A New Video Bit Rate Estimation Scheme using a Model for IPTV Services

Hye Jeong Cho[1], Dae-Young Noh[1], Seong Hwan Jang[2], Jae Cheol Kwon[2] and Seoung-Jun Oh[1]
[1] School of Electronic Engineering, Kwangwoon University, Seoul, Korea
[e-mail: {hjcho, kurtnoh, sjoh}@media.kw.ac.kr]
[2] Central R&D Laboratory, Korea Telecom (KT), Seoul, Korea
[e-mail: {shjang1234, jckwon}@kt.com]
*Corresponding author: Seoung-Jun Oh

Received March 31, 2011; revised July 25, 2011; accepted August 15, 2011; published October 31, 2011

Abstract

In this paper, we present a model-based video bit rate estimation scheme for reducing the bit rate while maintaining a given target quality in many video streaming services limited by network bandwidth, such as IPTV services. Each item of video content can be stored on a video streaming server and delivered with the estimated bit rate using the proposed scheme, which consists of the following two steps: 1) In the first step, the complexity of each intra-frame in a given item of video content is computed as a frame feature to extract a group of candidate frames with a lot of bits. 2) In the second step, the bit rate of the video content is determined by applying statistical analysis and hypothesis testing to that group. The experimental results show that our scheme can reduce the bit rate by up to 78% with negligible degradation of subjective quality, especially with the low-complexity videos commonly used in IPTV services.

Keywords: H.264, video streaming service, IPTV, rate control, frame complexity

The part of this paper was presented in the ICONI (International Conference on Internet) 2010, December 16-20, 2010, Philippines. This work was supported by the Central R&D Center Program of KT, the MKE/KIAT through the Workforce Development Program in Strategic Technology, and the Research Grant of Kwangwoon University in 2011.

DOI: 10.3837/tiis.2011.10.008

1. Introduction

Internet Protocol Television (IPTV) uses an IP broadband network to deliver television (cable TV type) services to end users. Traditional telecommunications service providers can utilize their IP networks (and broadband consumer access) to deliver broadcast TV, video on demand (VOD), and other Internet services to consumers [1].

The compression technology of serviced video is applied according to two business models: Managed Network and Open Internet. Service providers, such as Korea Telecom (KT), deliver H.264/AVC video contents through the Managed Network. In the Open Internet, the service is delivered over the public Internet and should enable access to IPTV content not only from TV sets but also from other home devices, such as portable multimedia players, laptop computers, and so on. The scalable video coding (SVC) technology enables the system to consider the available bandwidth for other devices. Additionally, a further study is needed on how to best control the SVC rate according to network resource availability [2]. The most developments launched IPTV are focused on delivering high resolution/quality video over the Managed Network with supporting quality of service (QoS).

The video technology adopted by many service providers is the H.264/AVC standard. The newest international video coding standard, H.264/AVC, has recently been approved by ITU-T as the recommended H.264 and by ISO/IEC as the MPEG-4 part10 advanced video coding (AVC) international standard [3]. The emerging H.264/AVC significantly achieves better performance in both peak signal to noise ratio (PSNR) and visual quality at the same bit rate compared with prior video coding standards. The H.264/AVC can save up to 39%, 49%, and 64% of the bit rate, when compared with MPEG-4, H.263, and MPEG-2 [4]. The standard is used in many video streaming services limited by network bandwidth, such as IPTV services. Video streaming services save video content on a streaming server after performing MPEG-2 to H.264 transcoding. The compressed video content is usually delivered through constant bit rate (CBR) channels. The bit rate channels needed for SDTV and HDTV video can be as high as 2–3Mbps and 10–12Mbps, respectively. **Fig. 1** shows the conventional scheme in an IPTV system over the Managed Network. TV programs are encoded through a transcoder and then serviced to users with many interesting contents via a streaming server. A web server manages user information such as authentication and accounting. Finally, the delivered contents are passed on to the set-top box (STB) in the home, in which they are decoded and then displayed on multimedia appliances.

