Modality Conversion in Content Adaptation for Universal Multimedia Access

Truong Cong Thang, Yong Ju Jung, and Yong Man Ro

Multimedia Group, Information and Communications University (ICU),
Yuseong, Daejon, PO Box. 77, 305-732, Korea

ABSTRACT

Modality conversion is an important part in the content adaptation process of a Universal Multimedia Access system. The decision on modality conversion is dependent on terminal/network conditions and on the user preference. We propose a systematic and comprehensive approach to make decision on modality conversion in a UMA system. Our approach includes three integral components, (1) an overlapped content value model, (2) a flexible specification of user preference on modality conversion, and (3) a resource allocation method to distribute resource to multiple contents.

Keywords: multimedia, UMA, transcoding, modality conversion, resource allocation

I. INTRODUCTION

Universal Multimedia Access (UMA) is currently a new trend in multimedia communications [2]. In a UMA system, content adaptation is the most important process to provide the best possible presentation under constraints of various kinds of terminals and network connections available today. Content adaptation has two aspects: one is *modality conversion* that converts content from one modality (e.g. video) to different modalities (e.g. image), the other is *content scaling* that changes the bit rate (or quality) of the contents without converting their modalities. It should be noted that, in the literature, sometimes the term *transcoding* means content scaling only, however in some cases it also covers the meaning of modality conversion. To avoid the confusion, we would like to use the term *content scaling* instead of content transcoding within a single modality.

Modality conversion is obviously first needed when the terminal cannot support certain modalities; we call this kind of support the *modality capability* of terminal. Besides, when one or more total resource constraints - e.g. total data size at terminal or bit rate of network connection - are small, modality conversion (together with content scaling) will be used to reduce the resource requirements of contents. Further when the user prefers or even can hardly perceive (e.g. visually impaired users) some modalities, modality conversion is also necessary.

So far, most researches on content adaptation have focused on transcoding of contents within a single modality [2][3], or on a single type of modality conversion, e.g. video to images [4]. Modality conversion may be supported in the approach of [5], yet this approach works with only one content item, resulting in little practical use. The approach in [6] is one of few adaptation approaches that can handle multiple modalities and multiple contents, however its resource allocation method is not quite suitable for making decision on modality conversion, and this will be more explained later. Especially, user preference on modality conversion has not been examined in those researches.

In this paper, we propose a systematic and comprehensive approach that can support modality conversion. This approach can handle multiple contents of a composite document and accommodate different constraints from terminal/network as well as user preference.

The paper is organized as follows. Section II formulates the content adaptation process with focus on the modality conversion aspect. Section III describes the overlapped content value model that shows the interrelationship of different modalities. The user preference on modality conversion and its integration into adaptation process is presented in section IV. The selection of resource allocation method is discussed in section V. Section VI presents some experiment results, and finally section VII concludes the paper.

II. MODALITY CONVERSION IN CONTENT ADAPTATION

2.1 Overview of content adaptation

Content adaptation process can be considered as the heart of a UMA system. The conceptual description of content adaptation, as given in figure 1, includes three main parts: decision engine, modality conversion engine and content scaling engine.

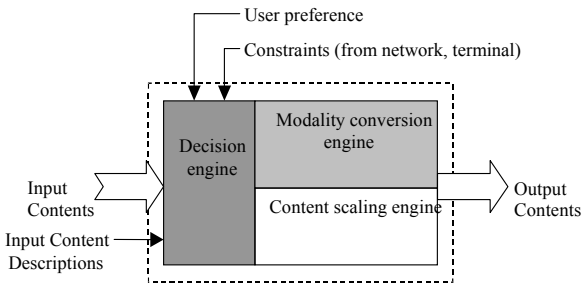

Figure 1: Content adaptation process

The decision engine takes contents and associated description (metadata) of the contents, user preference and other information about resource constraints (from network, terminals, costs...) as its inputs. Here, we focus on the modality conversion aspect then only user preference on modality conversion, also called modality conversion preference, is considered. The decision engine analyzes the content description, user preference, constraints and makes optimal decision on content modality conversion and content scaling, so as the adapted contents have the most value when presented to user. The modality conversion engine and the scaling engine include the specific operations to adapt the content according to instructions from decision engine. It should be noted that these two engine can be either offline or online. In the offline case, the decision engine just selects the appropriate content versions (of certain modality and quality) that are created in advance. In the online case, the content versions are created on the fly. Our current solution targets at the decision engine, and it can be used for both online and offline cases of content adaptation.

