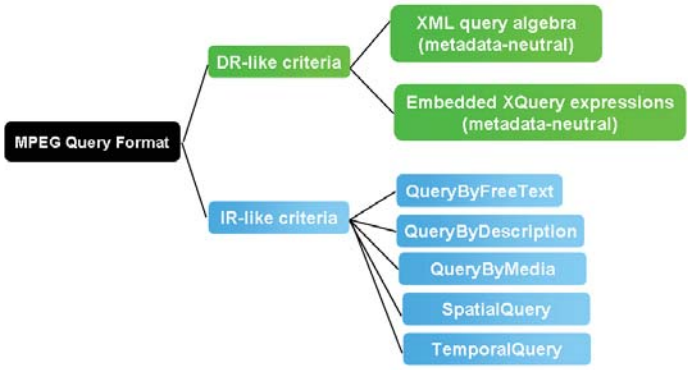

Figure 2. MPEG Query Format Outline

spatial- relationships, query-by-temporal-relationships and query-by-relevance-feedback. Regarding DR-like criteria, MPQF offers its own XML query algebra for expressing conditions over the multimedia related XML metadata (e.g., Dublin Core, MPEG-7 or any other XML-based metadata format) while also offering the possibility to embed XQuery expressions (see Fig. 2).

A valid MPQF document (according to the MPQF XML schema) always includes the Mpeg-Query element as the root element. Below the root element, an MPQF document includes the Input element or the Output element, depending on the fact that if the document is a client query or a server reply (there is only one schema and one root element). The part of the language describing the contents of the Input element is usually named as the Input Query Format (IQF), which it mainly allows specifying the search condition tree (Fig. 3) and also the structure and desired contents of the server output. The part of the language describing the Output element is usually named as the Output Query Format (OQF), and it specifies what the valid outputs are from the server to the client. IQF and OQF are used to facilitate understanding only, but do not have representation in the schema.

3.1.1. Multimedia Content Search and Retrieval Algorithms

A search and retrieval service able to respond to generic queries expressed with MPQF needs to be integrated with effective search and retrieval algorithms.

As stated above, technology that has proven to guarantee good performances in different use cases is the analysis of textual metadata (keywords, textual descriptions, plots, actors, user comments, etc.): there are several standards (like MPEG-7) describing information related to multimedia contents in a textual and interoperable format. In a real scenario, however, video content is not always accompanied by other corresponding information. There is the need, therefore, to provide innovative ways

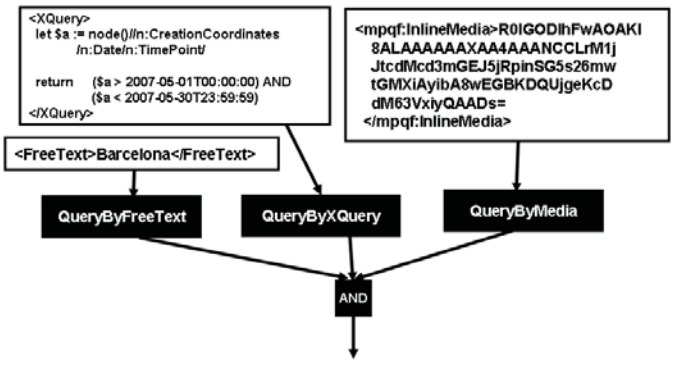

Figure 3. Example Condition Tree

to allow users to search for content exploiting all the available information. A solution can be identified in the automatic analysis of visual information and MPEG-7 represents a standard way to describe a set of low level features in an interoperable XML format. Nevertheless, one has to get over the above mentioned semantic gap. A way to address the problem and improve the overall performances, still maintaining the standard compatibility is to analyse MPEG-7 descriptors related to low level features jointly with textual metadata, whenever available.

