Enriching search and retrieval of multimedia information with DRM services and context-based content adaptation

Anna Carreras, Ruben Tous, Eva Rodríguez, Jaime Delgado
Universitat Politècnica de Catalunya, Campus Nord, Mòdul D6, Jordi Girona 1-3, E-08034 Barcelona, Spain
annac@ac.upc.edu

Giovanni Cordara, Gianluca Francini, Skjalg Lepsoy
Telecom Italia Lab, Via Reiss Romoli, 274 – 10148 Torino, Italy
giovanni.cordara@telecomitalia.it

Safak Dogan, Hemantha Kodikara Arachchi
I-Lab, CCSR, University of Surrey, Guildford, GU2 7XH Surrey, UK
S.Dogan@surrey.ac.uk

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. The most innovative part of the proposed work comes from the integration of 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 low-level features.

1. Introduction

The increasing complexity of today’s multimedia scenarios is leading to an overflow of new application-specific technologies with a common objective: to enable Universal Multimedia Access (UMA) while enriching user experience. Even if a substantial amount of work has been done in this area, it is possible 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, 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 analysis of these textual descriptors has been proven to guarantee good performances in different use cases, but in real scenarios they are not always available.

• 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. This is the most difficult problem to solve since it is intrinsically embedded in the nature of video digital contents and strictly related to human interpretation.

• The use of standards. A common standardized representation of metadata and context, and the definition of unified languages and interfaces to accept and respond to requests for multimedia content searches would facilitate the interoperability between different multimedia systems.

Furthermore, the affordability of consumer electronic devices such as MP3 and recordable players, digital cameras, etc., allow users to also become content producers as well as consumers, creating new and very interesting challenges from the perspective of enforcing Digital Rights Management (DRM) during content access and consumption.

Following the background information about the different lines of research activities integrated in the theme of the proposed work (Section II), 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 III).
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) [1] as the most suitable solution as it satisfies those requirements. Finally, we will also refer to different search and retrieval 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 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-by-example (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 it is expected that MPQF will become 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 standardizing 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 algorithms.

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: their main disadvantage relies however 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 [2]. 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 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.

2.3 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. There are two other initiatives, OMA (Open Mobile Alliance) and TV-Anytime, which are specifications from the Industry Fora, and intend to be standards. However, they are not official since they have not been produced by Standards Organisations.

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.

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 III-A. 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 III-B and Section III-C respectively.

![Figure 1. Proposed search and retrieval architecture for a complex and generic distributed multimedia scenario](image-url)
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, based on profiles elaborated in Section III-B, a flexible way of managing contextual information so as to enrich content adaptation services, as well as content search and retrieval, whenever necessary, is also presented.

3.1 Multimedia content search and retrieval service

3.1.1 MPEG Query Format. Required search functionalities amongst the modules in the proposed architecture vary depending on their roles. On 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 timestamp 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 (detailed in Section II) as the interface between parties in either of the two different situations described below. Though 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-media, query-by-feature-range, query-by-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).

![MPEG Query Format outline](image)

3.1.2 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 to allow users to search for content exploiting all the available information. A solution can be identified in the automatic analysis of visual information: nevertheless, one has to get over the abovementioned semantic gap.

A way to address the problem and improve the overall performances is to analyze 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, temporal and spatial features are extracted, the former related to the temporal evolution of the videos, revealing therefore, aspects of storytelling style; the latter, represented by a set of MPEG-7 descriptors, provides information about visual aspects;
- Analysis of textual information: Latent Semantic Indexing (LSI) technique has been used [3], 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 one can refer to [4].

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 II, MPEG-21 DIA is a complete standard that specifies the syntax and semantics of tools that assist in the adaptation of multimedia content.

In our system, contextual information would be expressed on MPEG-21 DIA UED profiles [5] and would be gathered directly by the Adaptation Engine. When a user selected a video, the adaptation operation would be performed based on context. Furthermore, it would first request the authorisation of this adaptation to the Digital Rights Management Service [6].

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.

4. Conclusions and future work

This paper has presented a standards-based architecture for search and retrieval in complex multimedia scenarios. To the best of our knowledge, this is the first piece of work trying to integrate MPQF with context-based adaptation and DRM.

The lack of metadata textual descriptors has been addressed by integrating video analysis algorithms for the automatic extraction of low-level features with a flexible language for multimedia search and retrieval based on MPQF.

In line with this study, we have proposed a context-based content adaptation service based on MPEG-21 DIA and a DRM service based on MPEG-21 REL to enrich the search and retrieval framework.

5. References