How MPEG Query Format enables advanced multimedia functionalities

Anna Carreras, Ruben Tous, Jaime Delgado

Distributed Multimedia Applications Group,
Universitat Politècnica de Catalunya,
c/Jordi Girona, 1-3, Mòdul D6, 08034 Barcelona, Spain
{annac, rtous, jaime.delgado}@ac.upc.edu

Abstract: In December 2008, ISO/IEC SC29WG11 (more commonly known as MPEG) published the ISO/IEC 15938-12 standard, i.e. the MPEG Query Format (MPQF), providing a uniform search&retrieval interface for multimedia repositories. While the MPQF's coverage of basic retrieval functionalities is unequivocal, it's suitability for advanced retrieval tasks is still under discussion. This paper analyzes how MPQF addresses four of the most relevant approaches for advanced multimedia retrieval: Query By Example (QBE), Retrieval trough Semantic Indexing, Interactive Retrieval, and Personalized and Adaptive Retrieval. The paper analyzes the contribution of MPQF in narrowing the semantic gap, and the flexibility of the standard. The paper proposes several language extensions to solve the different identified limitations. These extensions are intended to contribute to the forthcoming standardization process of the envisaged MPQF's version 2.

1 Introduction

In today's Multimedia Information Retrieval (MIR) systems, one of the main concerns is how to bridge the semantic gap between the machine-level audio-visual feature descriptors and the semantic-level descriptors directly interpretable by humans. The algorithms currently available in literature are not yet sufficient to assure good results, exploitable in commercial solutions. Of course, the problem arisen by the semantic gap is really difficult to solve since it is intrinsically embedded in the nature of digital contents and strictly related to human interpretation (for example, a picture of a beach at the sunset could be categorized as "sea" or "sunset", according to the mood and the sensitivity of the user). In this paper, we will try to address this issue from two different points of view.

First, from the “content provider side”, we will focus on two main retrieval approaches. On the one hand, we will consider QBE, which involves using an example of content to illustrate users' needs (Section 3.1). QBE is one of the most matures approaches for multimedia retrieval and it is based on similarity measures of specific Low Level Features (LLF) that have already been proved to give interesting results [Lux09]. On the other hand, the use of Semantic Indexing will also be addressed (Section 3.2). In this case, links between text-based search terms and semantic extracted descriptors need to be established; although this is a more recent area of research, a lot of work is currently being done on the automatic extraction of these descriptors using complex machine learning and pattern classification techniques.

Second, more related to the subjective perception of the user than to the nature of the digital content, the use of interactive retrieval based on Relative FeedBack (RFB) is the third multimedia retrieval approach that we are going to consider in this paper (Section 3.3). Finally, and because nowadays it is not enough to identify the right content but it is required to be presented in the most suitable way to the user, personalized and adaptive content retrieval will also be addressed in Section 3.4.

In the following section will briefly present the novel MPQF standard [MPQF07] as a possible unified query language. Subsequently, we will identify and separately evaluate its possible application in the four retrieval approaches previously identified: QBE, Semantic Indexing, Interactive Retrieval, and Personalized and Adaptive Content Retrieval. We consider that these four approaches adequately represent today's multimedia scenarios as they cover a broad part of the most relevant work done in this area of research.

Thus, in this paper we will not intend to present a survey on Multimedia Search and Retrieval (interesting works can also be found in [Ha08]), and neither describe the MPQF standard (as in [Dö08]); but analyze its use in some of the most relevant retrieval approaches.

2.1 MPEG Query Format Overview

The MPEG standardization committee (ISO/IEC JTC1/SC29/WG11) has developed a new standard, the MPEG Query Format (MPQF) [MPQF07], which aims to provide a standardized interface to multimedia document repositories. MPQF is an XML-based language which defines the format of queries and replies to be interchanged between parties in a multimedia information search and retrieval environment. MPQF can be used in standalone MMDBs, but it has been specially designed for scenarios in which several MMDBs and content aggregators interact (Fig. 1). Furthermore, MPQF does not make any assumptions about the metadata formats used by the target MMDBs, which can be MPEG-7 but also any other format (Dublin Core for example).

MPQF allows combining Information Retrieval (IR) criteria with Data Retrieval (DR) criteria. Regarding IR-like criteria, MPQF offers a broad range of possibilities that include but are not limited to Query By Example, Query By Feature Range, Query By Spatial or Temporal Relationships, and Query By Relevance Feedback. Regarding the DR criteria, MPQF offers its own XML query algebra, but also offers the possibility to embed XQuery expressions.