Before delving into the detailed problem formulation, it is necessary to clarify some basic terms used in this paper. From the highest level, a *multimedia document* is a container of multiple *content items*. A content item is an entity conveying some information, e.g. a football match that can be represented by any means, e.g. video, image, text, etc. Each content item can have many *content versions* of different qualities and modalities. A content version is a physical instance of the content item, e.g. it can be a video, an audio, etc, showing the information of a football match.

2.2 Problem formulation

Suppose we have a multimedia document consisting of multiple content items. To adapt this document to some resource constraints (e.g. total bit rate or total data size), the QOS-related decisions on modality conversion of the decision engine will answer simultaneously two basic questions for every content item:

1. When should modality conversion be made?
2. Which is the modality of output content item?

The first question implies the quality trade-off among modalities, that is, at what reduced quality level of current modality, the modality should be converted to maintain an acceptable level of QoS. And the second question is itself clear. Without answers to these questions, we cannot apply the appropriate operations of modality conversion (and content scaling) to adapt the contents. As described in the above, the answers to these questions will depend on three factors, modality capability of terminal, constraints, and modality conversion preference. To our best knowledge, there have been no systematic researches that can answer these two questions at the same time.

To tackle these questions, the decision-making process of the decision engine will be first represented as the traditional resource allocation problem as follows [6]. Let denote R_i and V_i the resource and content value of the content item i in the document. Here, the resource of content item can be the data size or the bit rate, and the content value means the amount of information conveyed by the content item.

The normal trend is that V_i is a non-decreasing function with respect to R_i. V_i is obviously dependent on the subjective evaluation of human being, which varies widely among different people. Also, when some certain modalities are not supported at terminal, the content of those modalities would become useless, that is, the content values become zero. So, the content value V_i is represented as a function of resource R_i, modality capability M, and user preference P_i:

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

Then we have the problem statement: given a resource constraint R_c, find the set of $\{R_i\}$ so as

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

To solve the problem stated above, our proposed approach will consist of three integral components:

1. A content value model: that gives the relationship between content value and resource.
2. A specification of user preference: that gives user a flexible way to have choices on modality conversion.
3. A resource allocation method: that distributes the total resource among multiple content items.

The process of the approach is as follows. First, each content item will be given a specific content value model relating its content value with its resource. The content value models are then modified according to user preference and terminal modality capability. After that,

the resource allocation method is used to distribute the resource among multiple content items. Mapping the allocated resources to content value models, we can find the appropriate qualities and modalities of adapted contents. In the following sections, we will present the detailed solution of the above problem.

III. OVERLAPPED CONTENT VALUE MODEL

Content value model is a variation of the traditional rate-distortion model [6][7], in which the distortion of a compressed signal is related to its bit rate. Here, content value model shows the relationship between the content value, which is the amount of information conveyed by the content, and its resource. As to a measure for information, content value is better than the distortion because it is not easy, if not impossible, to measure the distortion of contents in different modalities. Content value has been relatively considered in some recent researches.

In [5], the content value, which they call *quality value*, is computed from multiple quality axes, e.g. color depth, resolution, etc. The content value is related indirectly to its resource through the adaptation strategies, which are actually the techniques to adapt the content to some constraint. A set of ordered nodes, each consists of a content value and corresponding adaptation strategies, is used to search for the adapted version. However, this adaptation is just for a single content item. In [8], a metric of quality is extensively considered based on the compression ratio, which is well related to the amount of resource. However this quality metric is specifically for JPEG image. In [6], although their content representation scheme contains content versions of different modalities, the content value is related to the resource by some *single* analytic function (e.g. *log* function) or an arbitrary curve assigned by the creator or provider. This content value model cannot show the interrelation of content values in different modalities.

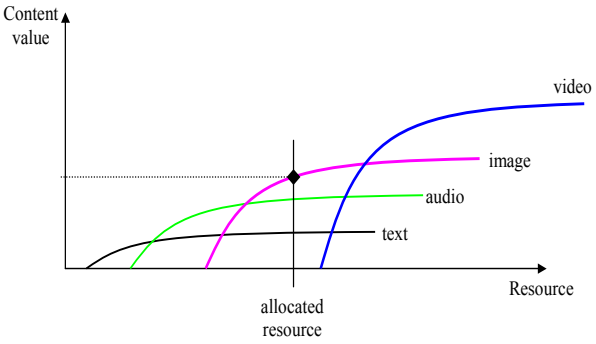

Figure 2: Overlapped content value model of a content item

In this paper, we propose a novel content value model, called the overlapped content value model, that can support the QoS-related decisions on modality conversion. Each content item has an overlapped content value model (figure 2) that represents the content value of different modalities versus the resource. The curve of each modality can be assigned by the content provider or be given by some analytic functions. Each point on a modality curve is corresponding to a version of that modality.

