CONTENT AND CONTEXT: TWO WORLDS TO BRIDGE

Fernando Pereira[1]

Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
fp@lx.it.pt

ABSTRACT

Networks, terminals and users are becoming increasingly heterogeneous. In this context, the growing availability and usage of multimedia content have been raising the relevance of content adaptation technologies able to fulfill the needs associated to all usage conditions without multiplying the number of versions available for the same piece of content. For efficient and effective adaptations, content and context descriptions must be available in order these two worlds are adequately bridged. This paper discusses the problem of content adaptation considering the major technologies which may have a role to play in achieving the 'one fits all' content provision paradigm with emphasis on descriptions/metadata.

1. INTRODUCTION

It is commonly said that we live today in a multimedia age. Until recently, and except for broadcast television and radio, voice was still the sole communication mechanism. However, the diffusion of digital processing algorithms and hardware has brought images, music, and video into everyday life. The availability of open standards (such as JPEG, MPEG-X, H.26X) has had a major impact on this progression. Such standards have made the creation, and communication of (digital) data aimed at our most important senses, sight and hearing, simple, inexpensive and commonplace.

While having efficient compression standards is very important, the efficient streaming, retrieval and filtering of the right content to the right user requires the availability of an adequate description of the content itself, the so called content description or content metadata. This content metadata typically represents important information about the content – 'the bits about the bits' – in a short amount of data which can be more easily processed than the full content data. The more it is know about the content through this content metadata, the better the content may be consumed by the users.

The growing heterogeneity of networks, terminals and users (see Figure 1) and the increasing availability and usage of multimedia content have been raising the relevance of content adaptation technologies able to fulfill the needs associated to all usage conditions without multiplying the number of versions available for the same piece of content while simultaneously maximizing the user satisfaction. While the availability of content metadata is essential to perform appropriate content adaptations, context metadata is also vital since in order the right adaptations for the right users are provided, it is crucial that the context where the content is to be consumed is described.

This paper addresses the problem of content adaptation/personalization using descriptions/metadata to bridge the two worlds that need to be matched: the content and the context.

2. TECHNOLOGIES FOR INTEROPERABLE ADAPTATION

In a heterogeneous world, the delivery path for multimedia content to a multimedia terminal is not straightforward. The notion of Universal Multimedia Access (UMA) calls for the provision of different presentations of the same content/information, with more or less complexity, suiting different usage environments (i.e., the context) in which the content will be consumed. 'Universal' applies here to the user location (anywhere) and time (anytime) but also to the content to be accessed (anything) even if that requires some adaptation to occur. This means content adaptation is proposed as the solution to bridge content authors and content consumers in the context of more and more diverse multimedia chains. Universal Multimedia Access requires a general and broad understanding of personalization, involving not only the user's needs and preferences, but also the user's environment's capabilities, e.g., the network characteristics, the terminal where the content will be

[1] The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org).

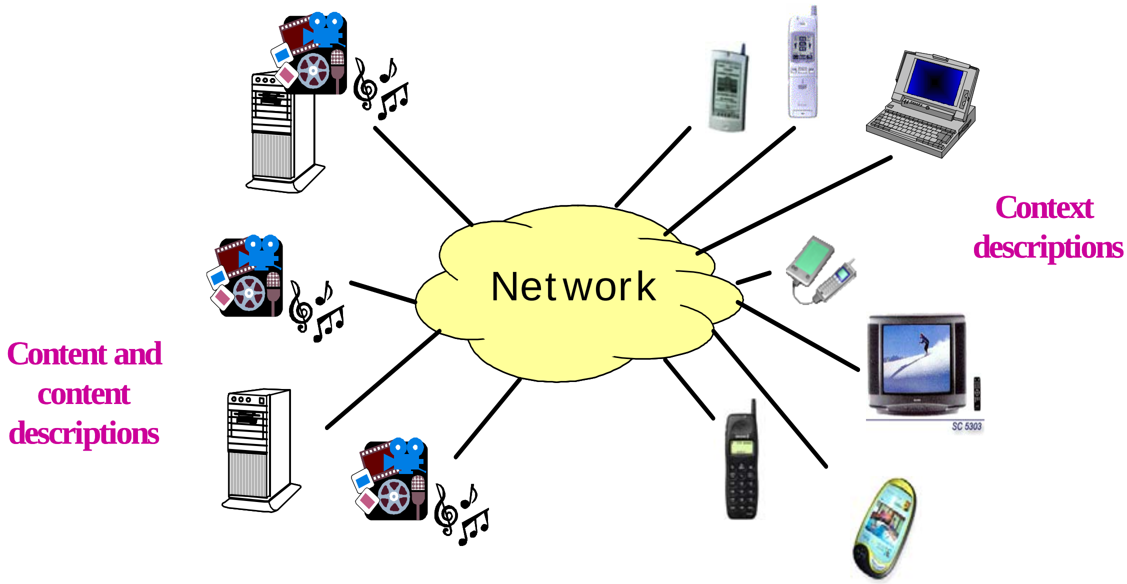

Figure 1 – Content adaptation framework

presented and the natural environment where a user is located, e.g., location, temperature and altitude.

