JOURNAL OF KOREA MULTIMEDIA SOCIETY, VOL. 9, NO. 12, DECEMBER 2006(pp. 1560-1571)

Adaptive Transcoding for Object-based MPEG-4 Scene using Optimal Configuration of Objects

Kyung-Ae Cha[†]

ABSTRACT

In order to transmit multimedia streams over the network with a timely changing channel bandwidth such as Internet, scalable video coding schemes have been studied to represent video in flexible bitstream. Much research has been made on how to represent encoded media(such as video) bitstream in scalable ways. In this paper, we propose an optimal selection of the objects for MPEG-4 bitstream adaptation to meet a given constraint. We adopt a multiple choice knapsack problem with multi-step selection for the MPEG-4 objects with different bit-rate scaling levels in the MPEG-4 bitstream. The bitstream adaptation based on the optimal selection result is then to fetch the necessary parts of the MPEG-4 bitstream to constitute an adapted version of the original MPEG-4 binary resource. The experiment results show that the optimal selection of MPEG-4 objects for a given constraint can promisingly be made which meets the given constraint.

Keywords: Object-based Transcoding, MPEG-4 Scene, Object-based Adaptation

1. INTRODUCTION

One of the problems encountered when a multimedia service is performed in a heterogeneous environment is the variation in the receiver's capability or the fluctuation of network throughput[1-6]. Therefore, providing an adaptive transmission scheme in the server is essential in rendering the appropriate contents to the specific receiver. This paper considers to transmit MPEG-4 contents making highly efficient use of the available resources.

MPEG-4 provides an object based scene representation framework in which multiple objects are represented and rendered in a single scene[8-15]. The scene information about the spatio-temporal composition of the audiovisual objects and text are represented as BIFS(Binary format for scene) with their rendering characteristics. The individual media corresponding to objects are carried in separate Elementary Streams(ES). The elementary streams for the same object may differ in encoded bit-rates, resolutions, etc. so that MPEG-4 also allows for transcoding of objects in terms of spatial, temporal and PSNR scalabilities by layered representation. That is, one single object can be represented with several elementary streams in different layers. So it is required that an optimal set of MPEG-4 objects need to be selected to meet given resource constraints.

At the presentation, an object included in the content increases the overall quality of the content according to its quality and decreases the available resources based on its resource requirement. Given pre-specified resource constraint that must not be exceeded, the problem can be defined as finding a

---

※ Corresponding Author : Kyung-Ae Cha, Address : (712-714) School of Computer Communication Engineering, Daegu University 15 Jinryang, Gyeongsan, Gyeongbuk, Korea, TEL : +82-850-6641, FAX : +82-850-6629, E-mail : chaka@daegu.ac.kr

Receipt date : Jane 7, 2006, Approval date : Nov. 27, 2006

[†] School of Computer and Communication Engineering, Daegu University, Korea

※ This work was supported in part by Daegu University Research Grant, 2006.

subset of objects to be included to maximize the overall quality[1]. The focus is then made on how to adaptively reconstitute an original rich MPEG-4 content according to a given set of constraints such as available network resources, terminal characteristics, user characteristics and natural environment characteristics etc. In this paper, the problem which is explained in [1] is extended to include multiple streams of each object to satisfy a given resource constraint such as network bandwidth.

Several researches[2-7] address the issue of resource variation in networks by providing an adaptive architecture with MPEG-4 streams.

A. Vetro et al. introduces a new framework for video content delivery that is based on the transcoding of multiple video objects[2]. Given the object-based framework, they present two kinds of strategies for supporting object-based transcoding in order to satisfy network conditions or user requirements in various ways. One is dynamic programming approach and the other is an approach that is based on available meta-data. They show the simulations with these two approaches provide insight regarding the bit allocation among objects and illustrates the tradeoffs that can be made in adapting the content. On the other side, we design the object-based transcoding scheme based on the multi-step media stream selection algorithm not using extra meta-data. Moreover we consider the object priorities to maintain higher priority objects to be served with higher quality.

Similar to this research, Goor, S.A. et al. propose an adaptive streaming system based on video object transcoding of MPEG-4 by applying priorities to the individual objects[3]. This research just mentions the MPEG-4 object-based features to be used the transcoding system.

