MPEG-21 DIA Standardization for Modality Conversion Preference

Truong Cong Thang*, Yong Ju Jung*, Yong Man Ro*, Jeho Nam**, Jin-Woo Hong**

* Information and Communication University (ICU)

** Electronics and Telecommunications Research Institute (ETRI)

ABSTRACT

When multimedia contents are adapted to different terminals in the ubiquitous computing environment, the contents' modalities can be converted variously. In this paper, we present an efficient user preference that enable user to specify his choices on the modalities of the adapted contents. We also propose the methods to systematically integrate the user preference into the content adaptation process. This modality conversion preference has been developed as a description tool for MPEG-21 Digital Item Adaptation (DIA).

I. INTRODUCTION

Universal Multimedia Access (UMA) is currently a new trend in multimedia communications. UMA system adapts the rich multimedia contents to various constraints of terminals and network connections, providing the best possible presentation to user. Meanwhile, the emergence of MPEG standards, especially MPEG-21 [1], facilitates the realization of UMA systems in an interoperable manner.

Actually, the best presentation to user is a very subjective concept. So, there should be some means to let user to customize the adaptation process. These means are called user preferences, the important inputs of a UMA system. This paper present an important user preference, the modality conversion preference, which we proposed and developed as a description tool in the standardization process of MPEG-21 Digital Item Adaptation (DIA) [2][1].

Modality conversion is obviously needed when the terminal or network cannot support the consumption or the transmission of certain modalities. For each content object, there would be many conversion possibilities, whereas, the user may prefer or even can hardly perceive (e.g. blind users) some modalities. The role of modality conversion preference is to let user specify his choices on modality conversion. Without this description tool, user has to accept any adaptation solution provided by the provider.

Our paper is organized as follows. In section II, we discuss the important features of the proposed description tool. In section III, we propose a systematic approach for integrating the user preference into the adaptation process. Some experiment results are presented in section IV, and finally section V concludes the paper.

II. MODALITY CONVERSION PREFERENCE

2.1 Overview

First, let us define some basic terms used in this paper. From the highest level, a *multimedia document* is a container of multiple *content objects*. A content object (or *object* for short) is an entity conveying some information, e.g. a football match. Each object may have many *content versions* of different modalities and qualities. A content version is a physical instance of the content object, e.g. a video or audio file showing a football match.

Basically, the content adaptation process includes three major modules: decision engine, modality converter and content scaler. The decision engine analyzes the content description, user preferences, constraints and then makes optimal decision on modality conversion and content (bitrate) scaling, so as the adapted contents have the most value when presented to user. Suppose we have a multimedia document consisting of multiple content objects. To adapt this document to some resource constraint (e.g. total bitrate or data size), the QOS-related decisions of the decision engine will answer simultaneously two basic questions for every content object:

1. Which is the modality of output object?
2. What is the content value of output object?

Without answers to these questions, we cannot apply the appropriate operations of modality conversion and content scaling to adapt the contents. However, it is not a good solution if the adapted document is fixed for all users.

The preference tools help user to personalize the content consumption. As for the decision engine, these are important inputs for it to answer the above basic questions. For any user preference tools, there are two basic interrelated issues, 1) the syntax to specify user's needs and 2) the method to integrate the preferences into the adaptation process.

The user preference should be flexible to support various practical cases and demands from user; we suppose the syntax of user preference need to support the following objectives:

1. Different ways to identify the contents: There may be two ways to identify the objects to give preferences. The first is the general way where the user wants to apply the preferences for all objects having some common features. The second is the specific way where the user wants to apply the preferences to some specific objects having some known URIs, e.g. via a highlight of the document.

2. *Once-for-all preferences*: user preferences may be provided just one time and then stored in the user profile for all future sessions. The preferences should be flexible so that user does not have to provide preferences again every time a change of network/terminal occurs.

2.2 Important features

Preference on modality-to-modality conversions

First we see that the user preference should not allow only the fixed selections of modality conversion. For example, the user may request that all videos be converted to audios. However, if the terminal cannot support audio modality, the contents will be discarded. Second, the user preference should not allow only the preferences on the destination modalities, because for example audio may be the best alternative for a concert video, whereas text may be the best alternative for a news image. So we propose that, to flexibly support the various conditions of terminal/network, the user preference should support selections on the very conversions from modalities to modalities.

Two levels of preference

Also, to help answer the two basic questions above, user preference for a conversion is divided into two levels. First, user will specify the relative *order* of each conversion of an original modality. Second, user can further specify the numeric *weight* of each conversion.

Given an original modality, the orders of conversions help the decision engine to determine which should be the destination modality if the original modality must be converted. For example, with the original video modality, the "video-to-video conversion", that is non-conversion of video, may have the first order; the video-to-image has the second order, and so on.

