Adaptation for Object-based MPEG-4 Content with Multiple Streams

(다중 스트림을 이용한 객체기반 MPEG-4 컨텐트의 적응 기법)

차 경 애*
(Kyung-Ae Cha)

Abstract In this paper, an adaptive algorithm is proposed in streaming MPEG-4 contents with fluctuating resource amount such as throughput of network conditions. In the area of adaptive streaming issue, a lot of researches have been made on how to represent encoded media(such as video) bitstream in scalable way. By contrast, MPEG-4 supports object-based multimedia content which is composed of various types of media streams such as audio, video, image and other graphical elements. Thus, it can be more effective to provide individual media streams in scalable way for streaming object-based content to heterogeneous environment. The proposed method provides the multiple media streams corresponding to an object with different qualities and bit rate in order to support object based scalability to the MPEG-4 content. In addition, an optimal selection of the multiple streams for each object to meet a given constraint is proposed. The selection process is adopted a multiple choice knapsack problem with multi-step selection for the MPEG-4 objects with different scalability levels. The proposed algorithm enforces the optimal selection process to maintain the perceptual qualities of more important objects at the best effort. The experimental results show that the set of selected media stream for presenting objects meets a current transmission condition with more high perceptual quality.

Key Words : Object-based Adaptation, Adaptive streaming, MPEG-4 Content, Multimedia streaming

1. INTRODUCTION

With the rapid development of the broadband communication networks, multimedia streaming is becoming a dominant application. For examples, video clips are often serviced via commercial streaming video players with mobile phone. However, due to the inherent resource sharing nature of the wire or wireless network, it is very difficult to ensure the reserved bandwidth to a certain connection. Hence, for the multimedia streaming, the instantaneous data rate must frequently be tailored to fit the current available bandwidth.

Scalable video coding schemes are able to provide a simple and flexible way for the video streaming by changing the bit rate of original video data. On the other hand, MPEG-4 provides an object based scene representation framework in which multiple objects are represented and rendered in a single scene[1-4].

* 대구대학교 정보통신공학부

An MPEG-4 scene is composed of individual audio-visual objects that are carried in separate Elementary Streams(ES)[5]. Each object in an MPEG-4 scene may be realized by its corresponding ES transmitted from its sender. The elementary streams for the same object may differ in encoded bit-rates, resolutions, etc., so that MPEG-4 also allows for scalability of objects in terms of spatial, temporal scalabilities by layered representation. That is, one single object can be represented with several ESs in different layers.

In this paper, we propose an optimal way of adapting MPEG-4 scenes with multiple media streams for elementary streams of objects. At the presentation, an object included in the content increases the overall quality of the content according to its quality (e.g., signal-to-noise ratio, priority, etc.) and decreases the available resources based on its resource requirement (e.g., encoded bit rate, required buffer size, etc.). Given pre-specified resource constraint (e.g., available network resources, terminal characteristics, user characteristics, natural environment characteristics, etc.) that must not be exceeded, the problem can be defined as finding a subset of objects to be included to maximize the overall quality[6]. P. Batra[6] presents mathematical formulations for modeling object-based scalability using the knapsack problem and its solving algorithm.

As mentioned above, the elementary streams for an object may differ in terms of the encoding characteristics which require different bit rates, buffer sizes, decoding overload and so on. The focus is then made on how to optimally reconstitute an original rich MPEG-4 content according to a given set of constraints by finding optimal set of object's elementary streams to be transmitted.

Based on the model in [6,7], we propose an adaptation method to find a feasible solution for an MPEG-4 content. In our system multiple elementary streams for the objects are provided and selected a set of streams for which the combination of the streams gives the maximization presentation quality of the content under the constraint. To achieve this, we define multi-step selection process which make different maximization conditions and constraints. Each step of the optimization process, the objects with higher priorities are more importantly considered in the selection. To this end, the objects with lower priorities are less likely to be selected so that the entire quality of the adapted content can be maintained with higher perceptual quality.

Several other researches address the issue of resource variation in networks by providing an adaptive architecture with MPEG-4 streams.

A. Mahajan et al.[8] proposed an adaptive MPEG-4 stream service system for wireless networks. Their system is developed based on an end-system aware architecture for improving QoS in wireless networks. They use a layered encoding feature provided by MPEG-4 and H.26x video standards. The adaptation is achieved by using information which is transmitted from a client. The client periodically sends feedback about bandwidth availability and user preference values to the server. Based on the values, the system dynamically selects the appropriate copy of audio or video stream. Thus, the proposed system provides user level adaptive streams with avoiding the network congestion. However, comparing with our system, A. Mahajan et al. treated the adaptation issue more simply. That is, the system selects a stream among pre-encoded streams based on the values of available bandwidth and user preference on scalability levels for specific audio or video streams. In case that an enhancement layer required from a user does not exist, the corresponding enhancement layer stream can not be transmitted. One of the major distinctions on our proposed method is to treat the MPEG-4 scene with multiple objects and to adapt the objects at different scalability levels.