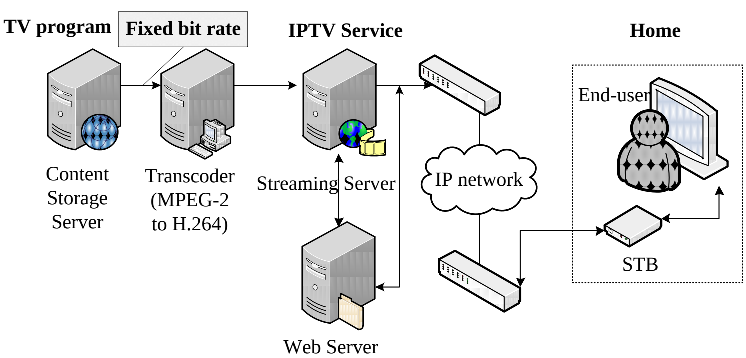


Fig. 1. The conventional scheme in IPTV system.

Each item of video content on a CBR channel does not take into account the content's characteristics because it is encoded by two different fixed bit rates; however, the serviced video content varies from low-complexity video to high-complexity video. The former can be encoded with a bit rate less than the fixed bit rate without degradation in subjective quality. In other words, the conventional scheme based on fixed bit rate causes bandwidth loss and requires a huge amount of storage space on a streaming server. When the open IPTV service is activated later, IPTV service providers can deliver the content, which, unlike specific companies' customized content, is a network resource that anyone can access. In order to deliver a lot of content on a CBR channel, it is important to select an efficient bit rate.

Solving this problem requires a scheme capable of finding an appropriate bit rate for video content while maintaining subjective quality equivalent to that of a scheme that uses a fixed bit rate. Employing this scheme requires determining a bit rate for video content prior to encoding that video content. A video transcoder can provide an additional controller that can also estimate the bit rate. A simple technique to estimate video content's bit rate is to vary the bit rate step in the H.264 encoder part of the transcoder. Visual quality should be verified at each encoding pass. Even though that method can provide an accurate bit rate, it is a very time-consuming process. The time required to estimate the bit rate should be minimized to meet the video streaming service requirements.

In this paper, we propose a new scheme for estimating the bit rate without multiple full encoding using a model based on a given target quality. Some contents in the subjective quality analysis can represent that the quality degradation is not recognized, even though they are encoded at a lower bit rate than the fixed bit rate. Ultimately, the contents encoded at a variable bit rate (VBR) coding mode using constant quantization parameters (QP) in the source coding are able to deliver a lower bit rate if the average bit rate is lower than the fixed bit rate. To minimize the time required for the encoding process without multiple full encoding, a candidate frame set is extracted by statistical analysis, which investigates the distribution of parameters highly correlated with the number of bits. Using the candidate frame set, the bit rate is estimated by mathematical models based on a predefined target quality. In addition, a fast estimation scheme is desgined by reducing the number of candidate elements. The processing time can be reduced through analyzing the temporal characteristic between GOPs including the candidate frames. This paper is organized as follows. Section 2 explains the analysis of subjective quality. Section 3 proposes the scheme for bit rate estimation based on the statistical analysis and the mathematical model. Section 4 explains the key-GOP extraction method to improve the processing speed up. Section 5 presents the experimental results. Then, we conclude the paper in Section 6.

2. Analysis of Subjective Quality in Low-Complexity Video

The purpose of this analysis is to examine the human perceived quality corresponding bit rate of video. We evaluate the subjective quality of H.264/AVC encoded video, in which low-complexity content category such as lecture is coded at bit rates from 1.0Mbps to 2.5Mbps. The evaluation is performed using the Double-Stimulus Continuous Quality Scale (DSCQS) method of ITU-R Rec. BT.500-7 [5]. All of the coded stimuli are rated by each viewer. Five viewers participated and general conclusions were based on their quality rating of the presented stimuli. The main idea of measuring DSCQS is to determine the Differential Mean Opinion Score (DMOS) between the reference encoded at 2.5Mbps and the test sequences averaged by all viewers. The task is to assess the degradation of the test sequence with respect to the reference sequence. If the DMOS value is near “0”, then the test sequence is similar to the reference sequence.