As for the weights of conversions, they help the decision engine to determine when conversion should be made. The conversion boundaries between modalities are determined by the perceptual qualities of different modalities. Meanwhile that quality is very subjective. So, the user's weights can be used to scale the qualities of different modalities, resulting in the changes of conversion boundaries of a content object.

The detailed specifications of the modality conversion preference can be found in [1]. In the next section we focus on the use of these user preferences in content adaptation process.

III. USING THE MODALITY CONVERSION PREFERENCE

3.1. Problem formulation

To tackle the above basic questions, the decision-making process of the decision engine will be first represented as the traditional resource allocation problem [3][4]. Let denote R_i and V_i the resource and content value of the content object i in the document. The content value V_i can be represented as a function of resource R_i, modality capability M, and modality conversion preference P_i^m :

$$V_i = f_i(R_i, M, P_i^m). \quad (1)$$

Then the problem of content adaptation is that: given a resource constraint R_c, *find the set of* $\{R_i\}$ *so as*

$$\sum_i V_i \text{ is maximum, and } \sum_i R_i \le R_c. \quad (2)$$

To solve this problem we first provide each content object with a content value model relating its content value with its resource. The content value models are then modified according to user preference and terminal capability. After that, a resource allocation method is used to distribute the resource among multiple contents. Mapping the allocated resources back to content value models, we can find the appropriate qualities and modalities of adapted contents.

3.2 Overlapped content value model

Content value model shows the relationship between the content value, i.e. the amount of information conveyed by the content, and its resource. To systematically support modality conversion, we propose the *overlapped content value model*. Each content object will be given an overlapped content value model (Fig. 1) representing the content values of different modalities versus the resource.

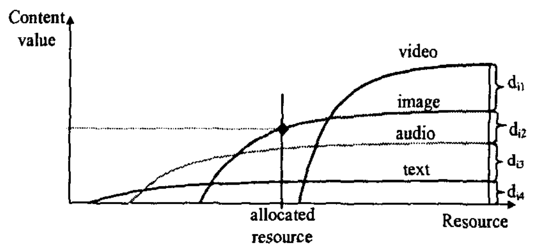

Fig. 1: Overlapped content value model of a content object

The number of curves in the model is the number of *modalities the content object may have.* Each point on a modality curve corresponds to a version of that modality. The final content value function will be the upper hull of the model, and the intersection points of the model represent the conversion boundaries between modalities. Denote J_i as the number of modalities and VM_{ij} as the content value curve of modality j of the content object i,

$j=1\ldots J_i$. The content value of a content object can be mathematically represented as follows:

$$V_i = \max\{VM_{ij}\} \text{ with } j = 1\ldots J_i \qquad (3)$$

The value j=1 indicates the original modality of the content object. The content value is obviously subjective and changes variously according to different users. In the following, the user preference is used to modify the content value model.

3.3 Using the user preference

The content value models are the important inputs of the adaptation process. In our approach, user preferences and are used to modify the content value models, resulting in appropriate changes at the output. For completeness purpose, the terminal capability is also considered.

Modifying according to modality capability

It is clear that when a terminal cannot support some modalities, the content values of those content versions at the terminal become zero. That means, the curves of the non-supported modalities need to be removed in the adaptation process. Then we have $V_i = \max\{VM_{ij}\}$, where j's are now indexes of the supported modalities

Modifying according to user preferences

Now consider the user preference on modality conversion. In fact, with a predefined content value model, there are already the orders of conversions. For example in Fig. 1, the video-to-video has the first order, video-to-image has the second order, and so on for the other conversions. These can be considered as the orders assigned by provider. User's orders of conversions may change the existing sequence of orders. In our approach all the curves that have violated orders (of user and provider) are removed. As mentioned above, the content values of modalities are very subjective. In our solution, the weights of conversions are used to scale the "distances" d_{ij} between the modality curves (as Fig. 1). The result of this scaling is the changes in the intersection points, or the boundaries between the modalities. If the weight of a curve increases, the operating range of the corresponding modality (delimited by the intersection points) will be broadened. The scaled distances d^s_{ij} are calculated as follows:

$$d^s_{ij} = \frac{w_{ij} d_{ij} \sum_j d_{ij}}{\sum_j w_{ij} d_{ij}} \qquad (4)$$

3.4 Resource allocation

A content value function can be continuous or discrete. If it is continuous, we may discretize it because the practical transcoding is done in the unit of bit or byte. In the following we implicitly suppose it is discrete, originally or after discretization. Meanwhile, function (1) is inherently non-concave, thus the above optimization can be solved optimally by Viterbi algorithm (VA) of dynamic programming.