R. S. Ramanujan et al.[9] presented the design of an adaptive service for streaming MPEG video over IP environment. The research aspect of the design is a video adaptation algorithm that performs the adaptation delivering the best quality for the video playback for a given QoS delivered by the underlying infrastructure. In this research, the receiver end maintains a video buffer from which the video player retrieves video frames for playback. The receiver buffers a predetermined number of video frames before allowing the player to start playback. The server end of the streaming service dynamically performs any necessary adaptation to the transmitted video stream. Similar to A. Mahajan, to reduce the computational requirements required to perform on-the-fly adaptations of the video stream, an off-line process is used to convert an MPEG file into a layered MPEG format. Thus, the main performance of this research brings focus into determination of a subset of the stored layers that is transported to the receiver end of the adaptive service. As mentioned above, this is no investigation using the MPEG-4 object based composition scheme for content adaptation like our research. That is, the above research mostly considers the audio or video streams but not scenes that are composed of multiple audio or video streams.

Recently, there has been much research on how to advance FGS. Such an effort is being made within MPEG-21 standardization as scalable video coding(SVC). However, most of research has been focused at scalable representation of a single video stream.

This paper is organized as follows: In Section 2, we describe a conceptual structure of MPEG-4 content in terms of object-based composition. Moreover some preliminaries are described. A way of optimally reconfiguring an MPEG-4 scene is introduced based an optimization algorithm for adapting MPEG-4 bitstream in order to deliver it under a given bandwidth constraint in Section 3. We show experimental results based our proposed methods in Section 4 and we summarize our work and discuss about the future work in Section 5.

2. OVERVIEW OF MPEG-4 CONTENT

2.1 Object-based Scalability of MPEG-4 Content

MPEG-4 enables the author of a multimedia scene to accurately render his composition by parsing a scene description, which effectively defines the structure of the different types of media objects in space and time[10-13]. That is, the scene consists of multiple objects, each of which can be of any type such as rectangular video, arbitrary shaped video, natural audio, synthetic audio, speech, text and graphics, etc. The scene information about the spatio-temporal composition of these audiovisual objects and text are represented as a scene description with their rendering characteristics. Therefore the objects have their own schedules in predefined manners in the scene description.

In addition to the object based scene representation, MPEG-4 also allows for scalability of objects in terms of spatial, temporal and peak signal-to-noise ratio (PSNR) scalabilities by layered representation. Thus, one single object can be represented with several elementary streams in different layers. For example, a video object may be represented with three different elementary streams in different encoded bitstreams such as 128kbps, 256kbps and 512kbps. At the rendering time, the selected stream among the three differently encoded sources will be presented for the video object. Depending on the processing capability, available network bandwidth or a user's specific requirement, appropriate elementary stream can be selected for the video object.

The format of MPEG-4 contents consists of the following data: IOD (initial object descriptor) which carries information about a list of points for Object Descriptor streams; OD (Object Descriptor) is used to achieve the mechanism linking media streams (Elementary Streams) and the scene description; BIFS (BInary Format for Scene) which describes a scene information and rendering characteristics of each object in the scene; OD that contains a list of ESD (Elementary Stream Descriptor); ESD that actually points elementary streams of each object in the MPEG-4 scene and ESs that are encoded bitstreams for the MPEG-4 objects.

The scene description is mapped into the BIFS which is depicted as scene tree. In the BIFS, the nodes are associated with the individual MPEG-4 objects. An OD which may contain several ESDs for multiple scalability levels. Each layer provides the scalability in terms of both encoded bit ratio and perceptual quality. When the bitstream is adapted to specific constraints such as network throughput, an appropriate ES is selected from multiple stream layers for an object.

2.2 Preliminaries

The MPEG-4 scene is composed of a set of audiovisual objects. Thus an MPEG-4 scene which is denoted as can be represented as following set.

M = {O_i | O_i is an audiovisual object, i = 1,2,...,n} where audiovisual objects are detected by parsing the BIFS data which is identified by the Initial Object Descriptor (IOD).

The elementary streams corresponding objects are delivered according to their temporal interval, thus the MPEG-4 stream itself can be fluctuated. The temporal attribute of an object is represented as (*st*, *et*) where *st* and *et* denote the time instances of the start and the end of the temporal duration, respectively[7]. Moreover the object should identify its OD by referencing a unique identification number (*OD_ID*). A priority value (denoted as *P*) can be assigned on each object by content authors manually or content authoring system automatically. Therefore, an object $O_i \in M$ is represented as a tuple (*OD_ID*, *st*, *et*, *P*).

A dot notation(.) is used to indicate the elements of the tuple. The OD of an object O_i which may contain a set of ESDs can be accessed by referencing O_i.*OD_ID*. The ESDs identify the ES and describe the characteristics of the ESs in terms of perceptual quality and encoded bitrate. The ESD is represented as a tuple (*ES_ID*, *BER*, *PQ*). *ES_ID* denotes the unique number identifying the ES layer, while *BER* and *PQ* denote the encoded bitrate and relative perceptual quality of the corresponding ES, respectively.

The quality of an ES is measured as a PSNR value. According to the PSNR, the perceptual quality assigned to each ES layer of an object is in the unit interval [0,1].

We denote O_{ij} as the j^{th} ES layer of the object O_i, which is identified by the j^{th} ESD. The j^{th} ESD also contains the perceptual quality P_{ij} and the encoded bitrate S_{ij}, respectively. P_{ij} and S_{ij} are assigned from the *ESD.PQ* and *ESD.BER*, respectively.

3. OBJECT BASED ADAPTATION OF MPEG-4 CONTENT

This section describes the trade-off method to find a subset of objects' elementary streams included in a scene so as to maximize the presentation quality with network throughput efficiency optimization.