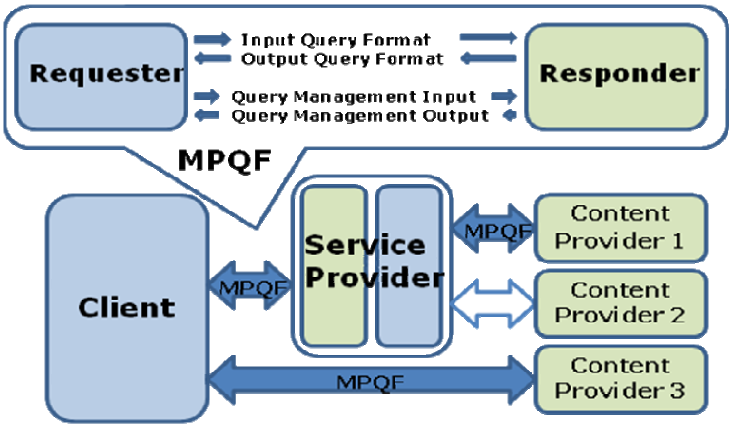

Fig. 1. Possible scenario of use of MPEG Query Format

3 Advanced Multimedia Functionalities with MPQF

3.1 MPQF for Query By Example

The main objective of this section is to report and evaluate the use of MPQF for QBE retrieval. The first thing we need to take into account is that there exist many different ways of implementing QBE algorithms. The detailed analysis of these algorithms is out of scope of this paper, nevertheless, a detailed publication about the QBE algorithms that have been used to raise our conclusions can be found in [Vis07]. In general terms, these algorithms can be based on different types of Low Level Features (LLF) more or less suitable depending on the multimedia content type. For example, when working with entire videos, Temporal LLF and LLF Distributions should be used, while, *ColourStructure* and *HomogeneousTexture* descriptors would be more useful when dealing with image frames. Furthermore, depending on the computation power or even the storage capacity of the system, it would also involve different types of pre-processing techniques along the retrieval process.

In this sense, it is important that service providers are able to query for desired capabilities, and that in turn, content providers are able to communicate their capacities to the service provider. This issue is suitably addressed by the MPQF through Query Management tools which include service discovery, querying service capability, and service capability description (see Query Management Input and Output in Fig. 1).

Moreover, depending on the application scenario, different types of QBE may be required. For example, in a Video Surveillance application scenario it would be interesting to detect similar faces (Query By Region Of Interest), while a user browsing movies similar to his/her favorite ones may require a completely different QBE algorithm (Query By Temporal or Spatial Similarity).

According to ISO-15938-12:2008, the technique of QBE is understood as the combination of different condition expressions such as *QueryByMedia*, *QueryByDescription*, *QueryByROI, SpatialQuery and TemporalQuery*. All these MPQF's condition types are based in the provision of an example (media, media region or media metadata description) in order to express the user information need. These condition types are selected or combined depending on each situation in order to return the best results.

QBE similarity searches are techniques of *content based multimedia retrieval* (CBIR, etc.) which allow expressing the user information need with one or more example digital objects (e.g. an image file). Even though the usage of low-level features description instead of the example object bit stream is also considered QBE, in MPQF these two situations are differentiated, naming *QueryByMedia* (or *QueryByROI*) to the first case (the digital media itself) and *QueryByDescription* the second one. This differentiation is important because in the first case is the query processor who decides which features to extract and use, and in the second case is the requester who perform the feature extraction and selection. In this work we will focus on the first one, as we consider that the *QueryByDescription* is sufficiently well addressed in [Dö08].

The MPQF's *QueryByMedia* type offers multiple possibilities to refer to the example media, as just including the media identifier (a locator such as an URL pointing to an external or internal resource) or directly embedding the image bit stream in Base64 encoding within the XML Query. When the *QueryByMedia* type is used, it's up to the query processor to extract the proper low-level features to perform a similarity search over the index. One of the limitations identified in this work is that MPQF does not standardise a set of parameters or algorithms to be used, leaving this totally open with the consequent lack of interoperability. One possibility could be using MPEG-7 descriptors such as *ScalableColor*, *ColorLayout* or *EdgeHistogram*. The standard should allow expressing different weights to each one of the different descriptors in order to tune the similarity algorithm. Currently these weights are sent to the query processor within non-standard attributes in the MPQF query. The inclusion of non-standard parameters is allowed in MPQF.

Overall, we can conclude that the MPQF offers the necessary tools for performing effective QBE, while maintaining the system network agnostic and media agnostic. Furthermore, it covers all the possible application scenarios we could think of. However, we consider a limitation the fact that MPQF does not standardise a set of parameters or algorithms to be used, leaving this totally open. Currently this information is sent to the query processor within non-standard attributes in the MPQF query, which severely constraints query interoperability.