One of the research considering the layered streams to transmit adaptive content is presented by A. Mahajan et al. They 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[4]. 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. Based on the client's feedback, 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. The system simply selects a stream among pre-encoded streams based on the values of available bandwidth and user preference. The research considers the audio or video streams but not scenes that are composed of multiple audio or video streams like our research.

This paper is organized as follows: In Section 2, we describe objects' features of MPEG-4 content and some definitions which are used to design the proposed adaptive algorithm. A way of optimally reconfiguring an MPEG-4 scene is introduced based an our algorithm for adapting MPEG-4 bitstream in order to deliver it under a given bandwidth constraint in Section 3. We show experimental results in Section 4 and we summarize our work and discuss about the future work in Section 5.

2. OPTIMAL CONFIGURATION OF MPEG-4 STREAM USING MULTIPLE ELEMENTARY STREAMS

In addition to the object based scene representation, MPEG-4 also allows for object-based transcoding in terms of spatial, temporal and peak signal-to-noise ratio (PSNR) scalabilities by layered representation. For examples, 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.

Given pre-specified resource constraints such as bandwidth and buffer size, *Batra* proposed a method of finding an optimal subset of the MPEG-4 objects to maximize a given figure of merit[1]. This work had been focused in solving the multiple choice knapsack problem (MCKP) in an efficient way for the objects at different perceptual quality levels in the MPEG-4 scene. In our approach, we consider object priorities in selecting the objects. Therefore, 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. Notice that the previous method is to select a set of objects for which the object combination gives the maximization condition under the constraint without considering the relative importance of the objects.

Figure 1 shows a conceptual structure of an MPEG-4 bitstream. Each ES layer provides different bit-rate scaling 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. In Figure 1, the ES 2 is selected for the video object 1, and ES 1 for the video object 2. The ESs that are not selected will be removed from the bitstream, so that the optimal ESs set to adapt given network bandwidth can be transmitted to a receiver without transmission delay.

3. OBJECT BASED TRANSCODING OF MPEG- 4 SCENE

In this section, we describe an optimization technique to adapt an MPEG-4 scene with multiple objects for which some of objects may be represented in an object-based transcoding at different layers.

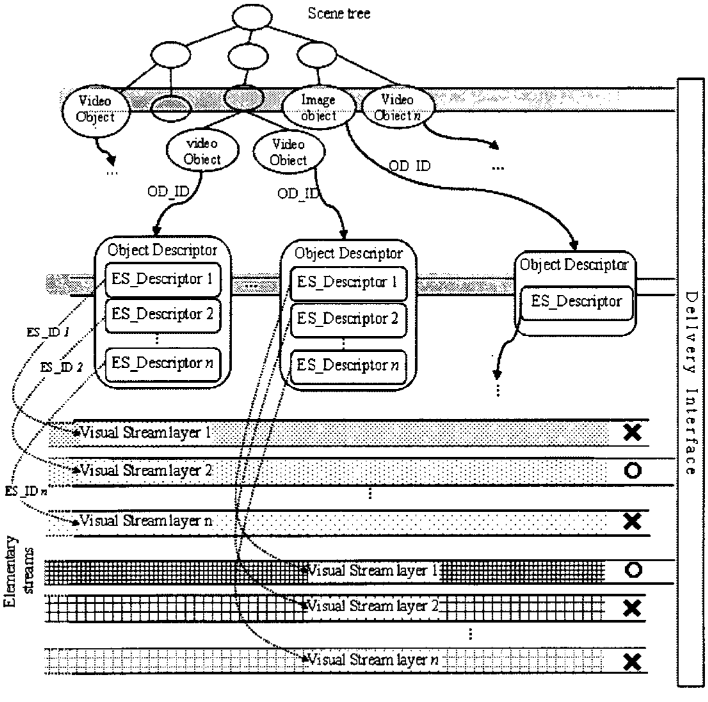

Fig. 1. Conceptual Structure of an MPEG-4 Bitstream.

3.1 MPEG-4 Scene Structure

The MPEG-4 scene is composed of a set of audio-visual objects. Thus an MPEG-4 scene M can be represented as a set $\{O_i | O_i \, is\, an\, audiovisual\, object, i = 1,2,...,n\}$ 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. Moreover the object should identify its Object Descriptor(OD) by referencing a unique identification number(denoted as *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]. All those values that are used to compute constraints can easily be read from the bitstream syntax description such as ES descriptors.