3.1 Optimal Object Configuration Algorithm

In our system, we assume that the elementary streams which participated in an MPEG-4 content are layered in the encoded with different perceptual quality. Layered encoded streams are useful in order to cope with the heterogeneity of user access rates and with the competing traffic in the links between server and client. Moreover these off-line process is used to derive information that can be used by the server to determine which streams to apply to maximize the perceptual quality of the content for the available resources that is supported by network and receiver[14]. At transmission time, the set of objects which are to be transmitted at time t is represented as $M_t = \{O_i.st < t < O_i.et, O_i \in M, i=1,2,...,k,n\}$.

Note that the maximum requirement bit rate of an object $O_i \in M_t$ is identified by S_{i1}. The transmission bitrate of M_t, $BR(M_t)$ is calculated by the equation (1).

$$BR(M_t) = \sum_{O_i \in M_t} BR(O_i), i = 1,2,...,n \quad (1)$$

The objective of the adaptation method is to find a stream layer of each object to maximize the overall quality of transmitted streams satisfy a given constraints. This problem can be categorized in the known as knapsack problem[6]. To solve the problem, our proposed adaptation method is performed in multiple steps: First, the objects are sorted in terms of their priorities and partitioned into two groups, high and low priority sets. An object with a high priority value means that the object priority value exceeds a certain threshold. The low priority objects' priority values don't exceeds a certain threshold. The threshold used to determine priority set is defined on the authoring stage or by user's requests. Here, $HP(M_t)$ is a set of the first ES layers of the objects with high priorities and is defined as following.

$$HP(M_t) = \{ O_{i1} \mid O_i.P \text{ is in the range of high priority}, O_i \in M_t, i = 1,2,k,...,n\}$$

The knapsack problem is solved by checking whether the objects in the high priority set can be all selected into the optimal set or not. If so, then the remaining objects belonging to the low priority set compete for the selection based on the different constraints. We use the backtracking method to solve the proposed multiple choice knapsack problem on each step.
In order to divide the problem into several sub-problems, different feasible sets are defined with predefined variables as follows: N is the number of objects participated in M_t. n_i is the number of layers of the i^{th} object and the index j indicates the ES layer of the object. Moreover the available resource which play as the role of a constraint is denoted as C.

First, a set of objects is selected in the sense that the sum of all perceptual qualities of their ESs is maximized without any object drop under the condition that the sum of the bit rates does not exceed the available bandwidth at a time instance. The following two feasible sets guarantee that all objects are to be appeared in the MPEG-4 scene presentation although some elementary streams are degraded with lower perceptual qualities under the current resource constraint.

A feasible set FSM_t is found by maximizing the sum of the selected ES layer's perceptual quality. In defining following feasible sets, X_{ij} is a decision variable where X_{ij} = 1 means that the j^{th} ES layer of the i^{th} object is selected.

$FSM_t = \{O_{ij} \mid X_{ij} = 1, i = 1,2,k,...,N, j = 1,2,k,...,n_i\}$ satisfies the following constraints, (2):

$$\text{Maximize} \sum_{i=1}^{N}\sum_{j=1}^{n_i} P_{ij}X_{ij}.$$

$$\text{Subject to} \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \leq C \qquad (2)$$

$$\sum_{j=1}^{n_i} X_{ij} = 1 \text{ and}$$

$X_{ij} \in \{0,1\}$ for i = 1,2,k,..., N and j = 1,2,k,..., n_i.

If the above feasible sets is not found, then the second step is to find a set of objects in that the sum of the products of the object priorities and perceptual qualities is maximized without any object drop under the same constraint.

$FSMP_t$ can be found also by maximize the sum of the multiplications of the values of the selected ES layer's perceptual qualities with the priority value of the corresponding object.

$FSMP_t$ = $\{O_{ij} \mid X_{ij} = 1, i = 1,2,k,...,N, j = 1,2,k,...,n_i\}$ satisfies the following constraints,(3):

$$\text{Maximize} \sum_{i=1}^{N}\sum_{j=1}^{n_i} P_{ij}X_{ij}O_i.P$$

$$\text{subject to} \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \leq C \qquad (3)$$

$$\sum_{j=1}^{n_i} X_{ij} = 1 \text{ and}$$

$X_{ij} \in \{0,1\}$ for i = 1,2,k,..., N and j = 1,2,k,..., n_i.

If the above condition can not be met, then object dropping is allowed to find a set of objects in the sense that the sum of all perceptual qualities of their ESs is maximized under the same constraint

The following set is the result to drop some objects from the scene with an effort of maximizing perceptual quality of overall scene.

$FSMZ_t$ = $\{O_{ij} \mid X_{ij} = 1, i = 1,2,k,...,N, j = 1,2,k,...,n_i\}$ satisfies the following constraints,(4):

$$\text{Maximize} \sum_{i=1}^{N}\sum_{j=1}^{n_i} P_{ij}X_{ij}$$

$$\text{Subject to} \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \leq C \qquad (4)$$

$$\sum_{j=1}^{n_i} X_{ij} \leq 1 \text{ and}$$

$X_{ij} \in \{0,1\}$ for i = 1,2,k,..., N and j = 1,2,k,..., n_i.

Finally we find a set of objects by only considering if the sum of the ESs at the first scalability levels (highest qualities) of objects can meet the constraint with object dropping.