The number of curves in the model is the number of modalities the content item would have. The final content value function will be the upper hull of the overlapped model, and the intersection points of the model represent the boundaries between modalities.

Suppose VM_{ij} is the content value curve of modality j of the content item i; $j = 1 \ldots K_i$ where K_i is the number of modalities of content item i. We also require $VM_{ij} \geq 0$ with $j = 1 \ldots K$. The content value of a content item can be mathematically represented as follows:

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

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 figure 3. Given an allocated resource of the content item, we can easily find the appropriate modality and content value of the content item. The dependence of content value on user will be considered in the next section, where the upper hull will be modified according to user preference and the modalities supported by terminal.

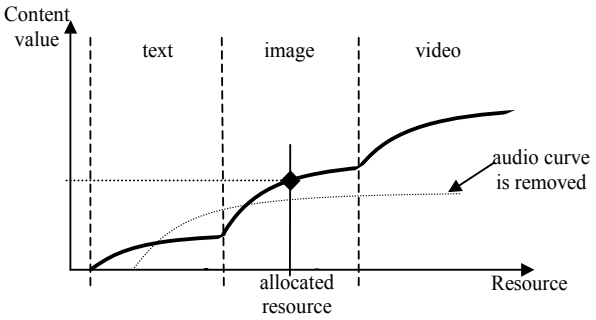

Figure 3: Example of a final content value function of a content item.

Of course the curve of each modality can be totally arbitrary, however it will be more effective if we can find some acceptable analytic functions to model the content value in each modality. In [6] the natural log function is used to relate content value to resource, regardless of content modality. Specifically $V = a*ln(R)$, where V is content value, R is the resource, and a is a scale factor.

However, let's consider an extreme case that the resource increases to infinity. It is practically clear that to user, the perceptual information will not be infinite, nevertheless the log function will give an infinite content value by its nature. Here we intentionally propose a simple analytic function for the curve of each modality as follows:

$$VM_{ij} = a_{ij}(R_i - b_{ij})/(R_i - b_{ij} + c_{ij}) \qquad \text{with } R_i \geq b_{ij} \qquad (4)$$

Figure 4 illustrates the analytic function for the case $a_{ij} = 1$, $b_{ij} = 50$, $c_{ij} = 100$.

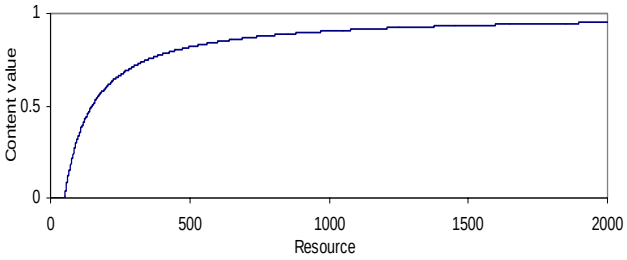

Figure 4: Example of analytic function of a modality curve

We can see that a_{ij} is the upper limit of the function, b_{ij} is the starting point of the function, and c_{ij} controls the slope of the function. Using this analytic function, we can model the different content items by tuning the parameters $\{a_{ij}, b_{ij}, c_{ij}\}$. The important point of this analytic function is its saturation when resource goes to infinity. Of course, this function may be extended in some way to accommodate more complicated content value models of some specific content types. Further modeling the content value both within a single modality and across different modalities will be reserved for our future research.

IV. MODALITY CONVERSION PREFERENCE

4.1 Preference on modality-to-modality conversion

The rationale of preference on conversions is discussed in detail in [1]. The basic point is that the user *should not specify the fixed choice of modality conversion* because content items may be discarded if the time-variant characteristics of terminal/network cannot support that fixed choice. Also, the user *should not specify the preference on the destination modalities alone* because the preference on output modality may depend on the input modality. So, to flexibly support the various conditions of terminal and network, we propose that the user *specifies preference on the very conversions from modalities to modalities*. It should be noted that the preference on modality-to-modality conversions can cover the cases of fixed choices and destination modalities.

4.2 Two levels of preference

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. We can see that the first level is qualitative and the second level is quantitative. The user may just select the orders of conversions and leave the weights to be default values.

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; and 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 item.