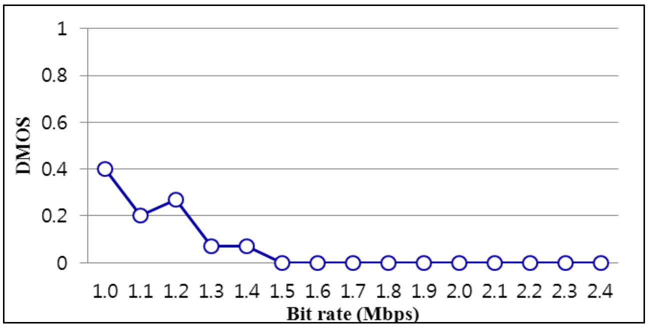


Fig. 2. Result of quality evaluation.

Fig. 2 shows the result of average for all DMOS in a low-complexity video. The quality degradation determined by the video encoded bit rate was on average 1.4Mbps. Therefore, the low-complexity video can encode a bit rate lower than 2.5Mbps with negligible degradation of subjective quality.

3. The Proposed Scheme for Bit Rate Estimation

In this section, we propose a bit rate estimation scheme to reduce the bit rate while maintaining the target quality in video streaming services limited by network bandwidth. **Fig. 3** shows a block diagram of the proposed scheme. In the transcoder, the input video is decoded by MPEG-2 and re-encoded by H.264 with the estimated bit rate. The proposed scheme is different from the conventional scheme, for there exists the bit rate estimator in it. To estimate the bit rate, a candidate frame set is extracted, which includes intra-frames that require a high number of bits. Because the proposed scheme does not encode full frames of video content, it is very important to determine the parameter that serves as an indirect measurement of frame’s bits. Using a candidate frame set, the bit rate is estimated by statistical analysis and a mathematical model based on target quality.

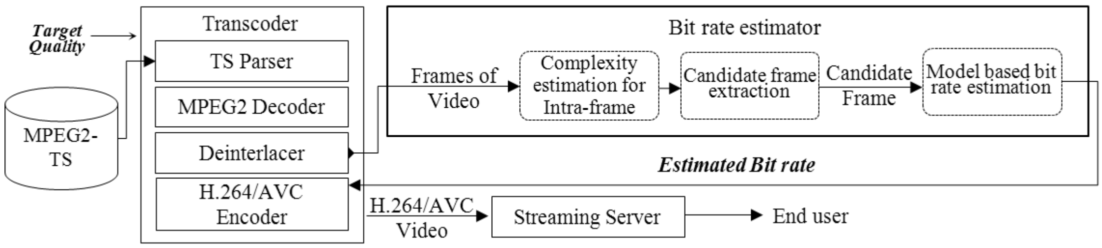


Fig. 3. The block diagram of the proposed scheme.

3.1 Coding Complexity Estimation for an Intra-frame

Some content complexity measurements for coding still images can be obtained without pre-encoding using variance-, edge-, and gradient-based methods [6]. Among them, the gradient-based method has been recently used to measure the complexity of an intra-frame [7]. In addition to the gradient information, the histograms of luminance and chrominance pixel values are also very useful when combined with the gradient to represent the content complexity.

Given the arbitrary *s*th test seuqence $\mathbf{Q_s}$, the set contains a number of groups of pictures (GOPs) specified the order in which intra- and inter-frames are arranged as follows:

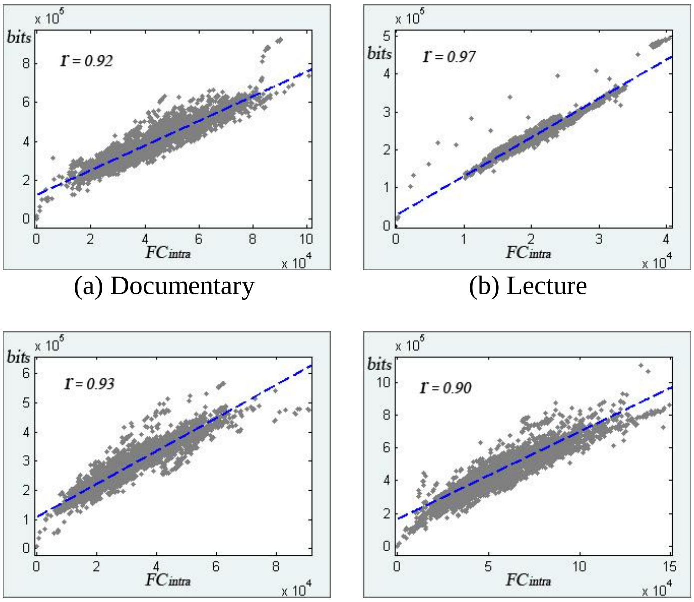
$$\mathbf{Q_s} = \{\{Q_{1,1}^s, \ldots, Q_{1,N}^s\}, \ldots, \{Q_{M,1}^s, \ldots, Q_{M,N}^s\}\} \tag{1}$$