$FSPZ_t$ = $\{O_i \mid X_i = 1, i = 1,2,k,...,N\}$ satisfies the following constraints, (5):

$$\text{Maximize} \sum_{i=1}^{N}\sum_{j=1}^{n_i} X_i O_i.P$$

$$\text{subject to} \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{i1}X_i \leq C \qquad (5)$$

$X_{ij} \in \{0,1\}$ for i = 1,2,k,..., N where X_i is the decision variable where X_i = 1 indicates that the i^{th} object is selected with its first ES layer.

A selection procedure for an optimal set of MPEG-4 objects is summarized in Figure 1. Note that the available resource value at time t is denoted as $B(t)$.

step (1): Compare the current MPEG-4 stream's resource requirement and current available resource.

step (2): Find a feasible set FSM_t among the low priority object set while maintaining the ES layers having high priority object as best quality.

step (3): When the feasible set in step (2) is not find, find a feasible set $FSMP_t$ among M_t under the resource constraint as an available bandwidth $B(t)$. In this step, the selection algorithm tries to select elementary stream layers considering object priority values in order to make ESs with high priority be more of high quality.

step (4): When the feasible set in step (3) is not found, find a feasible set $FSPZ_t$ among M_t

```
Input : Mt, Bavail(t)
Output : FSt
Initialize : FSt = ∅
Method
Begin
 If BR(Mt) > Bavail(t)
   if Bavail(t) − BR(HP(Mt)) > 0
       Find FSMt from Mt − HP(Mt)
                with C = Bavail(t) − BR(HP(Mt))
       FSt = FSMt
       if FSt = ∅
          Find FSMPt from Mt with C = Bavail(t)
          if FSMPt = ∅
             Find FSPZt from Mt − HP(Mt)
                    with C = Bavail(t) − BR(HP(Mt))
             if FSPZt = ∅
                Find FSMZt from Mt − HP(Mt)
                   with C = Bavail(t) − BR(HP(Mt))
                FSPZt = FSMZt
             end if
             FSt = FSPZt ∪ HP(Mt)
             exit
          end if
          else
             FSt = FSMPt
             exit
          end else
       end if
       else
          FSt = FSt ∪ HP(Mt)
       end else
   end if
 else
   Find FSMPt from Mt with C = Bavail(t)
   if FSMPt = ∅
        Find FSMPt from HP(Mt)
                    with C = Bavail(t)
        if FSMPt = ∅
            Find FSPZt from HP(Mt)
                    with C = Bavail(t)
            if FSPZt = ∅
                Find FSMZt from HP(Mt)
                        with C = Bavail(t)
                FSt = FSMPt
            end if
            else
                FSt = FSMPt
            end else
        end if
        else
        FSt = FSMPt
        end else
   end if
   else
        FSt = FSMPt
   end else
End
else
        Find FSMPt from Mt with constraint C = Bavail(t)
        if FSMPt = ∅
            Find FSMPt from CP(Mt)
                    with constraint C = Bavail(t)
            if FSMPt = ∅
                Find FSPZt from CP(Mt)
                    with constraint C = Bavail(t)
                if FSPZt = ∅
                    Find FSMZt from CP(Mt)
                    with constraint C = Bavail(t)
                    FSt = FSMPt
                    end if
                    else
                    FSt = FSMPt
                    end else
                    end if
            else
                    FSt = FSMPt
            end else
        end if
        else
            FSt = FSMPt
        end else
End
```

<Fig. 1> A selection procedure for an optimal set of MPEG-4 objects

− $HP(M_t)$ under $B(t)$.

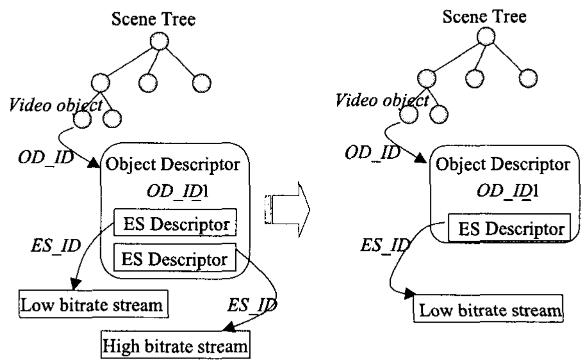
step (5): Find a feasible set $FSMZ_t$ among the low priority object set while maintaining the ES layers having high priority object as best quality when the feasible set in step (4) is not found. In this step, the selection algorithm considers the ES layers having high priority object as best quality by dropping lower priority objects.

step (6): Find a feasible set by the above feasible sets by turns following the steps from (1) to (5) only from the high priority object set because the available resource constraint can not accommodate the object streams in the best qualities with high priority.

3.2 Re- Configuration of Scene Structure

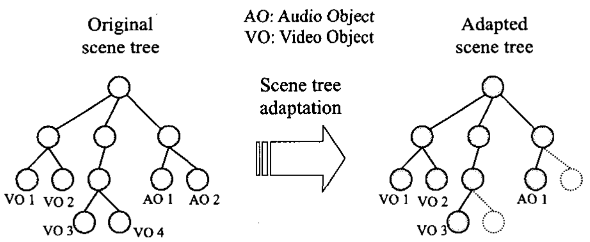
According to the result of the adaptation algorithm, the original scene description and their associated ODs should be modified.

If a video object, O_i is encoded in two different layers with the respective encoded bit rates and one ES in a layer among them is selected, then the ESs in the other layers are removed from the bitstream. The selected ES can be detected by referencing O_i.OD_ID.ES_ID. As described in section 2, the ES_ID identify the corresponding ES and describe the characteristics of the ES. Figure 2 depicts this modification process.