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 set from the *ESD.PQ* and *ESD.BER*, respectively. Figure 2 depicts the structure of ODs with ESDs.

3.2 Optimal Object Configuration Algorithm

At every transmission time, the set of objects which are to be transmitted at time t is represented as $M_t = \{O_i.st < t < O_i.et, O_i \in M, i = 1,2,...,k,n\}$. When an elementary stream is transmitted, each bit rate of the corresponding object is calculated using the following equation (1).

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

In equation (1), $BR(O_i)$ means the maximum requirement bit rate of an object $O_i \in M$ which is identified by S_{i1}. As defined 3.1, S_{ij} is the encoded bitrate of the j^{th} layer of the object O_i, thus S_{i1} means the first ES layer for the object O_i that is the original(highest) quality of ES requiring the maximum bit rate for the object O_i. $BR(O_i)$ is calculated by following equation (2).

$$BR(O_i) = \begin{cases} S_{i1}, \text{if } the\ object\ type\ is\ image \text{ or } audio \\ (t - O_i.st_I)S_{i1} - state(O_i)(t - O_i.st_I - 1)S_{i1}, \\ \qquad\qquad \text{if } the\ object\ type\ is\ video \end{cases} \tag{2}$$

where st_I is the starting time of I frame just before the current time instance t and

$$state(O_i) = \begin{cases} 0, \text{if } O_i\ is \in active \\ 1, \text{if } O_i\ is\ active \end{cases} \tag{2-1}$$

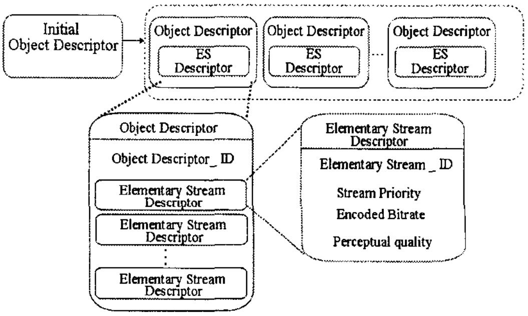

Fig. 2. Structures of ODs with ESDs.

The equation (2-1), $state(O_i)$ means that the status of the streaming objects such as video, audio and image.

Video stream is already started to be transmitted, then the status of the corresponding object node in the scene tree is changed to be *active* . So the result of $state(O_i)$ is 1. In the equation (2), $(t-O_i.st_I)S_{i1}-state(O_i)(t-O_i.st_I-1)S_{i1}$ is calculated as $(t-O_i.st_I)S_{i1}-(t-O_i.st_I-1)S_{i1}=S_{i1}$ that means the encoded bit rate of the video stream. On the other hand, the video stream is not transmitted on its starting time (*inactive*), the accumulated bit rate should be transmitted when the bandwidth is available. In this case $state(O_i)$ is 0, so $BR(O_i)$ is calculated as $(t-O_i.st_I)S_{i1}$ that means the accumulated bitrate of the video stream from its starting time.

For image object cases, the image data are transmitted at one single time and stored at the receiver side for their presentation times. Therefore, the sender does not need to care about the image type object after transmission. In the case of audio objects, the delayed stream that failed to be available at its starting rendering time is only necessary to be transmitted with removal of the delayed amount.

The objective of adaptation is to find a stream layer of each object to maximize the overall quality of transmitted streams. To do this, our proposed MCKP is solved in multiple steps: First, the objects are sorted in terms of their priorities and partitioned into two groups, high and low priority sets. 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 MCKP. If the objects in the high priority set cannot be all selected, they compete for the selection based on the MCKP. We use the backtracking method to solve the proposed multi-step MCKP.

In order to divide the problem into several sub-problems, different feasible sets are defined.

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.

*F*easible *S*et *M*aximizing the selected streams' perceptual quality at time t is defined as FSM_t. FSM_t is found by maximizing the sum of the selected ES layer's perceptual quality. Here the perceptual quality of an ES layer is one of the value from 0 to 1. As defined in 3.1, according to the PSNR of the ES, the perceptual quality assigned to each ES layer is in the unit interval $[0, 1]$.