The technical work conducted in this activity can be described as a sequence of different operations:

- Automatic extraction of low level features: for each video, a set of MPEG-7 descriptors is extracted; the obtained low level features are then processed in order to extract temporal and spatial features related to the whole content. The latter are represented by the MPEG-7 descriptors themselves, providing information about visual aspects. Each frame is associated to an element of a codebook made clustering the MPEG-7 descriptors extracted from a training set of videos, using the Generalized Lloyd Algorithm (GLA). A probability distribution related to the whole video is then computed. This distribution describes the visual aspects of the video.
- Starting from the analysis of MPEG-7 descriptors it is also possible to obtain information about the temporal evolution of the videos, revealing, therefore, aspects of storytelling style. After having compressed the total amount of data through the Singular Value Decomposition (SVD), we have chosen to develop new low level features that reflect the complexity of the temporal evolution of the principal components extracted by the data set: the spectral flatness and the fractal dimension. This is because the amount of change among frames can be inferred directly from the variation of the first few coefficients.
- The visual and temporal features undergo a fusion process to compute the overall similarity among contents;
- Analysis of textual information: Latent Semantic Indexing (LSI) technique has been used, a vector space technique that exploit co-occurrences between terms. Using LSI it is possible to discover similarities between texts even if they share few or no words;
- Construction of searchable indexes: The data extracted with textual and visual analysis are used jointly for creating tables of distances between contents in the repository. Such tables can be used in real time to provide answers to different kind of queries.

For further information about these algorithms one can refer to [8].

Once the searchable indexes are constructed, we can consider that this Content Provider (CP) includes not only the Content database (Fig. 1), but also these processing techniques and the associated database with the similarity measures between contents. When contents lack of metadata, or even for a more accurate result based on a user's Relevance Feedback (RFB), or combined with metadata, QueryByExample (QBE) is an interesting retrieval approach that needs to be addressed.

We already stated, at the beginning of this section, that MPQF is capable of expressing different types of queries for IR systems, as for example, query-byexample-media In Fig. 4 and Fig.5 we can see an example of the InputQuery and the OutputQuery, respectively, that could be used between the Service Provider (SP) and the CP previously identified. On the one hand, the

request includes the sample of content that is used to express user's interest and the description of the desired output (number of results, etc.). On the other hand, the response includes the list of the "most similar" items that have been retrieved.

Moreover, we could consider these messages as the ones exchanged between the user and the SP in a more specific application scenario.

3.2. Context-based Content Adaptation Service

In a complex multimedia scenario, many different adaptation operations may be performed on contents. As mentioned in Section 2, MPEG-21 DIA is a complete standard that specifies the syntax and semantics of tools that assist in the adaptation of multimedia content. It is used to satisfy transmission, storage and consumption constraints as well as Quality of Service management. The proposed context-based adaptation service for search and retrieval application is based on MPEG-21 DIA; a more detailed specification of a possible modular architecture of this service can be found in [9].

Due to the inherent complexity of the standards, their practicality in video search and retrieval domain is slim. In order to address this issue, a number of context profiles are defined. These profiles include User Profile, Network Profile, Terminal Profile, and Natural Environment Profile. They contain all the associated descriptors of MPEG-21 UED, and thus, cover a complete set of contextual information. A detailed description of them can be found in [9]. Furthermore, they could be extended to new types of context that could be identified by new sensors.

The use of profiles eases the introduction of standards while providing more flexibility and scalability to any architecture. In

```
<MpegQuery xmlns="urn:mpeg:mpqf:schema:2008" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpqf:schema:2008 mpqfv08.xsd" mpqfID="http://www.mpqf.org/id1">
  <Query>
    <Input>
      <OutputDescription maxItemCount="3" thumbnailUse="true">
        <SortBy xsi:type="SortByFieldType" order="decreasing">
          <Field>similarity_index</Field>
        </SortBy>
      </OutputDescription>
      <QueryCondition>
        <Condition xsi:type="QueryByMedia">
          <MediaResource resourceID="example">
            <MediaResource>
              <MediaUri>urn:frame:example</MediaUri>
            </MediaResource>
          </MediaResource>
        </Condition>
      </QueryCondition>
    </Input>
  </Query>
</MpegQuery>
```