4.3 Modifying the content value model

The content value models of content items are the important metadata inputs of the adaptation process. Any change in the content value models will result in the changes at the output of the adaptation process. Meanwhile, the adaptation process needs to take into account the modality capability and user preference. In our approach, these factors are used directly to modify the content value models, resulting in the appropriate changes at the output; and we try to keep the resource allocation algorithm as much independent on input information as possible. This separation helps to modularize the adaptation process.

4.3.1 Modifying according to modality capability

Let's consider the modality capability of terminal. As discussed above, when a terminal cannot support some modalities, the content versions of those modalities cannot be presented. The content values of those content versions at the terminal become zero, that means the curves of the non-supported modalities must be removed in the adaptation process. Then we have:

$$V_i = \max\{VM_{ij}\} \qquad (5)$$

where j's are indexes of the supported modalities

4.3.2 Modifying according to order of conversions

Now consider the user preference on modality conversion. The orders of conversions, which are the qualitative level, work similarly to the modality capability. 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 original orders and the user's orders of conversions.
2. Take a modality curve VM_{ij}, compare it with every curve $VM_{ij'}$ that has lower original 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 content item *i*.

4.3.3 Modifying according to 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 5. Note that the sum of d_{ij} is fixed and equal to the maximum content value of the content item i.

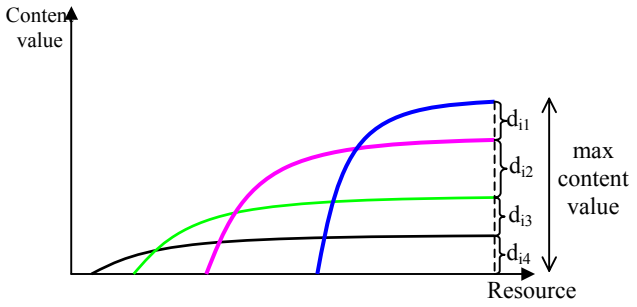

Figure 5: Scaling the curves according to the weights

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 $d'_{ij} = w_{ij} * d_{ij}$.

The relative lengths of these new distances d'_{ij} reflex the user preference, however these lengths still need to be scaled once more to keep the sum of distances unchanged. The final distances are:

$$d_{ij}^{s} = \frac{w_{ij} d_{ij} \sum_{j} d_{ij}}{\sum_{j} w_{ij} d_{ij}} \qquad (6)$$

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

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.

V. RESOURCE ALLOCATION METHOD

Given the content value models and resource constraint, we need to apply a resource allocation method that find the amount of resource for every content item, so as the adapted document has the maximum value to user. The problem of resource allocation has been well tackled for decades. This problem is often solved by two basic methods, the Lagrangian method and the dynamic programming method. The details of these two methods can be found in various references, e.g. [7].

In [6], the Lagrangian method is adopted to find the content versions having appropriate amounts of resource. However, the Lagrangian method is only suitable with the concave content value model. That is why the content value is represented by a single concave curve (e.g. *ln* function); and if the content value model is non-concave, it will be replaced by the concave hull of the model.

In our approach, the upper hull of the overlapped content value model is inherently non-concave. If we replace it by a concave hull, the subtle boundaries between different modalities will disappear. That is, the advantage of overlapped content value model in discriminating the modalities is eliminated. Especially when there are a large number of contents in the document, the differences between the non-concave hull and concave-hull of all content items will be added up, resulting in an adapted document that may be totally different from what expected. So we decide to employ the dynamic programming method for resource allocation [7]. The advantage of the dynamic programming is that it can work with the non-concave content value model, however it has disadvantage of more complexity compared to the Lagrangian method.

VI. 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 item, 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.

Due to the limited space, we just show some example simulated cases here. Figure 6 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 7 shows the adapted document when Rc 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 R_c is reduced as low as 450KBs, the video and all images are converted to audio as shown in figure 8.