*F*easible *S*et *M*aximizing *P*riority of the selected streams is defined as $FSMP_t$. $FSMP_t$ can be found 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.

If the above conditions 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. Thus, the feasible set is defined by following the concept of the Zero-one knapsack problems where there are only two cases in selection of the ES for an object, selecting of original(highest) quality of stream layer(one) or dropping the object(zero). The next two sets are defined considering these cases.

*F*easible *S*et *M*aximizing perceptual quality of the selected streaming by following the *Z*ero-one knapsack problem at time t is defined as $FSMZ_t$. The result from $FSMZ_t$ is to drop some objects from the scene with an effort of maximizing perceptual quality of overall scene.

We define a set of objects by only considering if the sum of the ESs at the highest qualities of objects can meet the constraint with object dropping. *F*easible *S*et maximizing *P*erceptual qualities of the selected streams by following

Zero-one knapsack problem at time t is defined as $FSPZ_t$.

The followings are represented the formal definition of above defined feasible sets. Here X_{ij} is a decision variable where $X_{ij}=1$ means that the j^{th} ES layer of the i^{th} object is selected. 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.

$FSM_t = \{O_{ij} | X_{ij} = 1,\ i = 1,2,k,\dots,N,\ j = 1,2,k,\dots,n_i\}$ satisfies the following constraints.

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

$$\text{subject to} \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \le C$$

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

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

$FSMP_t = \{O_{ij} | X_{ij} = 1,\ i = 1,2,k,\dots,N,\ j = 1,2,k,\dots,n_i\}$ satisfies the following constraints.

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

$$\text{subject to} \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \le C$$

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

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

The following sets are the results to drop some objects from the scene.

$FSMZ_t = \{O_{ij} | X_{ij} = 1,\ i = 1,2,k,\dots,N,\ j = 1,2,k,\dots,n_i\}$ satisfies the following constraints.

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

$$\text{subject to} \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{ij}X_{ij} \le C$$

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

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

$FSPZ_t = \{O_i | X_i = 1,\ i = 1,2,k,\dots,N\}$ satisfies the following constraints.

$$\text{Maximize} \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i} X_i\, O_i{\cdot}p$$

$$\text{subject to} \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i} S_{i1}X_i \le C$$

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

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 3. Note that the available resource value at time t is denoted as $B(t)$. $HP(M_t)$ is a set of the first ES layers of the objects with high priorities where $HP(M_t) = \{O_{i1} |\ O_i{\cdot}p$ is in the range of high priority, $O_i \in M_t$, $i = 1,2,\dots,n\}$. Here, an object with a high priority value means that the object priority value exceeds a certain threshold.

The followings explain the optimal object selection algorithm in more details.

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 $M_t - HP(M_t)$ under $B(t)$.

step (5): Find a feasible set $FSMZ_t$ among the

```
Input : $M_t$, $B(t)$
Output : $FS_t$
Initialize : $FS_t = \varnothing$
Method
Begin
calculate $BR(M_t)$
If $BR(M_t) > B(t)$
    if $B(t) - BR(HP(M_t)) > 0$
        Find $FSM_t$ from $M_t - HP(M_t)$
        with constraint $C = B(t)-BR(HP(M_t))$
        $FS_t = FSM_t$
        if $FS_t = \varnothing$
            Find $FSMP_t$ from $M_t$
            with constraint $C = B(t)$
            if $FSMP_t = \varnothing$
                Find $FSPZ_t$ from $M_t - HP(M_t)$
                if $FSPZ_t = \varnothing$
                    Find $FSMZ_t$ from $M_t$-$HP(M_t)$
                    $FSPZ_t = FSMZ_t$
                end if
                $FS_t = FSPZ_t \cup HP(M_t)$
                exit
            end if
            else
                $FS_t = FSMP_t$
                exit
            end else
        end if
        else
            $FS_t = FS_t \cup HP(M_t)$
        end else
    end if
    else
        Find $FSMP_t$ from $HP(M_t)$
        with constraint $C = B(t)$
        $FS_t = FSMP_t$
        if $FS_t = \varnothing$
            Find $FSM_t$ from $HP(M_t)$ with
            constraint $C = B(t)$
            if $FSM_t = \varnothing$
                Find $FSMZ_t$ from $HP(M_t)$
                if $FSMZ_t = \varnothing$
                    Find $FSPZ_t$ from $HP(M_t)$
                    $FSMZ_t = FSPZ_t$
                end if
                $FS_t = FSMZ_t$
                exit
            end if
            else
                $FS_t = FSPZ_t$
                exit
            end else
        end if
    end else
End
```

Fig. 3. A selection procedure for an optimal set of MPEG-4 objects.