<Fig. 2> Object Descriptor modification

When adapting an original MPEG-4 content by removing some objects, the scene composition information should be changed accordingly. Figure 3 shows a scene tree adaptation concept in which two video objects are removed after the scene tree adaptation. At the same time, the all regarding the object are removed from the original MPEG-4 binary resource. Moreover the addressing points of the ODs of the IOD and ESDs of the ODs should be corrected accordingly. The IOD is updated by pointing the remaining ODs in the OD stream.



<Fig. 3> Scene Tree modification

4. Experiments

This section presents the evaluation results and analysis of the proposed method of finding an optimal set of MPEG-4 objects for a given constraint.

For the experiment, we constitute example MPEG-4 content#1 as shown in Table 1.

<Table 1> Object composition and Temporal scenario of MPEG-4 content #1

stream type		H.263 video	G.723	H.263 video	mp4 audio	H.263 video	G.723	H.263 video	G.723
Object ID		1	2	3	4	5	6	7	8
Temporal Attributes	st	1	1	1	1	1	1	1	1
	et	20	20	20	20	20	20	20	20

Table 2 shows multiple streams for objects in the MPEG-4 content#1. Note that the values of relative perceptual quality (denoted as RPQ) for an object are assigned in proportion to the object's priority at the authoring stage. This RPQ is used as a measure that indicates the quality of the ES layers.

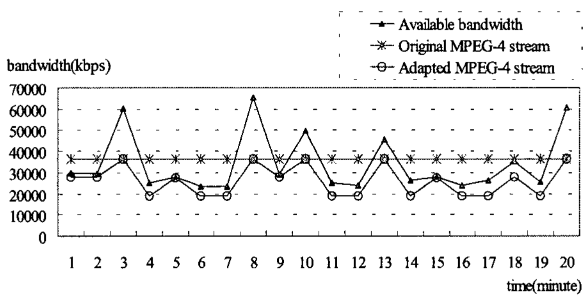
<Table 2> Multiple streams for example MPEG-4 content #1

object priority			high				low			
			8	8	7	7	6	6	5	5
stream type			H.263 video	G.723	H.263 video	mp4 audio	H.263 video	G.723	H.263 video	G.723
stream layer	1	bit-rate	9124 Kbps	6.3 Kbps	9124 Kbps	56 Kbps	9124 Kbps	6.3 Kbps	9124 Kbps	6.3 Kbps
		RPQ	1	1	1	1	1	1	1	1
	2	bit-rate	505.6 Kbps	5.3 Kbps	522.4 Kbps	48 Kbps	505.6 Kbps	5.3 Kbps	505.6 Kbps	5.3 Kbps
		RPQ	0.8	0.9	0.8	0.9	0.8	0.9	0.8	0.9
	3	bit-rate	112.1 Kbps		105.47 Kbps	32 Kbps	112.1 Kbps		112.1 Kbps	
		RPQ	0.7		0.7	0.8	0.7		0.7	
	4	bit-rate	56.7 Kbps		52.76 Kbps	28 Kbps	56.7 Kbps		56.7 Kbps	
		RPQ	0.6		0.6	0.7	0.6		0.6	
	5	bit-rate	22.81 Kbps		21.83 Kbps	24 Kbps	22.81 Kbps		22.81 Kbps	
		RPQ	0.5		0.5	0.6	0.5		0.5	

(RPQ : Relative perceptual quality)

The RPQ value on each ES layer is assigned a number between 0 and 1 where it is set to 1 for the ES layers with the highest encoded bitrate for its object. For example, the H.263 video stream with the priority value 8 is encoded in five scalability levels with different bitrates where the higher the bit rate is the higher RPQ value is. The top four ranked objects in their object priority are categorized into the high object priority level and the remaining four objects are considered to belong to the low object priority level. If an object has a high priority, an ES with a high bitrate is likely to be selected and to belong to the set that contains objects to be delivered at high quality levels.

Figure 4 shows that the variations of the available bandwidth in time for a delivery network in the line with triangles, and the required bandwidth to deliver the original MPEG-4 content #1 in the line with asterisks. After adaptation by our proposed method with the available network bandwidth as a constraint, the required bandwidth is plotted in a line with circles. During the adaptation process, a set of ESs is optimally selected which does not exceed the available network bandwidth in the way that maximum quality is obtained under the bandwidth constraint. It can be noticed that the bandwidth line after adaptation lies below the available bandwidth of the network so that the adapted MPEG-4 binary resource can be delivered under the bandwidth constraint of the network characteristics without transmission delay.

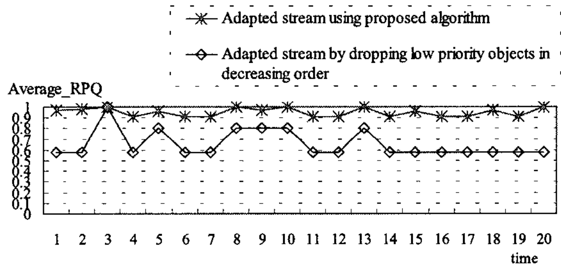


<Fig. 4> Bandwidth : Available bandwidth, original MPEG-4 stream of MPEG-4 content 2, adapted MPEG-4 stream of MPEG-4 content #1

Figure 5 shows the perceptual quality of the adapted MPEG-4 binary resource and that of the streams by performing the method of dropping off the objects with relatively low object priority values under a bandwidth constraint. The value of average relative perceptual quality is calculated the equation (6).

$$Average\_RPQ = \sum_{i=1}^{n} O_i.P \times RPQ_i \quad \sum_{i=1}^{n} O_i.P \qquad (6)$$

Although an object has a low object priority value, our method selects an ES with a low bitrate for that object, not dropping the object off from the MPEG-4 bitstream. So, the resulting perceptual quality of the adapted bitstream fluctuates in a small scale for our method.