Technologies that will allow a Universal Multimedia Access system to be constructed are starting to appear. Among the most relevant are adaptation tools that process content to fit the characteristics of specific consumption environments. These adaptation tools have to consider individual data types, e.g., video or music, as well as structured content, e.g., portals, and MPEG-21 Digital Items [1]; thus, adaptation extends from individual multimedia objects to multiple, structured pieces of content. Content and usage environment (or context) descriptions are central to content adaptation since they provide information that can control a suitable adaptation process. Today, Universal Multimedia Access service deployment is limited not only by network and terminals bottlenecks, but also by the lack of standard technologies that allow some services to hit mass markets at acceptable prices, e.g., mobile video streaming. For interoperable adaptation, some tools will need to be or are being standardized; examples are content and usage environment description, delivery protocols and rights expression mechanisms. In this context, the MPEG standardization group has been playing a central role with relevant technologies specified in all MPEG projects but mainly in MPEG-4 [2], MPEG-7 [3] and MPEG-21 [1]. Of course, some non-normative technologies are at least as important as the normative technologies such as the content adaptation rules/criteria, the content adaptation algorithms, and the usage of content descriptions for adaptation.

While universal multimedia delivery is still in its infancy it has already become clear that, as delivery technology evolves, the human factors associated with multimedia consumption increase in importance. In particular, the importance of the user and not the terminal as the final point in the multimedia consumption chain is becoming clear. The vision of mass delivery of identical content like in broadcasting is being replaced by one of mass customization of content centered on the user and on the user experience understood in a broader way [4].

3. CONTENT DESCRIPTION: THE MPEG-7 STANDARD

With the availability of the MPEG-1, MPEG-2 and MPEG-4 coding standards and, in general, the growing facility in acquiring, producing and distributing audiovisual content, it became evident that the increasing user difficulties in managing, retrieving and filtering audiovisual content had to be addressed. In fact, content has value only if it can be consumed, quickly and efficiently. After being a major responsible for the explosion of digital audiovisual content until the domestic user, MPEG recognized the need to address the problem of audiovisual content identification and management by specifying a standard way of describing various types of audiovisual information such as elementary pieces, complete works and repositories, irrespective of their representation format or storage medium. As a consequence, MPEG launched in 1996 the MPEG-7 project, formally called 'Multimedia Content Description Interface' [3]. Like previous MPEG standards, MPEG-7 answers to a set of requirements extracted from relevant applications, but this time the standard audiovisual representation to be developed does not target anymore to represent the data itself but data about the data, the so called metadata. MPEG-7 descriptions provide metadata solutions for a large set of application domains; moreover they are media and format independent, object-based, extensible and may express description capabilities with different levels of abstraction, from low-level, automatic and often statistical features, to high-level features conveying semantic meaning. The provision of a description framework that supports the combination of low-level and high-level features in a single description is a major MPEG-7 strength. In combination with the highly structured nature of MPEG-7 descriptions, this capability constitutes one of the essential differences between MPEG-7 and other available or emerging multimedia description solutions. For example, the MPEG-7 visual descriptors cover five basic visual features: color, texture, shape, motion (see Figure 2) and localization.

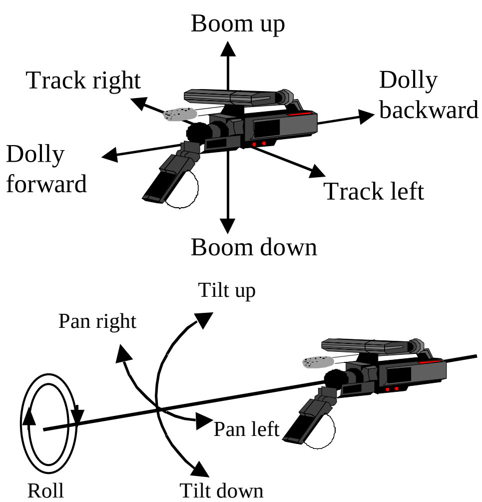

Figure 2 - Types of MPEG-7 camera motion

Following the principle, that MPEG standards must specify the minimum necessary, MPEG-7 only specifies the description format and its decoding but not the description creation and consumption engines leaving the application developers a lot of freedom.
MPEG-7 specifies two major types of tools:

- Descriptors (D) - A descriptor is a representation of a feature defining the syntax and the semantics of the feature representation; a feature is a distinctive characteristic of the data that signifies something to somebody. Examples are: a time-code for representing duration, color moments and histograms for representing color, and a character string for representing a title.
- Description Schemes (DS) - A description scheme specifies the structure and semantics of the relationships between its components, which may be both descriptors and description schemes. A simple example is: a movie, temporally structured as scenes and shots, including some textual descriptors at the scene level, and color, motion, and audio amplitude descriptors at the shot level.

MPEG-7 descriptions may be express in two ways: textual streams using the so called Description Definition Language (DDL) and binary streams using the so called Binary format for MPEG-7 data (BiM) which is basically a DDL compression tool.
In conclusion, MPEG-7 offers a powerful, flexible, and standard way of describing multimedia content.

4. CONTEXT DESCRIPTION: THE MPEG-21 DIA STANDARD

The MPEG-21 standard aims to enable the transparent and augmented use of multimedia data across a wide range of networks and devices [1]. In order to realize this goal, MPEG-21 provides a normative open framework for multimedia delivery and consumption. This framework will be of use for all players in the multimedia delivery and consumption chain. It will provide content creators, producers, distributors, and service providers with equal opportunities in the MPEG-21 enabled open market. This will also be to the benefit of content consumers, providing them access to a large variety of content in an interoperable manner.
Two concepts are central in the MPEG-21 Multimedia Framework: Digital Item and User. The 'What' in the framework is the Digital Item which is a structured digital object with a standard representation, identification and metadata within the MPEG-21 frame-work; Digital Items contain 'resources' which is the content, and metadata associated to the resources or to the overall Digital Item. The 'Who' is a User that interacts in the MPEG-21 environment or makes use of a Digital Item, including individuals, consumers, communities, organizations, corporations, consortia, governments and other standards bodies and initiatives around the world.
One of the goals of the standard MPEG-21 is to achieve interoperable transparent access to (distributed) advanced multimedia content by shielding users from network and terminal installation, management and implementation issues. This will enable the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared, always with the agreed/contracted quality, reliability and flexibility, allowing the multimedia applications to connect diverse sets of Users, such that the quality of the user experience will be guaranteed.
Towards this goal the adaptation of Digital Items is required where a Digital Item is subject to a resource adaptation engine, e.g. transcoding or transmoding (see Section 5), as well as a description adaptation engine, which produce together the adapted Digital Item. For the

adaptation, it is essential to have available not only the description of the content itself but also a description of its format and of the usage environment in order that content adaptation may be performed to provide the User the best content experience for the content requested with the conditions available. While the content description problem has been addressed by MPEG-7, the description of content format and usage environments has not been addressed and it has been the target of the MPEG-21 Digital Item Adaptation (DIA) specification [5].

A major component of the DIA specification is the so-called Usage environment description tools. These tools/descriptors describe several dimensions of the context, notably the terminal capabilities (codec capabilities, input-output capabilities, and device properties), the network characteristics (network capabilities, network conditions), the user preferences (user info, usage preferences, usage history, presentation preferences, accessibility characteristics, and location characteristics) and the natural environment (location and time, and audiovisual environment).

In conclusion, MPEG-21 DIA offers a powerful, flexible, and standard way of describing multimedia consumption contexts.

5. PROCESSING DESCRIPTIONS FOR ADAPTATION

There are typically three major types of reasons which may justify content adaptation:

- Technological limitations – Users are connected through networks and consume the content using devices/terminals with specific characteristics, e.g. bandwidth, spatial resolution.
- Semantic preferences – Users have content preferences among different content assets and within content assets, e.g., sports versus news, and football goals versus 'quiet periods'.
- Perception preferences and limitations – Users perceive content with their senses, their sensations, and even handicaps, e.g., color blind.

While content adaptation typically happens at the content server, it is also possible to perform adaptations at intermediate network nodes or gateways, where the content server asks for the adaptation services available at a certain (active) network node, or at the user terminal. This last solution may include the full or only partial adaptation at the terminal and, although rarely adopted, has at least the advantage of decreasing the relevance of privacy issues. It is also true that real-time adaptation is typically more problematic than off-line adaptation although the major problem remains the same: finding the best way to provide a user a certain content experience.

