Modality Conversion For Media QoS

Truong Cong Thang, Yong Ju Jung, and Yong Man Ro
Multimedia Group, Information and Communication University, Taejon, 305-732, Korea
Tel : +82-42-866-6279 Fax : +82-42-866-6245
E-mail: {tcthang, yjjung, yro}@icu.ac.kr

Abstract: We present modality conversion as an effective means for QoS management. We show that modality conversion, in combination with content scaling, would give a wider range of adaptation to support QoS at media level. Here, we consider modality conversion with respect to resource constraint and human factor. To represent modality conversion as well as content scaling, we present the overlapped content value (OCV) model that relates the content value of different modalities with resources. The specification of user preference on modality conversion is divided into qualitative and quantitative levels. The user preference is then integrated into the OCV model so that modality conversion correctly reflects the user's wishes. For the conversion of multiple contents, an optimization problem is formulated and solved by dynamic programming. The experiments show that the proposed approach is efficient to be applied in practice.

Keywords: Modality conversion, Content adaptation, QoS, User preference

1. INTRODUCTION

In multimedia communication, QoS is a crucial issue where the provider supplies, under different resource constraints, the best possible quality to the user. QoS is a broad concept which can be tackled from different levels. Usually, QoS is considered from two perspectives, the media perspective where source content can be adapted to the resource and the network perspective where the transmission protocol is adaptive to the resource and the content. These two perspectives result in two main research topics, media QoS and network QoS. This paper focuses on the media perspective.

The key technology to support media QoS is content adaptation [1]. Content adaptation has two major aspects: one is modality conversion that converts content from one modality to different modalities, the other is content scaling that changes the bitrates (or qualities) of the contents without converting their modalities.

So far, media QoS has been mostly supported by content scaling [2]. Intuitively, given some resource constraint of terminal/network, the provider will (down) scale the contents to meet the constraint while still providing the best possible quality to the user. However, in some cases, the quality of the scaled contents is unacceptable or not as good as that of a substitute of a different modality. That is, content scaling may not be able to handle the wide range of resource constraint variations. A possible solution for this problem is to convert the contents into other modalities. For example, when the connection bitrate is too low, sending a sequence of "important" images would be more appropriate than streaming a scaled video of low quality. This is a typical case of conversion from video modality to image modality. So, modality conversion, in combination with content scaling, would give a wider range of adaptation to support QoS at media level.

There are four main factors that may affect the decision on modality conversion. The first factor is the modality capability, which is the support for user's consumption of certain modalities. This factor can be determined from the characteristics of terminal (e.g. text-only pager), or surrounding environment (e.g. a too noisy place). The second factor is the user preference that shows user's levels of interest to different modalities. The third factor includes the resource constraints. The fourth factor is the semantics of the content itself. For instance, between an interview video and a ballet video, the provider would be more willing to convert the former to a stream of text.

Currently, modality conversion is often carried out only when some modality is not present in the modality capability (e.g. [3]). In this paper, modality conversion is considered with the factors of resource constraint and user preference. From the QoS point-of-view, two most important questions for modality conversion are "At which resource constraint should the current modality be converted?" and "What is the destination modality?" On the other hand, the best quality to user is a very subjective concept. For each content, there would be many conversion possibilities, whereas, the user may prefer or even can hardly perceive (e.g. blind users) some modalities. So, there should be some means to let user to customize the adaptation process. These means are called user preferences and considered as components of the user profile, which is an important input of a UMA system [1].

To this end, we present a systematic approach that support modality conversion while at the same time taking into account other contextual factors like user preference, terminal capability. We present the overlapped content value (OCV) model to represent the relationship between content values of different modalities and resource, helping find the conversion boundaries between modalities. We show how the user preference on conversion can be efficiently specified

and then incorporated into the adaptation process based on the conceptual OCV model. To handle the adaptation problem with multiple contents, we formulate the problem as a constrained optimization, and show that dynamic programming can be applied to optimally solve the problem.

The paper is organized as follows. In Section 2 the modeling of modality conversion is presented. The human factor in modality conversion, including the specification and the integration into adaptation process, is described in Section 3. Section 4 presents the solution to make decisions on scaling and conversion for multiple contents. The experiments are shown in Section 5, and finally the conclusion is provided in Section 6.

2. MODALITY CONVERSION MODELING

The process of content scaling can be represented by some "rate-quality" curve, which shows the quality of the scaled content according to the bitrate (or any resource in general). A recent trend in UMA is to use this rate-quality curve as the metadata to automate content scaling [1][2]. Usually, the rate-quality curve is obtained for a particular modality because each modality has its own characteristics. Extending this concept, we introduced the overlapped content value (OCV) model to conceptually represent both content scaling and modality conversion [4].