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, $FSPZ_t$ by the following steps 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.

4. SIMULATION RESULTS

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 an MPEG-4 scene with 6 video streams, 2 audio streams and 3 image data as shown in Table 1. For the bandwidth as a constraint, the traffic characteristics of the MPEG-4 stream vary in time.

Each object in the MPEG-4 content may contain multiple ESs with different qualities. In Table 2, the values of the perceptual quality and encoded bit rates are tabulated for the objects with several

Table 1. Example of scenario with an MPEG-4 content

Object ID	Stream type	Temporal intervals	
		Start times (*st*)	durations (*d*)
1	mp2 video	0	20
2	G.723 Audio	0	20
3	AAC Audio	1	7
4	H.263 video	1	7
5	H.263 video	8	12
6	H.263 video	2	3
7	H.263 video	8	3
8	H.261 video	8	9
9	jpg image	5	3
10	jpg image	11	3
11	jpg image	17	3

bit-rate scaling levels for the MPEG-4 content. Note that the values of *RPQ* for an object are assigned in proportion to the object's priority at the authoring stage. The *RPQ* value on each ES layer is assigned 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 mp2 video stream, with the priority value 11 is encoded in three different quality levels with different bit rates where the higher the bit rate is the higher RPQ value is.

The object priority level is determined by the object priority values. The top five ranked objects in their object priority are categorized into the *high* priority level and the remaining six objects are considered to belong to the *low* 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 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. Thus, high network throughput can be achieved efficiently. If the original MPEG-4 binary resource is not adapted to the available network bandwidth, it might cause transmission delays because the bitrates of the original stream exceed the available bandwidth at some transmission time instances (4, 5, 9, 10, 11, 12, 14, 15, 16 and 17).

In order to estimate the perceptual quality of the adapted MPEG-4 streams which will be decoded at the receiver, we refer to the perceptual quality

Table 2. Multiple streams of the MPEG-4 content

Object priority level			High					low					
Object priority			11	10	9	8	7	6	5	4	3	2	1
Object ID			1	2	3	4	5	6	7	8	9	10	11
Object			MPEG-2 video	G.723	AAC	H.263	H.263	H.263	H.263	H.261	JPEG	JPEG	JPEG
ES layers	1	Bit rate	150	6.3	64	112.1	256	256	256	128	1	1	1
		RPQ	1	1	1	1	1	1	1	1	1	1	1
	2	Bit rate	120	5.3	32	56.7	192	192	192	64	0.9	0.9	0.9
		RPQ	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	3	Bit rate	100		16	22.81	128	128	128		0.7	0.7	0.7
		RPQ	0.8		0.8	0.8	0.8	0.8	0.8		0.8	0.8	0.8
	4	Bit rate			8		64	64	64		0.5	0.5	0.5
		RPQ			0.7		0.7	0.7	0.7		0.7	0.7	0.7
	5	Bit rate									0.25	0.25	0.25
		RPQ									0.6	0.6	0.6

(RPQ: Relative perceptual quality)

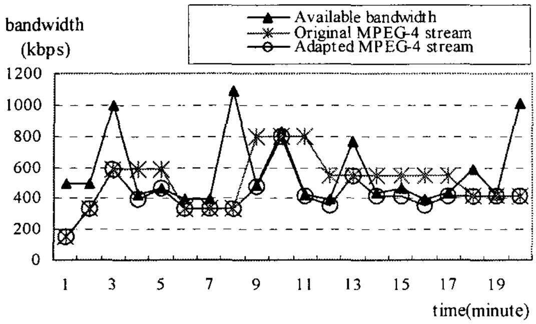

Fig. 4. Bitrate of the Adapted MPEG-4 stream for the MPEG-4 content.