For efficient and adequate content adaptation, the availability of content and usage environment (or context) description solutions is essential since content and context descriptions provide important information for the optimal control of a suitable adaptation process. The more complete, granular and precise are the descriptions, notably of the content, the more optimal may be the adaptation.

Content is more and more available as 'smart content' which is content structured and accompanied by metadata allowing its adaptation to different usage contexts in a much less complex way.

The adaptation of content typically includes two steps:

- Selecting from the crowd - The first step in content adaptation consists in selecting the right piece of content among those (many) offered, manually or automatically. For example, this step may involve selecting among broadcast TV channels and programs using a newspaper or electronic TV program guide or selecting and downloading songs from an Internet digital music service. As already said, to provide the effective access to the growing amounts of multimedia content, e.g., by means of software agents, it is essential to capture a (standard) representation of the user's preferences.
- Adapting the selected one - The second step involves the adaptation of the piece of content previously selected to the relevant context characteristics.

In general, there are three major ways by which an adaptation system may provide a user with adequate adapted content:

- Variation selection – One from the previously available content variations of the same content is selected, e.g., several variations at different bitrates directly coded from the original may be available.
- Content scalability - A scalable coded stream is adequately truncated since each scalable stream provides a set of coded representations, different in terms of one or more scalability dimensions, e.g., quality, spatial resolution.
- Content transformation - Content is transformed based on one of the available variations by means of transcoding, transmoding or semantic filtering, such as summarization.

Many times the term content adaptation is only used to refer to the last case. Of course, content adaptation may be performed on the compressed or uncompressed data domains and this may make a lot of difference in terms of associated complexity.

Variation selection

Variation selection is a widely used solution which is becoming less efficient as consumption heterogeneity is growing. Variation selection may be combined both with scalable content and content transformation, notably to decrease the complexity of on-line adaptations. Important questions associated to the adoption of an adaptation system based on variation selection are: How many variations should be available? Which variations should be available? If variations are produced on-line, should they become newly available variations? How should available variations be managed in time, e.g. by popularity?

Content scalability

Scalable content is 'born with heterogeneity in mind' to make coding independent from transmission conditions. The main goal of scalable coding is to produce a single compressed bitstream from which it is possible to extract multiple representations of the content, different according to one or more characteristics. There are several scalability dimensions, notably spatial resolution, temporal resolution and quality/SNR. A scalable representation usually corresponds to an efficient representation of the content at successively higher bitrates and enables many applications. In MPEG, the Video subgroup is now working towards the development of a new scalable codec that addresses some weaknesses of previous MPEG scalable video coding standards, especially in terms of rate-distortion performance compared to non scalable codecs. Also, there are today available bitstream description tools which allow a node 'ignorant' about the (scalable) bitstream format to truncate it in an efficient way, for example the MPEG-21 Bitstream Syntax Description Language (BSDL).

Content transformation

Content transformation is clearly the most exciting kind of content adaptation. In terms of content transformation, the following major cases have to be considered:

- Transcoding – Content is transformed keeping the modality and amount of information but changing fidelity, e.g., video to video conversion changing spatial resolution, coding format or just quality.
- Transmoding – Content is transformed changing modality but keeping as much as possible the amount of information, e.g., video to picture, text to speech.
- Semantic filtering – Content is transformed keeping the fidelity and modality but changing the amount of information, e.g., video summary according to some criteria, region of interest, adult/aggressive content filtering.

Content may be simultaneously subject to multiple forms of adaptation, e.g., semantic filtering and transcoding when you create a summary and reduce the spatial resolution of the adapted video.

In terms of transcoding, the video content may be transformed in terms of

- Component – e.g., color to black and white.
- Color depth – e.g., 8 to 4 bit/sample.
- Bitrate or quality – e.g., 256 to 32 kbit/s.
- Error resilience – e.g., low to high resilience.
- Spatial resolution – e.g., CIF to QCIF.
- Temporal resolution – e.g., 25 to 12,5 Hz.
- Bitstream statistics – e.g., CBR to VBR.
- Format or syntactic – e.g., MPEG-4 to MPEG-1 or MPEG-4 Advanced Simple Profile to MPEG-4 Simple Profile.
- Perception - Content appearance is transformed according to specific user characteristics in terms of perception, e.g., visual handicaps, such as color blind deficiencies, or specific preferences in terms of visual temperature.