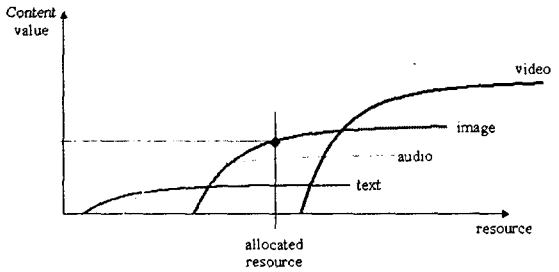

Fig. 1: Overlapped content value model of a content.

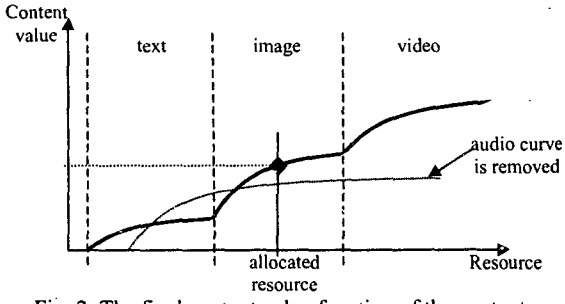

Fig. 2: The final content value function of the content.

Fig. 1 shows the example OCV model of a video content, which consists of the rate-quality curves of different modalities (called modality curves). These curves, provided manually or automatically, are normally non-decreasing and saturate when the resource is high. We can see that the intersection points of the modality curves represent the conversion boundaries between modalities.

By the proper estimation of content value for different modalities, we can put the modality curves into an OCV model. Denote $VM_j(R)$ the rate-quality curve of modality j of a content, $j=1...J$, where J is the number of modalities of the content; R is the resource. $VM_j(R) \geq 0 \ \forall j$. The content value function, which is the upper hull of the model, can be written as follows:

$$V = \max\{VM_j(R) \mid j=1...J\} \quad (1)$$

The content value is obviously subjective and changes variously according to different users. For example when the user is deaf, the audio curve should be excluded from the content value model and the final content value function is shown in Fig. 2. From the OCV model, given an allocated resource of the content item, we can easily find the appropriate modality and content value of the content item, so as maintaining an acceptable QoS.

3. HUMAN FACTOR IN CONVERSION

This section addresses the dependence of modality conversion on user. We first present the modality conversion preference, then we propose the methods to integrate the preference into the adaptation process, specifically to modify the OCV model.

3.1 User preference on modality conversion

To flexibly support the various conditions of terminal/network, the user preference should support selections on the very conversions from modalities to modalities. 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, has the first order; the video-to-image conversion 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. This is based on the fact that 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 user preference tool can be found in [4][5]. In the next section we focus on the integration of these user preference into the OCV model.

3.2 Integrating the User Preference into Adaptation

3.2.1 Using orders of conversions

In fact, with a predefined content value model, there are

already the orders of conversions. These can be considered as the orders assigned by provider. User's orders of conversions may change the existing sequence of orders, and the procedure to modify the content value model of content item i is as follows:

1. Check the existing orders and the user's orders of conversions.
2. Take a modality curve VM_{ij}, compare it with every curve VM_{ij*} that has a lower existing order. If the user's order of VM_{ij} is lower than the user's order of any VM_{ij*}, remove VM_{ij}.
3. Repeat step 2 for all modality curves of object i.

As an example, given the original model in Fig. 1, if the user requests the orders of conversions as follows: video-to-video is "first", video-to-image is "second", video-to-audio is "fourth", and video-to-text is "third", then the model depicted in Fig. 2 is also the resultant modified model in this case, in which the audio curve is removed by the above algorithm.

3.2.2 Using weights of conversions

Because the content value or quality of each modality is very subjective, the user can change the conversion points (intersection points) by some quantitative preference on the conversions. In our solution, the weights of conversions are used to scale the "distances" d_{ij} among the curves of modalities as shown in figure 3. Note that the sum of d_{ij} is fixed and equal to the maximum content value of the content item i.

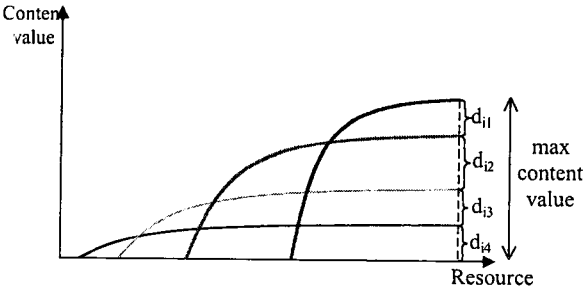

Figure 3: Scaling the curves according to the weights

Denote w_{ij} as the weight of conversion j of content item i, the scaled distances can be computed as follows. We first multiply the weights with the corresponding distances $d'_{ij} = w_{ij} \cdot d_{ij}$. The relative lengths of these new distances d'_{ij} reflect the user preference, however these lengths still need to be scaled once more to keep the sum of distances unchanged. The final distances are:

$$d^s_{ij} = \frac{w_{ij} d_{ij} \sum_j d_{ij}}{\sum_j w_{ij} d_{ij}} \qquad (2)$$

where d^s means the final scaled distance. And we can easily see that $\sum_j d_{ij} = \sum_j d^s_{ij}$.