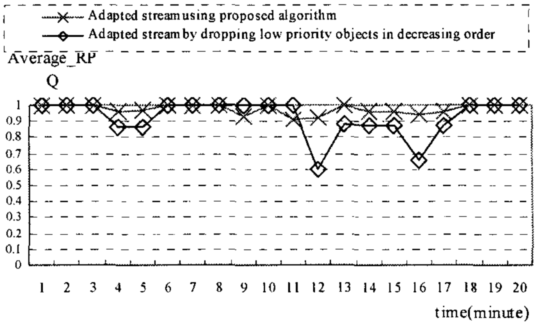

Fig. 5. Average relative perceptual quality: Adapted stream of the MPEG-4 content.

of transmitted streams as *Average_RPQ*. The *Average_RPQ* at time t is calculated as

$$Average\_RPQ = \sum_{i=1}^{n} O_i.P \times RPQ_i / \sum_{i=1}^{n} O_i.P$$

where RPQ_i indicates the relative perceptual quality of the selected ES layer of the object O_i and n is the number of the objects that are to be transmitted at time t.

Figure 5 shows the average RPQ for the adapted MPEG-4 binary resource for the proposed method and the method of dropping off the objects with relatively low object priority values under a bandwidth constraint. Since the proposed method takes into account the object priority in the bitstream adaptation process, the objects are less likely to be dropped off for the low priority objects. 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. When the required bandwidth unduly exceeds the available network bandwidth, it can be observed that the resulting perceptual quality of the adapted bitstream sharply decreases if an ES of that object with the lowest object priority value is dropped off, that is, not included in the adapted bitstream at time instance 12 and 16 in Figure 5.

Figure 6 shows the perceptual qualities for the high priority objects selected into the adapted bitstreams for the MPEG-4 content. In Figure 6, 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 6 that the quality of ES with high object priority is consistently maintained under a network with varying bandwidth. 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.

In order for multiple stream adaptation algorithm to be viable it need to be implementable with the following two major implementing issues : The selected stream layer of objects is maintained through the object's transmission period. The cur-

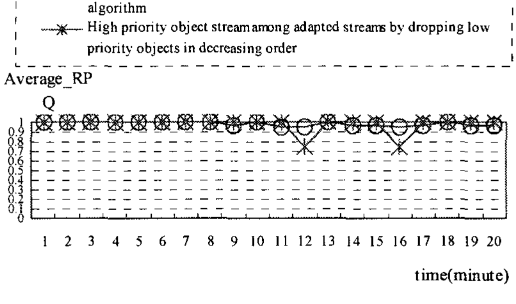

Fig. 6. Average perceptual quality: high priority objects among adapted streams of the MPEG-4 content.

rent bandwidth is estimated using previous evaluated values.

We also investigate the MPEG-4 content (refer to the Table 1) with the above assumption. Table 3 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.

Table 3. Average perceptual quality of the adapted MPEG-4 content

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	0.982
RPQ	1	1	1	1	1	0.8	1	1	1	1	1	

Figure 7 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 8 shows the estimated transmission delay when the adapted streams are delivered.

Figure 9 depicts the timeline of adapted MPEG-4 streams which shows in the above example. The timeline which is drawn with dotted line means that the original time duration of the object. As a result, we can see that high priority object streams are prepatched so as to maintain its quality and use available bandwidth fully. With this result, the higher priority object may be presented maintaining constant quality at the end user's terminal. Moreover, the lower priority object such

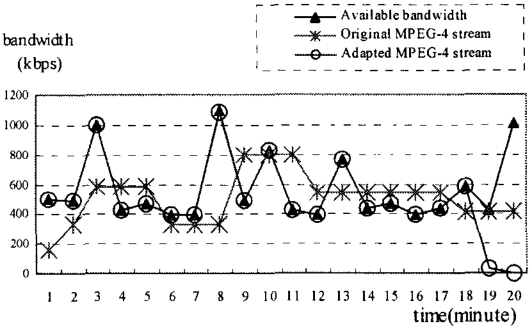

Fig. 7. Available bandwidth, original MPEG-4 stream of MPEG-4 content, adapted MPEG-4 stream of MPEG-4 content considering implementation issue.