3.2 MPQF for Retrieval through Semantic indexing

As introduced in Section 2, a MIR system has the particularity that it must combine Information Retrieval (IR) techniques, with techniques for querying metadata, which belong to the Data Retrieval (DR) area within the Databases discipline. Though there is a solid research basis regarding the Information Retrieval challenge, the necessity to face such problem appears, in fact, because of the difficulty of annotating the content with the necessary metadata and the difficulty of formalizing the end-user's semantic-level criteria. As a result, from the multimedia retrieval point-of-view, measures are needed to deal with uncertainty and the potential lack of search precision. However, in a vast number of scenarios, simple IR-like mechanisms like keywords-based search use to offer pretty satisfactory results even when the size of the target collections is big. There are, nevertheless, situations in which the end-user requirements, and/or the circumstances, motivate the efforts of producing higher-level semantic metadata descriptors and formalizing parts of the user's semantic-level criteria moving them to the Data Retrieval realm. An example could be the video surveillance scenario, in which a huge quantity of information is stored, and the query expressiveness and results precision are critical. This formalization task requires enhancing the metadata production layer but also implies offering to the user a richer interface or, in subsequent layers, post-processing the initial non-formalized query. This enrichment of the querying process is related to the improvement of the metadata-level query capabilities. The result is the starting point of what is known as semantic-driven MIR, whose evolution leads to the usage of semantic-specific technologies as those from the Semantic Web initiative.

Current practices in the metadata community show an increasing usage of Semantic Web technologies like RDF and OWL. Some relevant initiatives are choosing the RDF language (e.g. Dublin Core) for modelling semantic metadata because of its advantages with respect to other formalisms. RDF is modular; a subset of RDF triples from an RDF graph can be used separately, keeping a consistent RDF model. So it can be used in presence of partial information, an essential feature in a distributed environment. The union of knowledge is mapped into the union of the corresponding RDF graphs (information can be gathered incrementally from multiple sources).

As introduced in Section 2.1, MPQF is an XML-based language in the sense that all MPQF instances (queries and responses) must be XML documents, i.e. it has an XML serialization format. However, this fact is independent of the target metadata data model. Initially MPQF was designed to only address XML-enabled databases, Formally, MPQF is Part 12 of MPEG-7, which is an XML application, and at the very beginning MPQF was meant to target MPEG-7 repositories. Nevertheless, soon the query format was technically decoupled from MPEG-7 and became metadata-neutral, i.e. MPQF is not coupled to any particular metadata standard. However, the final standard (12/2008) still assumed that queries refer to metadata, at a logical level, as XML trees. The *EvaluationPath* element is probably the most important part of the standard as it identifies the results of the query based on the selected "branch" of this tree. Thus, MPQF expresses conditions and projections over the metadata using XPath expressions, i.e. privileging XML-enabled metadata repositories but restraining those based in other models, especially those based in RDF metadata.

This limitation was already identified in [TD08], and subsequently, and amendment to the MPQF entitled “Semantic Enhancement” [Amd08] was initiated during the 88th MPEG meeting (April 2009), and will be probably finalized during the next meeting (90th MPEG meeting, October 2009). This amendment is the necessary extension to allow the MPQF not only to manage metadata modelled with Semantic Web languages like RDF and OWL, but also to query constructs based on SPARQL.

3.3 MPQF for Interactive Retrieval

When retrieving multimedia content, an important issue that needs to be considered is the subjective perception of the user. Through the use of Relevance Feedback (RFB), the query is refined over stages in which the user indicates which retrieved examples match or do not match the user’s need. Based on this feedback, the system modifies its retrieval mechanism in an attempt to return a more desirable instance set to the user.

Once again, depending on the application scenario, the interaction between the user and the system may be different. For example, while a doctor could be very patient to find the most similar medical image within a database in order to make a diagnostic, a user browsing multimedia content in the web would be bothered in the early stages of the interaction.

The MPQF specifies the *QueryByRelevanceFeedback* type which describes a query operation that takes the result of the previous retrieval into consideration. It contains two elements: the *answerID* which identifies the result set where the relevance feedback should be performed; and the *ResultItem* which identifies the good examples that will be used as input for the next query. Althought it is also possible to discard bad results by combining the boolean NOT with the Query By Relevance Feedback operation, we miss the possibility of scoring the results. The MPQF offers the possibility of weighting the query conditions combined within the query “tree”, but it would be also interesting to score the different elements of the list of results.

We believe the MPQF is a little bit too simplistic when addressing interactive retrieval as it only allows distinguishing between good and bad results, while it would be much more interesting to know which has or have been the most relevant result/s in order to refine the query. Of course, this should be only an optional attribute suitable for some specific domains or application scenarios.

Nevertheless, we miss an important element we have pointed out earlier in this section: the number of iterations. We believe that the user (or even the Service Provider in some scenarios) should be able to specify the number of iterations she/he is going to perform beforehand, as this would facilitate the application of the most effective matching algorithms.