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.

4. MULTIPLE CONTENTS ADAPTATION

Suppose we have a multimedia document consisting of multiple content objects. To simultaneously adapt these multiple objects to some constraint, the system has to answer two basic questions for every object: 1) which is the modality of output object and 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.

Let denote R_i and V_i the resource and content value of the object i in the document. The content value V_i can be represented as a function of resource R_i, and user preference P_i:

$$V_i = f_i(R_i, P_i). \qquad (3)$$

The eq. (3) is obtained from the OCV model and the user preference as in the above sections. Then the problem of content adaptation for the given document can be represented as the constrained optimization problem as follows [6]:

Given a resource constraint R^c, find the set of $\{R_i\}$ so as

$$\sum_i V_i \quad \textit{is} \quad \textit{maximum}, \qquad (4a)$$

$$\textit{and} \quad \sum_i R_i \le R^c. \qquad (4b)$$

A content value function can be continuous or discrete. If the function is continuous, we may discretize it because the practical transcoding is done in the unit of bits or bytes. So we can implicitly suppose the function is discrete. Meanwhile, function (3) is inherently non-concave, thus the above problem can be solved optimally by dynamic programming [6]. The disadvantage of dynamic programming is the high complexity. In order to speed up the allocation process, Viterbi algorithm [7] and heuristic approximations are applied for dynamic programming.

5. EXPERIMENTS

We have developed a test-bed for providing the multimedia services over heterogeneous networks [8]. The adaptation engine is built on a Windows2000 server with Pentium IV 1.7GHz and 256MB RAM. In our system, the content transcoding is simply done offline. Scaling operations include reducing the spatial size for video and image modalities, reducing bandwidth for audio, and truncating the words for text. The resource constraint is datasize constraint D^c.

For experiments, we employ a multimedia document consisting of six objects: one video, one audio, three images, and one text paragraph. The original datasizes of the objects are respectively 1500KBs, 480KBs, 731KBs, 834KBs, 813KBs, and 8KBs. The OCV model for each content object is obtained from subjective tests. The highest possible content value for any objects is 10, yet the actual maximum content value of each object is assigned according to its relative importance. The objects are transcoded in unit of kilobytes (KBs), so the content value functions can be discretized by the uniform step size of 1KBs.

Table I: Default orders of conversions in experiment

from \ to	Video	Image	Audio	Text
Video	first	second	third	fourth
Image	unsupported	first	second	third
Audio	unsupported	unsupported	first	second
Text	unsupported	unsupported	second	first

First, all user preferences are set to be default with the *weights* of 1 and the *orders* given in Table I. To check the response of the adaptation system, we vary the datasize constraint D^c. Table II shows the document versions adapted to different values of D^c. In this Table, the first column is D^c; each object has two columns, one for the datasize and the other for the modality; the last column is the total content value of adapted document. Here, *Mod* means modality and *V, I, A, T* mean video, image, audio, text modalities respectively. We can see that as D^c decreases, the datasizes of the objects are reduced to satisfy the datasize constraint of the whole document. Also, at some points, the modalities of the objects are converted to meet the constraint and to give the highest possible total content value.

From Table II, we note that as D^c is reduced, the image objects are converted to audio modality and then text modality as the result of default orders of conversions. Now, the user may prefer that images be converted to text modality rather than to audio. Using modality conversion preference the user sets the order of image-to-audio to "third", and the order of image-to-text to "second". The corresponding adaptations are shown in Table III in which the user's need is satisfied. Now if the image objects need to be converted, they will be converted to text modality.

Also, we see in Table II that object 1 (originally of video modality) is converted to image modality when D^c is reduced to 130KBs. Suppose the user wishes to watch the video modality beyond the normal limit, then the user increases the weight of the video-to-video to 1.2 for example. Table IV shows the adaptations subject to this preference. We see that now the video modality, obviously with lower quality (datasize of 60KBs), is retained at D^c = 130KBs. Anyway this tradeoff between modality and quality is acceptable to the user.

The above experiments show that the system adapts accurately and flexibly to different values of resource constraint as well as user preferences.

For the current document, the maximum processing time is below 0.4 second. One important feature of multimedia document is that the number of content objects in a document is not many, often not more than a few dozens. Our experiments also show that the processing time for such numbers of objects is just several seconds, which is acceptable for the practical Internet services.