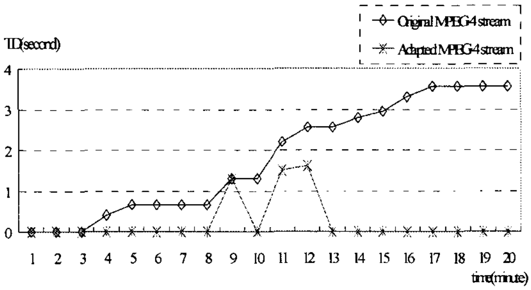

Fig. 8. Transmission delay : adapted MPEG-4 stream vs. original MPEG-4 stream of MPEG-4 content considering implementation issue.

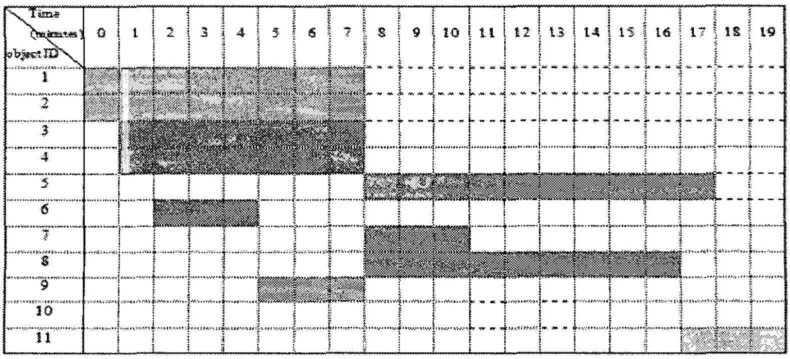

Fig. 9. Timeline of the object in the Adapted MPEG-4 content.

as object ID 10 can be delivered during its presentation time if the bandwidth become enough to cover its bitrate with other streams that already started to be transmitted.

5. CONCLUSIONS

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 in order to support an object-based transcoding 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.

6. REFERENCES

[1] P. Batra, "Modeling and efficient optimization for object-based scalability and some related problems," *IEEE Trans. on Image Processing*, Vol. 9, No. 10, pp. 1677-1692, October 2000.

[2] A. Vetro, H. Sun, and Y. Wang, "Object-Based Transcoding for Adaptable Video Content Delivery," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 3, pp. 387-401, March 2001.

[3] Goor, S.A. and Murphy, L., "A Priority Based Adaptive MPEG-4 Streaming System Using Pre-Encoded Video Objects," *Irish Telecommunications Systems Research Symposium* 2003(IEI/IEEE ITSRS).

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

[5] D. Turner and K.W. Ross "Optimal streaming of a synchronized multimedia presentation with layered Objects," *IEEE International Conference on Multimedia and Expo*, pp. 263-266, July 2000.

[6] Ronda, J.I., Eckert, M., Jaureguizar, F., and Garcia, N. "Rate control and bit allocation for MPEG-4," Vol. 9, Issue 8, pp. 1243-1258, December 1999.

[7] L.P. Kondi and A.K. Kataggelos, "An operational rate-distortion optimal ngle-pass SNR scalable video coder," *IEEE Trans. image processing*, Vol. 10, No. 11, pp. 1613-1620, November 2001.

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

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

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

[11] ISO/IEC JTC1/SC29/WG11, "ISO/IEC 21000-1 PDTR Second Edition: Vision, Technology and Strategy," *MPEG2003/* N6269, Hawaii, USA, Dec. 2003.

[12] J. Signes, Y. Fisher, and A. Eleftheriadis, "MPEG-4's binary format for description," *Signal Processing Image Communication*, 15 (4-5), pp. 321-345, January 2000.

[13] J. Kneip, B. Schmale, and H. Miller, "Applying and implementing the MPEG-4 multimedia standard," *IEEE Micro*, Vol. 19, No. 6, pp. 64-74, 1999.

[14] R. Koenen, "Profiles and levels in MPEG-4: approach and Signal Processing," *Image Communication*, Vol. 15, No. 4-5, pp. 463-478,

January 2000.

[15] B. Prabhakaran, "Multimedia authoring and presentation techniques," *Systems*, Vol. 8, issue 3, pp. 157, 2000.

Kyung-Ae Cha

Kyung-Ae Cha received the M.S and Ph.D degrees in Computer Science from Kyungpook National University, Korea in 1999 and 2003, respectively. She is currently a professor at the School of Computer Communication Engineering of Daegu University, Korea. Her research interests include Multimedia Systems, Multimedia contents authoring and presentation, Digital Multimedia Broadcasting.