In terms of transmoding, the content may be transformed in terms of video to images, images to video or text, text to speech, etc. While some automatic transmoding transformations may be rather simple such as video to images which may just reduce to keyframe extraction, other types of transmoding such as images to text may become rather complex since they may imply some kind of semantic understanding of low-level features.

In terms of semantic filtering, the content may be transformed in terms of

- Time – The amount of information is reduced by decreasing the duration of the content, e.g., a summary according to some filtering criteria (see Figure 3).
- Space - The amount of information is reduced by focusing on a specific spatial area of the content, e.g., a region of interest in a picture.
- Scene composition - The amount of information is reduced by decreasing the number of objects in a scene, e.g., only high priority objects in an object-based scene.

While many adaptation tools are already available, adaptation services are just emerging and still have to take full benefit of the range of tools at hand. Anyway what the type of adaptation, the availability of adequate content and context descriptions is a must.

6. FINAL REMARKS AND FUTURE WORK

Heterogeneity is a growing characteristic of terminals, networks, environments, interfaces, sensors, and users. Since multimedia content cannot be available in all possible flavors for all possible usage combinations, it is necessary to adapt it. Scalable coding, content description and usage environment description are three major signal processing related technologies relevant for content adaptation asking for a normative specification; in this context, MPEG worked to provide the necessary standards.

Anyway content adaptation will not be the panacea for all problems in terms of multimedia services ... While the

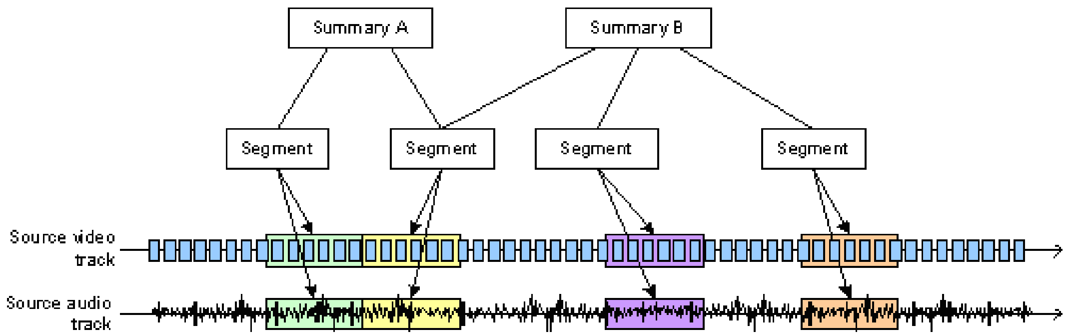

Figure 3 – Two summaries based on different criteria [6]

center of the multimedia experience. The key to future multimedia applications is in the quality of experience, notably through advanced interfaces. Experiences regard the relationship between senses and information; thus the man-information interface plays a fundamental role … and not only for sight and hearing! The psychological state of the user may be essential to select the right piece of content from those available, adapt the audio melody and amplitude or adapt the video brightness and color temperature. More than accessing the content (which should be easier and easier in the future), the true challenge of content adaptation may lie in the psychological and sensory dimension of the experience, not the technological dimension!

In conclusion, while content adaptation is an issue tackled by a growing community, and many relevant tools have been developed recently, there is still a significant number of questions to be answered before content adaptation fulfills all the expectations created.

7. REFERENCES

[1] I.Burnett, R.Van De Walle, K.Hill, J.Bormans, F.Pereira, "MPEG-21: Goals and Achievements", IEEE Multimedia, vol.10, nº 4, pp.60-70, October-December 2003.

[2] F.Pereira, T.Ebrahimi (ed.), The MPEG-4 Book, Prentice Hall, 2002.

[3] B.S.Manjunath, P.Salembier, T. Sikora (ed.), Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley & Sons, LTD, 2002.

[4] F.Pereira, I. Burnett, "Universal Multimedia Experiences for Tomorrow", IEEE Signal Processing Magazine, vol. 20, nº 2, pp. 63-73, March 2003.

[5] A. Vetro, "MPEG-21 Digital Item Adaptation: Enabling Universal Multimedia Access," IEEE MultiMedia, vol. 11, no. 1, Jan.-Mar. 2004, pp84-87.

[6] J.Magalhães, F.Pereira, "Using MPEG standards for multimedia customization", Signal Processing: Image Communication, vol. 19, nº. 5, pp. 437-456, Maio 2004