6. CONCLUSIONS

We have studied the modality conversion as a solution to control the QoS of multimedia services. The conversion between modalities is considered with respect to resource and user preference, based on the conceptual overlapped content value model. By comparing the content values of different modalities in the model, the adaptation engine can quantitatively make decisions on modality conversion as well as content scaling. For the conversion of multiple contents, we show that the optimal decisions can be obtained from a constrained optimization, using dynamic programming. In the future, we will focus on the efficient estimations of content values across various modalities.

References

[1] A. Vetro, "MPEG-21 digital item adaptation: enabling universal multimedia access", *IEEE Multimedia*, vol. 11, pp. 84-87, 2004.

[2] A. Vetro, C. Christopoulos, H. Sun, "Video transcoding architectures and techniques: an overview", *IEEE Signal Processing Magazine*, vol. 20, pp. 18-29, 2003.

[3] A. Kaup, "Video analysis for universal multimedia messaging", Proc. 5th IEEE Southwest Symp. Image Analysis and Interpretation, pp. 211-215, 2002.

[4] Thang, T.C. et al.: 'CE report on modality conversion preference: part-I', ISO/IEC JTC1/SC29/WG11 M9495, Pattaya,.2003.

[5] ISO/IEC FDIS 21000-7, 'Information Technology – Multimedia Framework – Part 7: DIA', 2003

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

[7] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268-278, Mar. 1973.

[8] Y. J. Jung, T. C. Thang, J. Lee, and Y. M. Ro, "Visual media adaptation system for active media," in *Proc. 2003 Int. Conf. on Imaging Sci., Syst., and Tech.*, Las Vegas, 2003.

Table II. Results of the adapted documents with different values of constraint (default preferences)

	Object 1		Object 2		Object 3		Object 4		Object 5		Object 6		Content
D^c (KBs)	D_1 (KBs)	Mod	D_2 (KBs)	Mod	D_3 (KBs)	Mod	D_4 (KBs)	Mod	D_5 (KBs)	Mod	D_6 (KBs)	Mod	value
3000	1041	V	260	A	544	I	571	I	576	I	8	T	28.30
1000	343	V	90	A	179	I	189	I	191	I	8	T	24.08
300	100	V	30	A	52	I	56	I	57	I	5	T	14.88
200	85	V	26	A	20	A	47	I	21	A	1	T	11.38
130	45	I	25	A	19	A	20	A	20	A	1	T	8.47
110	36	I	21	A	17	A	17	A	18	A	1	T	7.59
90	34	I	5	T	16	A	17	A	17	A	1	T	6.61
70	14	A	5	T	16	A	17	A	17	A	1	T	5.61
10	1	T	3	T	1	T	2	T	2	T	1	T	1.27

Table III: Results of adapted documents when "order" of image-to-text conversion is "second" and "order" of image-to-audio conversion is "third".

	Object 1		Object 2		Object 3		Object 4		Object 5		Object 6		Content
D^c (KBs)	D_1 (KBs)	Mod	D_2 (KBs)	Mod	D_3 (KBs)	Mod	D_4 (KBs)	Mod	D_5 (KBs)	Mod	D_6 (KBs)	Mod	value
300	100	V	30	A	52	I	56	I	57	I	5	T	14.88
200	67	V	22	A	34	I	38	I	38	I	1	T	11.13
130	15	A	5	T	34	I	37	I	38	I	1	T	7.82
110	12	A	5	T	28	I	32	I	32	I	1	T	6.81
90	13	A	5	T	2	T	34	I	35	I	1	T	5.75
70	1	T	5	T	2	T	30	I	31	I	1	T	4.70
10	1	T	3	T	1	T	2	T	2	T	1	T	1.27

Table IV: Results of the adapted documents when the "weight" of video-to-video conversion is 1.2

	Object 1		Object 2		Object 3		Object 4		Object 5		Object 6		Content
D^c (KBs)	D_1 (KBs)	Mod	D_2 (KBs)	Mod	D_3 (KBs)	Mod	D_4 (KBs)	Mod	D_5 (KBs)	Mod	D_6 (KBs)	Mod	value
300	100	V	30	A	52	I	56	I	57	I	5	T	14.88
200	85	V	26	A	20	A	47	I	21	A	1	T	11.38
130	60	V	20	A	16	A	16	A	17	A	1	T	8.44
110	34	I	22	A	17	A	18	A	18	A	1	T	7.46
90	14	A	22	A	17	A	18	A	18	A	1	T	6.54
70	1	T	20	A	16	A	16	A	16	A	1	T	5.57
10	1	T	3	T	1	T	2	T	2	T	1	T	1.27