where M is the total number of GOPs, and N is the number of frames in a GOP. $Q_{i,j}^s$ means the *j*th frame of the *i*th GOP. Thus, $Q_{i,1}^s$ is the intra-frame in the *i*th GOP. Our objective is to measure the complexity for the intra-frame in $\mathbf{Q_s}$. In order to measure a frame complexity, we use the complexity measurement defined in [8]. We call it FC_{intra}. The value of FC_{intra} for $Q_{i,1}^s \in \mathbf{Q_S}$, $CC(Q_{i,1}^s)$, can be computed by Eq. (2).

$$\begin{aligned}
&CC(Q_{i,1}^s) = Grad_i^s \times SOH_i^s \\
&\text{where} \\
&Grad_i^s = \sum_{x=0}^{K_Y-2}\sum_{y=0}^{L_Y-2}(|Y_{x,y}^{s,i} - Y_{x,y+1}^{s,i}| + |Y_{x,y}^{s,i} - Y_{x+1,y}^{s,i}|)/(K_Y L_Y) \\
&\qquad + \sum_{x=0}^{K_U-2}\sum_{y=0}^{L_U-2}(|U_{x,y}^{s,i} - U_{x,y+1}^{s,i}| + |U_{x,y}^{s,i} - U_{x+1,y}^{s,i}|)/(K_U L_U) \\
&\qquad + \sum_{x=0}^{K_V-2}\sum_{y=0}^{L_V-2}(|V_{x,y}^{s,i} - V_{x,y+1}^{s,i}| + |V_{x,y}^{s,i} - V_{x+1,y}^{s,i}|)/(K_V L_V), \\
&SOH_i^s = \sum_{l=0}^{255}(\log_2 Hist_Y^{s,i}[l] + \log_2 Hist_U^{s,i}[l] + \log_2 Hist_V^{s,i}[l])
\end{aligned} \tag{2}$$

In Eq. (2), $Grad_i^s$ and SOH_i^s are the gradient and the statistic of histogram information of the *i*th intra-frame, respectively. $Y_{x,y}^{s,i}$ is a luminance value of a pixel (*x*,*y*) in the *i*th frame. $U_{x,y}^{s,i}$ and $V_{x,y}^{s,i}$ are the chrominance values correspondingly. $K_Y \times L_Y$, $K_U \times L_U$, $K_V \times L_V$ are the sizes of Y-, U-, V-frames in $Q_{i,1}^s$. $Hist\_Y^{s,i}[l]$ is the histogram of the luminance level *l*, and $Hist\_U^{s,i}[l]$ and $Hist\_V^{s,i}[l]$ are the histograms of chrominance level *l* correspondingly.

To investigate the relationship between the actual number of encoded bits and FC_{intra}, we have extensively encoded various test sequences using intra coding mode under constant quantization parameters (QPs), and recorded the number of encoded bits and FC_{intra} for each frame. **Fig. 4** shows the scatter plots of the number of bits versus FC_{intra} at different QPs in our test contents where each dot represents a frame. **Fig. 4** also shows the accuracy of the linear approximation in blue dotted lines by the correlation coefficient *r*, which is an indicator that reveals how closely the approximated linear relationship is to the actual data. The value of *r* is between -1 and 1. For the test sequences, the value of *r* between the number of bits and FC_{intra} is on average 0.93. When the value of *r* is at or near 1, the approximated linear relationship is the most reliable. The linear relationship also exists in other sequences with different slopes. Therefore, Eq. (2) can be used accurately to estimate the number of bits for intra-frames.