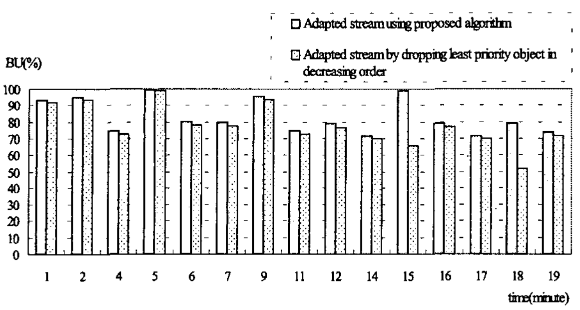


<Fig. 5> Average relative perceptual quality : adapted streams of MPEG-4 content #1

The bandwidth utilization of adapted streams, $BU(t)$ at time t, can be evaluated by the following. $B_{avail}(t)$ is represented as available bandwidth at time t.

$$BU(t) = 100 \times BR(t) \quad B_{avail}(t) \qquad (7)$$

Figure 6 depicts the bandwidth utilization (calculated by equation (7)) with adapted streams using proposed algorithm and the others by dropping least priority objects in decreasing order. As a result, the performance of bandwidth utilization of the adapted streams using multiple streams provides very high efficiency which may provide more better result if the multiple streams of each object more smaller bit rate scale.



<Fig. 6> Bandwidth utilization : adapted stream of MPEG-4 content #1

While the content#1 has the constant bitrate for its own stream, the second example content, denoted as MPEG-4 content#2 is composed with the fluctuation of traffic characteristics in time by the temporal attributes of the objects.
Table 3 and table 4 show the object types and their temporal scenario of the MPEG-4 content#2 and the multiple streams for the content, respectively.

<Table 3> Object composition and Temporal scenario of example MPEG-4 content #2

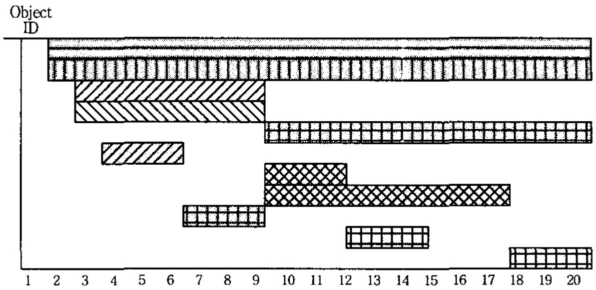
stream type		mp2 video	G.723	AAC	mp4 video	H.263	H.263	H.263	H.261	jpeg	jpeg	jpeg
Object ID		1	2	3	4	5	6	7	8	9	10	11
Temporal Attributes	st	1	1	2	2	9	3	9	9	6	12	18
	et	20	20	8	8	20	5	11	17	8	14	20

<Table 4> Multiple streams of the MPEG-4 content #2

object priority		high						low				
		11	10	9	8	7	6	5	4	3	2	1
stream type		mp2 video	G.723	AAC	mp4 video	H.263	H.263	H.263	H.261	jpeg	jpeg	jpeg
stream layer 1	bit-rate	150 Kbps	6.3 Kbps	64 Kbps	112.1 Kbps	256 Kbps	256 Kbps	256 Kbps	128 Kbps	1 Kbyte	1 Kbyte	1 Kbyte
	RPQ	1	1	1	1	1	1	1	1	1	1	1
stream layer 2	bit-rate	120 Kbps	5.3 Kbps	32 Kbps	56.7 Kbps	192 Kbps	192 Kbps	192 Kbps	64 Kbps	0.9 Kbyte	0.9 Kbyte	0.9 Kbyte
	RPQ	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
stream layer 3	bit-rate	100 Kbps		16 Kbps	22.81 Kbps	128 Kbps	128 Kbps	128 Kbps		0.7 Kbyte	0.7 Kbyte	0.7 Kbyte
	RPQ	0.8		0.8	0.8	0.8	0.8	0.8		0.8	0.8	0.8
stream layer 4	bit-rate			8 Kbps		64 Kbps	64 Kbps	64 Kbps		0.5 Kbyte	0.5 Kbyte	0.5 Kbyte
	RPQ			0.7		0.7	0.7	0.7		0.7	0.7	0.7
stream layer 5	bit-rate									0.25 Kbyte	0.25 Kbyte	0.25 Kbyte
	RPQ									0.6	0.6	0.6

(RPQ : Relative perceptual quality)

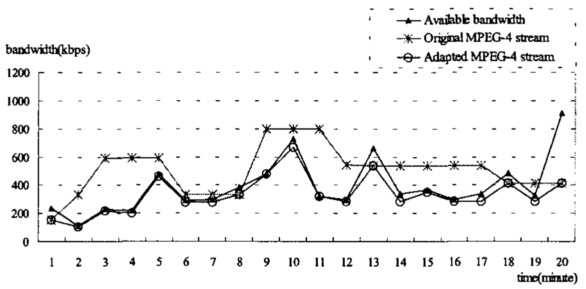
Figure 7 shows as timelines the temporal scenarios of the objects appearing in the MPEG-4 content #2.