Figure 4. Example of an MPQF QueryByMedia Input Query

```
<MpegQuery xmlns="urn:mpeg:mpqf:schema:2008" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpqf:schema:2008 mpqfv08.xsd" mpqfID="http://www.mpqf.org/id1">
  <Query>
    <Output>
      <ResultItem xsi:type="ResultItemType" recordNumber="1" confidence="0.9">
        <Thumbnail>urn:thumbnail:frame1</Thumbnail>
        <MediaResource>urn:frame1</MediaResource>
      </ResultItem>
      <ResultItem xsi:type="ResultItemType" recordNumber="2" confidence="0.8">
        <Thumbnail>urn:thumbnail:frame2</Thumbnail>
        <MediaResource>urn:frame2</MediaResource>
      </ResultItem>
      <ResultItem xsi:type="ResultItemType" recordNumber="3" confidence="0.7">
        <Thumbnail>urn:thumbnail:frame3</Thumbnail>
        <MediaResource>urn:video3</MediaResource>
    </ResultItem>
    <SystemMessage>
      <Status>
        <Code>001</Code>
        <Description>Query was successful</Description>
      </Status>
    </SystemMessage>
  </Output>
  </Query>
</MpegQuery>
```

Figure 5. Example of an MPQF QueryByMedia Output Query

our proposal, their use would definitely enrich the search and retrieval service while guaranteeing interoperability. Not only the user can identify the content he/she wants, but also will receive it in the most optimized way thanks to the Context-based Content Adaptation Service

3.3. Digital Rights Management Service

The DRM Management service will ensure that multimedia copyrighted content is used according to the terms stated by content creators or rights holders. This service will inform to the user the operations that he/she can perform with the videos found by the audiovisual search and retrieval service. It provides functionalities to obtain the licenses governing a digital resource, in this case a video, and provides information content usage information, according to licenses governing the selected video, to the user. Then, the user will select the operation he/she wants to perform, and if necessary will purchase the appropriate license.

This service provides two operations: The first one obtains the licenses associated to the video selected by the user and the second one determines the user's permissions and constraints of content usage. Next are described the two operations in detail:

- *getLicenses*: It receives as parameters a video label and returns an XML file containing the set of MPEG-21 REL licenses governing the image. These licenses will specify the rights that the user can exercise and the conditions that he previously has to fulfil. Moreover, it also returns the rights that the use could exercise if he previously purchases the appropriate license.
- *verifiyRights*: It receives as parameters the user's licenses governing the video and an XML file containing information about the usage that the user has previously done with this concrete video, and determines if the user can exercise the requested operation. This operation implements a license verification algorithm, based on the MPEG-21 REL Authorization Model [x], which verifies if an entity was authorized to perform the requested operation over a video.

3.3.1. Adaptation Authorization

In order to govern content adaptations for protected contents, licenses integrating MPEG-21 REL and MPEG-21 DIA should also be used. The main reason for integrating both standards is that, due to the increasing complexity of adaptations, we require more detailed descriptions about content adaptation in order to govern them. A detailed work on the adaptation authorisation can be found in [10].

4. Social Networks application scenario

Recently, online social networking sites are experimenting a dramatic growth in their use. Users of these sites form social networks and share contents (photos, videos, etc.) and personal contacts. In certain occasions users can wish to protect their personal information and the contents they share for privacy issues. Online social networking sites can be accessed from a broad diversity of devices (PDAs, mobile phones, PCs, laptops, etc.) and in different network conditions (fixed, mobile, local area, wide area, etc), then an efficient adaptation of the contents is required. Due to the huge amount and diversity of contents shared by users, online social networking sites also require efficient search and retrieval solutions. Furthermore, these solutions also can take advantage of automatically extracted features and textual metadata.

In online social networking sites, users will benefit of the proposed solution, since service providers can collect metadata descriptions from content providers and content retailers should be able to deploy value-added services offering fine-grained access to digital items, and advanced search and retrieval capabilities. Moreover, access to these sites can be done from a broad range of consumer electronic devices, since the framework provides efficient and useful adaptations. Finally, users will be able to protect their personal information, contacts as well as the contents they provide to other users.