In fact, the problem of resource allocation represented as a constrained optimization is often solved by two basic methods, Lagrangian method and dynamic programming method [4]. In [3], Lagrangian method is adopted to find the allocated amount of resource. However, Lagrangian method is only suitable with concave content value functions. If our content value function is the concave hull of the overlapped content value model, the advantage in discriminating the modalities is eliminated. The advantage of the dynamic programming is its support for the non-concave functions, however its disadvantage is the high complexity.

IV. EXPERIMENT RESULTS

We have deployed a trial system to test the efficiency of the proposed approach. The system includes a multimedia server and various types of clients such as PCs, Laptops, PDAs. For each content object, the server stores multiple versions of different modalities and resolutions. The current resource constraint R_c for the content adaptation is the total data size at client, measured by Kilo Bytes.

Figure 3: Adaptation when video modality is not supported

Due to the limited space, we just show some example simulated cases here. Figure 2 shows an adapted document when R_c=1100 (KBs). In this case there is no modality conversion. We see that this document has one video, three images, one text paragraph, and one audio. Figure 3 shows the adapted document when R_c is still 1100 (KBs) but video is not supported by the terminal. We see that in this case the video is converted to a sequence of images. In case all modalities are supported, but Rc is reduced as low as 450KBs, the video and all images are converted to audio as shown in figure 4.

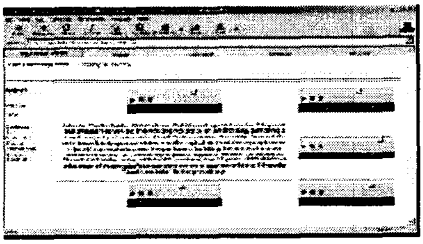

Figure 4: Adaptation with R_c = 450KBs

Figure 5: Adaptation when video modality is not supported and order of video-to-audio is second

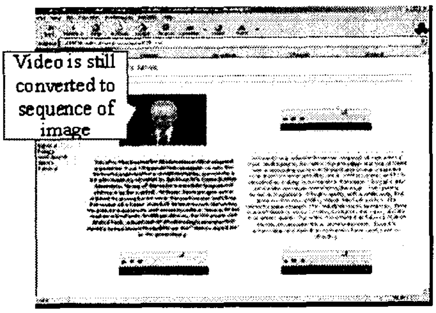

Figure 6: Adaptation when the weight of video-to-image is increased from 1 to 3. (R_c = 450KBs)

Consider the case of figure 3 again. In this case the sequence of conversion orders is default, that is, order of video-to-video is the first, order of video-to-image is the second, order of video-to-audio is the third. Now the user wants that, if the video must be converted, it should be converted to audio first, i.e. order of video-to-audio is the second and order of video-to-image is the third. The newly adapted document with this user preference is shown in figure 5. We can see that the video is now converted to audio, not sequence of images.

Again with the case of figure 4 where the weights of conversions actually have default value of 1. Now if the weight of video-to-image is increased to 3, that means the operating range of video-to-image is broaden, we have the newly adapted document as shown in figure 6. We can see that the video is now converted to a sequence of image, not audio.

The above experiment results show that the system can adapt dynamically and efficiently to different conditions of terminals, resource constraints. Especially, the user preference is shown to be very helpful for user to customize his content consumption.

V. CONCLUSION

In this paper, we have described the modality conversion preference, which is crucial in providing customized content presentations to different users. We also proposed a systematic approach for integrating the user preference. The combination of the overlapped content value model, user preference, and the dynamic programming method, provides a comprehensive solution for content adaptation, for both modality conversion and content scaling. Our future works will be carried out in three main directions. The first is exploring more efficient ways to smoothly combine the user preference into the content adaptation process. The second is further modeling the content value of contents within a single modality and across multiple modalities. And the third is extending to consider a combination of different resource constraints such as bandwidth, data size, screen size, etc.

REFERENCES

[1]. "MPEG-21 Digital Item Adaptation SCD", ISO/IEC JTC1/SC29/WG11 N5612, Pattaya, Thailand, Mar. 2003.

[2]. T. C. Thang, et al., "Additional Display Presentation Preferences in MPEG-21 DIA", ISO/IEC JTC1/SC29/ WG11 M8970, Shanghai, China, Oct. 2002.

[3]. R. Mohan, J. R. Smith, C.-S. Li, "Adapting Multimedia Internet Content for Universal Access", IEEE Trans. Multimedia, Vol. 1, No. 1, pp. 104-114, Mar. 1999.

[4]. A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression", IEEE Signal Processing Magazine, pp. 23-50, Nov. 1998.