<Fig. 7> Timeline of MPEG-4 content #2

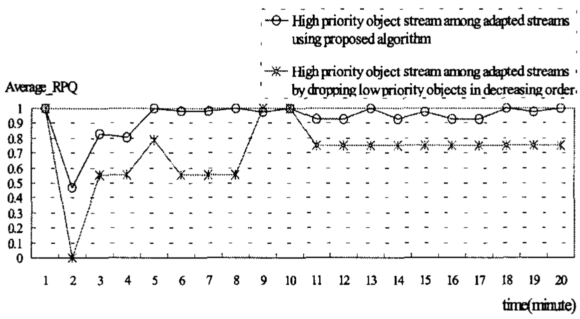
Now, we evaluate the scenario of MPEG-4 content#2 over lower bit rate bandwidth environment than the above evaluation environment for MPEG-4 content#1.

The adaptation result is shown in figure 8 that represents the bitrates of the adapted stream do not exceed the available bandwidth through the transmission time. For the bandwidth as a constraint, the traffic characteristics of the MPEG-4 stream vary in time.



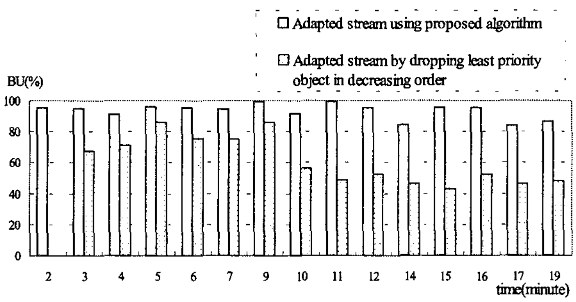
<Fig. 8> Bandwidth : Available bandwidth, original MPEG-4 stream of MPEG-4 content 2, adapted MPEG-4 stream of MPEG-4 content #2

The proposed algorithm enforces the optimal selection process of the objects to maintain the perceptual qualities of more important objects at the best effort. It can be observed from Figure 9 that the quality of ES with high object priority is consistently maintained under a network with varying bandwidth except the beginning time. This is because the objects with high object priority are likely to be included as ESs with high bitrates into the adapted MPEG-4 binary resource. Figure 9 shows the perceptual qualities for the high priority objects selected into the adapted bitstreams for the MPEG-4 content #2.



<Fig. 9> Average relative perceptual quality : adapted streams of high priority objects of MPEG-4 content #2

Since the objects are dropped off at beginning time due to the very low available bandwidth, it can be observed that the perceptual quality curve sharply decreases at that time. However the bandwidth utilization value shown in figure 10 represents that the adapted bitstream uses the network throughput efficiently at the second time instance. In figure 10, time instance 1, 8, 13, 18 and 20 are skipped because the available bandwidths are higher than the required bitrate of the MPEG-4 stream at those times.



<Fig. 10> Bandwidth utilization : adapted stream of MPEG-4 content #2

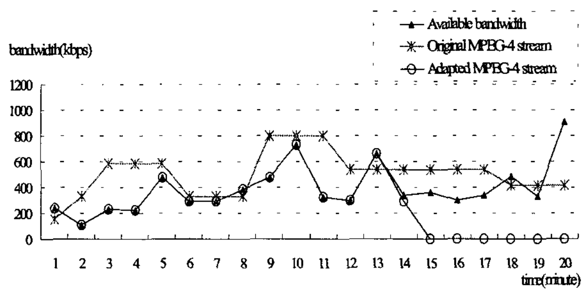
In order for the adaptation algorithm to be more viable, we consider the following two implementing issues : The selected stream layer of objects is maintained through the object's transmission period. The current bandwidth is estimated using previous evaluated values.

We investigate the MPEG-4 content#2 with the above assumption. Table 5 represents the selected stream layer of each object and the average perceptual quality of total media streams. The perceptual quality of most objects are transmitted with their original thus, the average relative perceptual quality is evaluated very high. Note that the highest average perceptual quality is 1.

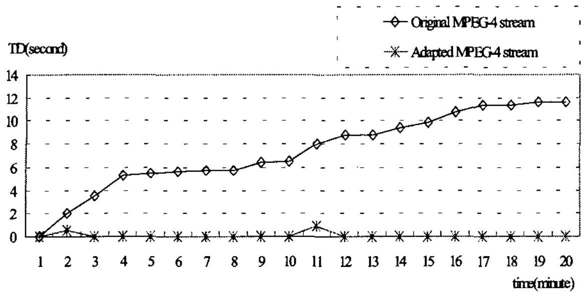
<Table 5> Average perceptual quality of the adapted MPEG-4 content #2

Object ID	1	2	3	4	5	6	7	8	9	10	11	Average _RPQ
Priority	11	10	9	8	7	6	5	4	3	2	1	0.8106
RPQ	1	1	0.9	0.8	0.7	0	0.7	0.9	1	1	1	

Figure 11 depicts the comparison of transmission bit rate of adapted stream, available bandwidth and original MPEG-4 stream. As we can see through the figure, the adapted streams are delivered using the full available bandwidth. Thus the transmission delay can be occurred. Figure 12 shows the estimated transmission delay when the adapted streams are delivered.