5. Conclusions and Future Work

This paper has presented a standards-based generic approach for complex multimedia management.

First of all, several weaknesses when dealing with context and metadata for multimedia management have been identified. The novelty of our proposed solutions comes from the fact that these problems are addressed in the most generic way, as the authors consider it is the best approach to exploit the maximum potential of both types of descriptors.

On the one hand, a flexible and extensible way of representing context is used to enrich content adaptation, content search, and Digital Rights Management. On the other hand, the use of a flexible language for multimedia search and retrieval based on MPEG Query Format is the key point trying to narrow the semantic gap, as it gives all the required functionalities. Furthermore, the lack of metadata textual descriptors has been addressed by integrating video analysis algorithms for the automatic extraction of low-level features with this flexible language for multimedia search and retrieval. The use of standards, such as MPEG-21 DIA, MPEG-21 REL, MPEG-7 and MPQF, is also mandatory to guarantee interoperability with similar systems.

We will continue working on the instantiation of the approach presented in this paper in an ongoing project, the XAC2 project (sequel of XAC, Xarxa IP Audiovisual de Catalunya, Audiovisual IP Network of Catalonia); XAC2 is a network for digital assets interchange among TV channels and content producers. It is worth mention that it is expected that from this work it will emerge the first known implementation of an MPEG Query Format processor. Currently, parts of the ongoing implementation are being contributed to the MPEG standardisation process in the form of Reference Software modules.

Acknowledgments

This work has been partially supported by the Spanish government through the project MCM-LC (TEC 2008-06692-C02-01)

References

[1] Wang, J.-G., Kim, S.-F., Chang, Kim, H.-M. (2007). Utility-Based Video Adaptation for Universal Multimedia Access (UMA) and Content-Based Utility Function Prediction for Real-Time Video Transcoding, *IEEE Trans. Multimedia,* 9 (2) 213220, February 2007.

[2] ISO/IEC/SC29/WG11/N9341. (2007). *ISO/IEC 15938-12 FCD MPEG Query Format*, October.

[3] Xu, L.L., Li, Y. (2003). Video classification using spatialtemporal features and PCA. In *Proceedings of the 2003 international Conference on Multimedia and Expo -Volume 3 (ICME '03) -Volume 03* (July 06 -09, 2003). ICME. IEEE Computer Society, Washington, DC, 485-488.

[4] Sorvari, A., Jalkanen, J., Jokela, R.,. Black, A., Koli, K., Moberg, M. , Keinonen, T (2004). Usability issues in utilizing

context metadata in content management of mobile devices", *In*: Proc. of the third Nordic conference on Human-computer interaction (NordiCHI'04), Tampere, Finland.

[5]. Aktas, M. S., Fox, G. C., Pierce, M. (2005). Managing Dynamic Metadata as Context, *In*: Proc. Of the 2005 Istanbul Internacional Computational Science and Engineering Conference (ICCSE2005), Istanbul, Turkey, 27-30 June.

[6] Wallace, M., Akrivas, G., Mylonas, Ph., Avrithism, Y., Kollias, S. (2003). Using Context and Fuzzy Relations to Interpret Multimedia Content, *In:* Proc. of the Third International Workshop on Content-Based Multimedia Indexing (CBMI), Rennes, France, Sep.

[7] Dey, A. K. (2000). Providing MPEG Query Format, ISO/IEC 15938-12:2008, December 2008.

[8] IST-1-038398 -Networked Audiovisual Media Technologies -VISNET II. (2007). Deliverable D2.2.5: First set of developments and evaluation for search systems for distributed and large audiovisual databases. November.

[9] Andrade, M. T., Kodikara Arachchi, H., Nasir, S., Dogan, S., Uzuner, H., Kondoz, A. M., Delgado, J., Rodríguez, E., Carreras, A., Masterton, T., Craddock, R (2007). Using context to assist the adaptation of protected multimedia content in virtual collaboration applications, *In: Proc. IEEE Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2007),* New York, USA, 12-15 Nov.