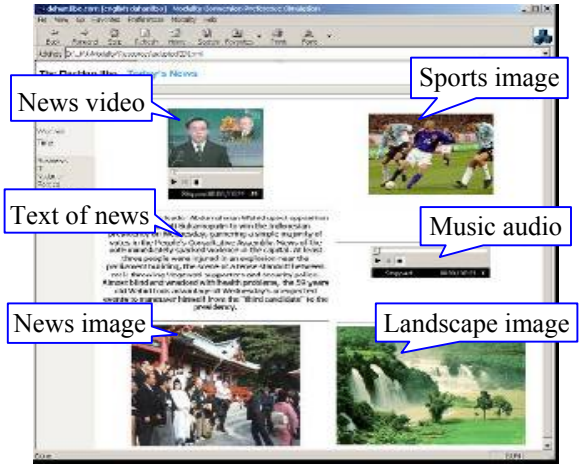

Figure 6: Adaptation with R_c = 1100KBs

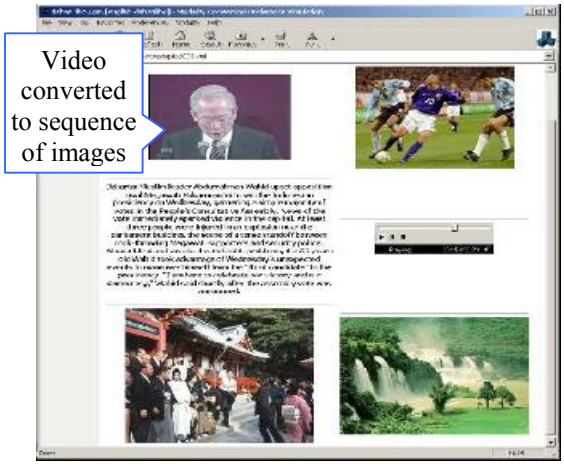

Figure 7: Adaptation when video modality is not supported

Figure 8: Adaptation with R_c = 450KBs

Consider the case of figure 7 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 9. We can see that the video is now converted to audio, not sequence of images.

Figure 9: Adaptation when video modality is not supported and order of video-to-audio is second

Again with the case of figure 8 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 10. 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. Besides, the user

preference is shown to be very helpful for user to customize his content consumption.

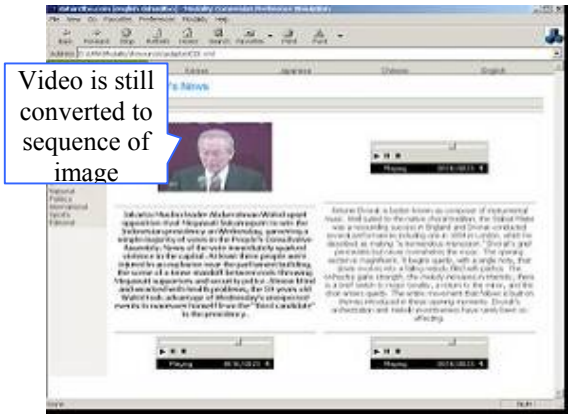

Figure 10: Adaptation when the weight of video-to-image is increased from 1 to 3. (R_c = 450KBs)

VII. CONCLUSIONS

In this paper, we propose a systematic approach for making decisions on modality conversion, an important function in content adaptation. The combination of three integral components - overlapped content value model, modality conversion preference, and the dynamic programming method – is the crucial point of the approach. The proposed approach allows to determine which destination modalities would be, and especially when conversion should be made. It can also handle multiple contents and accommodate a flexible specification of user preference. Our future works will be carried out in two main directions. The first is quantifying the content value of contents within a single modality and across multiple modalities. The second is extending to consider a combination of practical resource constraints such as bandwidth, data size, screen size, etc.

Acknowledgement:

The research reported herein was supported in part by Digital Media Lab.

Reference

[1]. T. C. Thang, Y. J. Jung, Y. M. Ro, J. Nam, M. Kimiaei, J.-C. Dufourd, “CE report on Modality conversion preference”, ISO/IEC JTC1/SC29/WG11 M9495, Pattaya, Thailand, Mar. 2003.

[2]. N. Bjork and C. Christopoulos, "Video Transcoding for Universal multimedia Access", Proceedings of ACM Multimedia 2000, pp. 75-79, Nov. 2000.

[3]. K. Lee, H. S. Chang, S. S. Chun, H. Choi, S. Sull, “Perception-based image transcoding for universal multimedia access, International Conference on Image Processing, pp. 475-478, 2001.

[4]. Kaup, S. Treetasanatavorn, U. Rauschenbach, J. Heuer, “Video analysis for universal multimedia messaging”, Fifth IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 211-215, 2002.

[5]. W. Y. Lum and F. C. M. Lau, “A QoS-sensitive content adaptation system for mobile computing”, Computer Software and Applications Conference, pp. 680–685, 2002.

[6]. 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.

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

[8]. S. Chandra, C. S. Ellis and A. Vahdat, “Application-level differentiated multimedia web services using quality aware transcoding”, IEEE Journal on Selected Areas in Communications, Vol. 18, No. 12, pp. 2544 –2565, Dec. 2000.