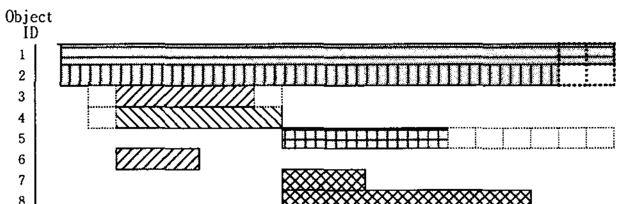


<Fig. 11> Available bandwidth, original stream of MPEG-4 content#2, adapted stream of MPEG-4 content#2



<Fig. 12> Transmission delay : adapted stream of MPEG-4 content#2 vs. original stream of MPEG-4 content #2

As we can see through the experimental results, the adapted streams are delivered using the full available bandwidth. Thus the transmission delay can be occurred as shown in figure 12. However, on the terms of the trade-off, the portions of timelines for object ID 1,2,3,4,5 and 10 in figure 13 which depicts the object timeline of the adapted content. The portions drawn with dotted line means that the media stream corresponding to the object is not transmitted any more. That is, the streams are pre-patched so as to maintain their quality and use available bandwidth fully even though occurring transmission delay. As a result, we can see that our proposed adaptation method make an effort of keeping the more high quality streams for the high priority objects.



<Fig. 13> Timeline of the Adapted objects for MPEG-4 content #2

5. Conclusion

In this paper, we propose an optimal selection of the objects for MPEG-4 bitstream adaptation to meet a given constraint. We adopt a multi-step selection for the objects with different scalability levels for the MPEG-4 bitstream. At each step, the proposed algorithm enforces the optimal selection process of the objects to maintain the perceptual qualities of more important objects at the best effort. Thus, the bitrate of the content can be adapted(not exceeded) to the given bandwidth with maintaining perceptual quality of the decoded scene as high level.

For the future works, we are going to consider more efficient mechanism to adapt contents when huge number of streaming objects are participated in the transmitted content.

참 고 문 헌

[1] WG11(MPEG), MPEG-4 Overview document, ISO/IEC JTC1/SC29/WG11 N4030, March 2001.

[2] ISO/IEC 14496-1:1999 Coding of audio-visual Objects - Part 1: Systems ISO/IEC JTC1/SC29/WG11 N2501, 1999.

[3] ISO/IEC 14496-1:2000 Coding of Audio-Visual Objects - Part 1:ISO/IEC JTC1/SC29/WG11 N2501, 2000.

[4] WG11(MPEG) MPEG-4 Overview (V.21 - Jeju Version) document, ISO/IEC JTC1/SC29/WG11 N4668, March 2002.

[5] C. Herpel and A. Eleftheriadis : "MPEG-4 Systems: elementary stream management, "Signal Processing: Image Communication, 15 (4-5), pp. 299-320, January 2000.

[6] P. Batra, "Modeling and efficient optimization for object-based scalability and some related problems," IEEE Trans. Image Processing, vol. 9, no. 10, pp. 1677-1692, October 2000.

[7] Kyung-Ae Cha and Sangwook Kim, "Adaptive Scheme for Streaming MPEG-4 contents to Various Devices," IEEE Transactions on Consumer Electronics, vol. 49, no. 4, pp. 1061-1066, November 2003.

[8] A. Mahajan, P. Mundur and A. Joshi, "Adaptive multimedia system architecture for improving QoS in wireless networks," LNCS2532, Springer, December, 2002, pp. 713-719.

[9] R.S. Ramanujan, J.A. Newhouse, M.N. Kaddoura, A. Ahamad, E.R. Chartier and K.J. Thurber, "Adaptive streaming of MPEG video over IP Networks," IEEE Conference on Local Computer Networks (LCN'97), pp. 398-409, 2-5 November 1997.

[10] J. Kneip, B. Schmale and H. Mller, "Applying and implementing the MPEG-4 multimedia standard," IEEE Micro, vol. 19, no. 6, pp. 64-74, 1999.

[11] R. Koenen, "Profiles and levels in MPEG-4: approach and" Signal Processing: Image Communication, vol. 15,Issues 4-5, pp. 463-478, January 2000.

[12] B. Prabhakaran, "Multimedia authoring and presentation techniques,"Multimedia Systems, vol. 8, issue 3, pp. 157, 2000.

[13] K.A. Cha and S. Kim, "MPEG-4 STUDIO: an object-basedsystem for MPEG-4 contents," Multimedia Tools andvol. 25, no.1, pp. 111- 131, January 2005.

[14] D. Turner and K.W. Ross, "Optimal streaming of a synchronizedpresentation with layered Objects," IEEE International Conference on Multimedia and Expo, New York, pp. 263-266, July 2000.

차 경 애 (Kyung-Ae Cha)

- 정회원
- 1996년 2월 경북대학교 컴퓨터과학과 (이학사)
- 1999년 2월 경북대학교 컴퓨터과학과 (이학석사)
- 2003년 8월 경북대학교 컴퓨터과학과 (이학박사)
- 2001년 3월 ~ 2004년 2월 경북대학교 초빙교수
- 2004년 3월 ~ 2005년 2월 한국정보통신대학교 연구교수
- 2005년 3월 ~ 현재 대구대학교 정보통신공학부 전임강사
- 관심분야 : 멀티미디어시스템, 디지털방송, 컨텐트 저작 및 스트리밍