3.4 MPQF for Personalized and Adaptive Content Retrieval

The main idea on personalized and adaptive retrieval is to use contextual information (from the user or usage environment) in order to provide effective multimedia information retrieval. Of course, this can be considered under the big umbrella of context-awareness area of research. On the one hand, user preferences can help in the identification of retrieved multimedia content, and on the other hand, information about the characteristics and capabilities of the terminal, the network or the natural environment may be used to improve the user's Quality of Experience (QoE) by adapting the content efficiently.

MPQF allows expressing few preferences on the presentation of multimedia content results set. This is done through the *OutputDescription* descriptor included in the Input Query. Nevertheless, it only specifies few listing and sorting options that could easily be extended. We believe personalization is a complex multimedia retrieval service, and as such, it should be considered in the MPQF management tools first. The management part of the MPQF copes with the task of searching for and choosing desired multimedia services for retrieval. This part includes service discovery, querying for service capabilities, and service capability descriptions. We miss the possibility of detecting and selecting a context-aware adaptation service. The MPQF standard can detect services such as authentication, or billing, but does not include context-aware services.

For example, a content provider may offer an integrated service including multimedia contents and the adaptation service. The delivery of most of the contextual information could be done in a separate channel than the query itself as proposed in [ETH09], but probably it would be more useful to integrate the user preferences inside the input query. This could be done by specifying a new query type named "*QueryByUserPreferences*", or even "*QueryByUserContext*" if we think of extending the user preferences with the user historical data for example. Of course, different standards representing contextual information (i.e. MPEG-21, UAProf, etc.) could be used, in the same way MPQF is metadata neutral. Another possibility would be to include this information on the Output Description element.

4 Conclusions

This paper has presented an analysis of the MPQF standard in four relevant areas of research within multimedia search and retrieval applications, namely, query by example, query by semantic indexing, interactive retrieval, and personalized/adaptive retrieval. We can conclude that the first one, QBE, is well addressed, but we consider a limitation the fact that MPQF does not standardise a set of parameters or algorithms to be used. Currently these data are sent to the query processor within non-standard attributes in the MPQF query, which severely constraints query interoperability.

The other three retrieval approaches require further extensions of the standard in order to fully exploit today's application scenarios (the second one is already being addressed trough an Amendment). Probably, the detected limitations are due to the fact that the editors of MPQF have tried to maintain a quite simple standard in order to potentiate its use within the research community. Nevertheless, we believe the specification of MPQF profiles for concrete application scenarios could help to further develop the parts that have been identified as too simplistic, such as personalization, and interactive retrieval.

Finally, it would be very interesting to give the opportunity to the users (or even to the Service Provider) of deciding whether they allow or not the use of advanced retrieval functionalities, such as personalization, or semantic indexing. Usually, these kinds of techniques involve the use of personal information (previous queries, preferences, etc.) that the user may want to protect.

All the identified limitations and proposed solutions are intended to contribute to the forthcoming standardisation process of the envisaged MPQF's version 2. And of course, some evaluation work will be done as soon as a finalised version of a software module based on MPQF exists, which for the moment is not the case.

Acknowledgements

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

References

[Lux09] Lux, M.; Caliph & Emir: MPEG-7 Photo Annotation and Retrieval; Winer of the 2009 ACM Multimedia Open Source Software Competition, October, 2009, Beijing, China. ACM 978-1-60558-608-3/09/10.

[MPQF07] Gruhne, M., Tous, R., Doeller, M., Delgado, J., Kosch, H.; MP7QF: An MPEG-7 Query Format. 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS 2007), Barcelona, November 2007. IEEE Computer Society Press. ISBN 0-7695-3030-3. p. 15-18.

[Ha08] Hanjalic, A., Lienhart, R., Ma, W.-Y., Smith, J. R.; The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?; IEEE Proceedings, Special Issue on Multimedia Information Retrieval, vol.96, no.4, pp. 541-554, 2008.

[Dö08] Döller, M., Tous, R., Gruhne, M., Yoon, K., Sano, M., Burnett, I. S.; The MPEG Query Format: Unifying Access to Multimedia Retrieval Systems; IEEE Multimedia, vol. 15, no.4, pp. 82-95, 2008.

[Vis07] IST-1-038398 - Networked Audiovisual Media Technologies - VISNET II, "Deliverable D2.2.5: First set of developments and evaluation for search systems for distributed and large audiovisual databases". November 2007.

[TD08] Tous, R., Delgado, J.; Semantic-Driven Multimedia Retrival with the MPEG Query Format; in proc. of the Third International Conference on Semantic and Digital Media Technologies (SAMT'08), LNCS 5392, pp.149-163, 2008.

[Amd08] WD 1.0 of ISO/IEC 15938-12:2008 AMD2 Semantic Enhancement.

[ETH09] Eberhard, M., Timmerer, C., Hellwagner, H.; Fully Interoperable Streaming of Media Resources in Heterogeneous Environments. MXM Developer's day 2009.