[10] Carreras, A., Delgado, J. (2008). A new type of contextual information based on the adaptation authorisation, *In:* 9th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2008), Klagenfurt, Austria, 7-9 May.

Authors biographies

Anna Carreras obtained her B.Sc. degree in Telecommunication Engineering in 2006 and her M.Sc. degree in Computer Architectures, Networks and Systems in 2008, both from Universitat Politècnica de Catalunya (UPC) in Barcelona (Spain). Member of the Distributed Multimedia Applications Group (DMAG) since 2005. Her research interests include Digital Rights Management, and the use of context and metadata in distributed multimedia applications.

Ruben Tous received his Ph.D degree in Computer Science and Digital Communication from Universitat Pompeu Fabra, Barcelona, Spain, in 2006. Since 2006 he is a researcher at Distributed Multimedia Applications Group of the Department of Computer Architecture of Universitat Politècnica de Catalunya, and an Assistant Professor in the same department. He is an expert for the Asociación Española de Normalización y Certificación (AENOR) and has been participating as spanish delegate in ISO/MPEG and ISO/JPEG. His research interests include semantic-driven multimedia indexing and retrieval, knowledge representation and reasoning for multimedia understanding (multimedia ontologies) and semantic alignment.

Eva Rodríguez obtained her Ph.D. degree in Computer Science in 2007 and her B.Sc. degree in Telecommunication Engineer in 2001. She is an Associate Professor at the Department of Computer Architecture in the Universitat Politècnica de Catalunya (UPC). Member of the Distributed Multimedia Applications Group (DMAG) since 2002. Participant in several IST projects in the areas of multimedia publishing, electronic commerce and distributed applications. Her research interests include Digitals Rights Management, content protection, MPEG-21 Intellectual Property Management and Protection, and multimedia framework. Active participation, since 2003, in MPEG-21 standard. Author of several published papers in international journals and conferences.

Jaime Delgado received his Ph.D. degree in Telecommunication Engineering in 1987. Since September 2006, Professor at the Computer Architecture Department of the Universitat Politècnica de Catalunya (UPC) in Barcelona (Spain). Previously, Professor at the Universitat Pompeu Fabra (UPF), since 1999. Head and founder of the Distributed Multimedia Applications Group (DMAG).

Project Manager of several European and national research projects. Active participation, since 1989, in International standardisation, as editor of standards and chairman of groups. Evaluator and reviewer for the European Commission since 1989. Advisor for the Spanish Ministry of Science. Author of several hundreds of published papers and books.

Giovanni Cordara graduated in Telecommunications Engineering from the Polytechnic of Turin, in October 2000. In 2001 he joined Telecom Italia LAB, the corporate research centre of Telecom Italia, where he was originally involved in the design and implementation of MPEG-4 solutions for interactive Television and E-learning. In 2003, he was involved in the design of a DRM platform for the protection of Premium contents distributed over heterogeneous networks and terminals. Since 2003 he has been involved in the image/video processing field, carrying out research activities related to video coding, automatic multimedia analysis, advanced search and retrieval.

Gianluca Francini received his degree in Computer Science from the University of Torino, Italy, in 1994. In 1996 he joined the Multimedia research group of CSELT and received his Master in Telecommunication. He is a senior researcher in the Telecom Italia group. His research interests include computer vision, image and video retrieval, 3D reconstruction and recommendation systems.

Diego Gibellino graduated in Computer Science Engineering from the Polytechnic of Turin, in December 2000. In 2001 he joined Telecom Italia LAB, the corporate research centre of Telecom Italia, where he was involved in the design and implementation of Telecom Italia's MPEG-4 based Media Platform, focusing on the media coding architecture. He has then been involved in the design and implementation of enhanced Video Communication embedded systems based on MPEG-4/AVC video coding standards. He is currently involved the definition of the next generation IPTV services and IPTV platform for Telecom Italia.