(a) Documentary (b) Lecture

(c) Religion (d) Sports

Fig. 4. Scatter plots of the number of encoded bits versus FC_{intra}.

3.2 A Hypothesis Test for Candidate Frame Extraction

Although the correlation between FC_{intra} and the number of bits is high, the maximum FC_{intra} frame does not always have the maximum number of encoded bits. Thus, we need to extract the candidate intra-frame in $\mathbf{Q_s}$. The candidate frame set $\mathbf{T_s}$ contains D intra-frames, and each element in $\mathbf{T_s}$ is a frame which requires more than a certain number of encoded bits, which is specified in Eq. (3):

$$\mathbf{T_s} = \left\{ \begin{array}{l} I^s_{c,1} \mid \theta(c) = i \text{ and } c = c+1, \ \text{ if } CC(Q^s_{i,1}) \ge \tau_s \\ \qquad\qquad \text{where } \tau_s = \mu_s + k\sigma_s,\ 1 \le c \le D \le M,\ \text{and } 1 \le i \le M \end{array} \right\} \tag{3}$$

where $I^s_{c,1}$ is a candidate intra-frame, and $\theta(\cdot)$ is a nondecreasing mapping function from the integer set $\{1,\ldots,M\}$. If $CC(Q^s_{i,1})$ is greater than the content-adaptive threshold τ_s, the ith intra-frame $Q^s_{i,1}$ is extracted as $I^s_{c,1}$. τ_s is the threshold which is decided according to the test sequence. μ_s and σ_s are the mean and standard deviation of FC_{intra} values in the sth sequence, respectively. k is a constant which can be chosen using statistical analysis. To determine k, we firstly need to decide an FC_{intra} distribution. Using the goodness-of-fit between FC_{intra} and theoretical values of each candidate probability density function (PDF), we assume that FC_{intra} approximately has a discrete Gaussian distribution as following:

$$f_X(x_i) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x_i-\mu_X)^2/2\sigma_X^2} \tag{4}$$

where a random variable X is distributed with a mean μ_X and variance σ_X^2 under the test sequences. We use a statistical method based on hypothesis testing to decide k of τ_s. The null hypothesis H_0 is that the frame with the maximum number of encoded bits is not an element of $\mathbf{T_s}$. The alternative hypothesis H_1 is that its frame is an element of $\mathbf{T_s}$. Let h be the number for the intra-frame with the maximum number of encoded bits, the hypotheses denoted by Eq. (5).

$$\begin{aligned} H_0 &: I_{h,1}^{s} \notin \mathbf{T_s}, \\ H_1 &: I_{h,1}^{s} \in \mathbf{T_s} \end{aligned} \tag{5}$$

Given μ_s and σ_s in Eq. (3), we can calculate k_s satisfing the equation $CC(Q_{h,1}^{s}) = \mu_s + k_s\sigma_s$ in $\mathbf{Q_s}$. Among k_s's, we choose the maximum value of k that rejects the H_0 in every test sequences. We set to k=1.2 empirically because of $0 \le k_s \le 1.2$.

3.3 Model-based Bit Rate Estimation

Using the set of $\mathbf{T_s}$, the bit rate can be estimated via statistical analysis and a mathematical model. To estimate the bit rate while maintaining the given PSNR quality, a PSNR-Q model is proposed in this paper, which is derived from the H.264/AVC Quantization process [9]. With this model, an estimated QP is determined and is finally applied to the bit rate estimation. The relationship between the quantization step size (*Qstep*) and QP is

$$Qstep = \frac{2^{qbits} \times PF}{MF} \tag{6}$$

where *PF* and *MF* are a post-scaling and a multiplication factor in the H.264/AVC standard and *qbits* = 15+floor (QP/6). When uniform quantization is applied to uniformly distributed inputs, the mean square error (*MSE*) is as follows.

$$MSE = \frac{1}{Qstep}\int_{-Qstep/2}^{Qstep/2} u^2 du = \frac{Qstep^2}{12} \tag{7}$$

According to Eqs. (6) and (7), the PSNR can be derived as:

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) = a \times QP + b \tag{8}$$

where a and b are constant values obtained by linear regression. As a result, the value of QP can be estimated as:

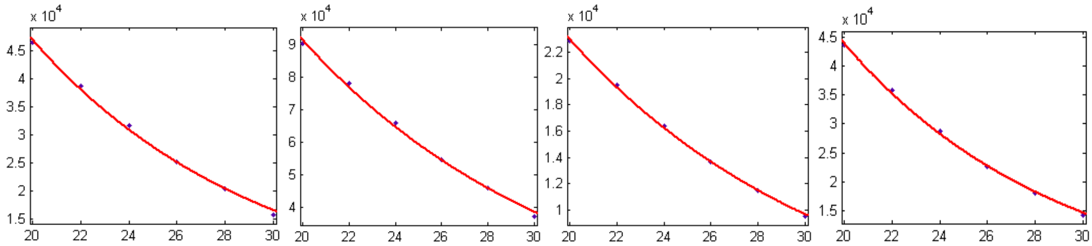
$$QP_e = \frac{b - PSNR_t}{a} \tag{9}$$

where $PSNR_t$ is a given target PSNR, and QP_e is an estimated QP.

Using QP_e, the number of intra-frame bits is firstly estimated. Some parameters obtained via intra-frame estimation are used to estimate the number of inter-frames bits in a GOP. Based on those results, the bit rate for an item of video content can be estimated. To estimate the number of intra-frame bits, we use a simple but effective Rate-Quantization (R-Q) model. An exponential relationship between the actual number of encoded bits and QP was modeled by Zhou and his colleagues [8]. For simplicity, the following R-Q model for intra-frames is defined as:

$$R_{c,1}(QP_e) = \alpha \times e^{(-\beta \times QP_e)} \tag{10}$$

where $R_{c,1}(QP_e)$ is the number of encoded bits for the cth candidate intra-frame at QP_e, and α and β are the model's parameters [8]. To reveal the relationship between the number of encoded bits and QP, **Fig. 5** shows several examples of curve-fitting results for intra-frames, which are the mathematically approximated curves where each small dot represents the actual number of encoded bits of an intra-frame at each QP. Because α and β can be obtained by exponential regression, $R_{c,1}$ can also be calculated by Eq. (10).



(a) Music (b) Lecture (c) Sports (d) Documentary

Fig. 5. R-Q curves for test video sequences.

It is difficult directly to estimate the number of inter-frame bits in H.264/AVC. Thus, the bit rate conversion method introduced in [10] is used with the value of QP_e instead of using the intra-frame R-Q model. The bit rate conversion is defined as:

$$R_{c,j+1}(QPP) = R_{c,j+1}(QP_s) \times 2^{(QPP - QP_s)/6}, \quad 1 \le j \le N-1 \tag{11}$$

where $R_{c,j+1}(QPP)$ is the number of encoded bits for the (j+1)th inter-frame in the cth GOP at QPP, N is a GOP size. As defined in Eq. (11), this method requires encoding a GOP at a certain value of QP, QP_s, as a reference, that is, $R_{c,j+1}(QP_s)$ is computed in advance. In our experiment, we use QP_s=26. Furthermore, QPP is set to QP_e+1 here since an inter-frame QP is an intra-frame QP+1 in H.264/AVC rate control. After estimating the number of intra- and inter-frame bits, the total number of bits for each GOP, R_c can be estimated using Eqs. (10) and (11) as following:

$$R_c = R_{c,1} + \sum_{j=1}^{N-1} R_{c,j+1} \cdot \tag{12}$$

Finally, the bit rate for a given video content is estimated using the GOP that is expected to have the maximum number of encoded bits among all GOPs. The number of bits per second (*BPS*) can be calculated as follows:

$$BPS = FPS \times \frac{\arg\max_c \{R_c \mid 1 \le c \le D\}}{N}, \tag{13}$$

where *FPS* is the number of frames per second, *N* is a GOP size.

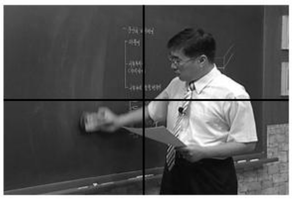
4. The key-GOP Extraction Method to Improve the Processing Speed

The bit rate estimation for a sequence is a time-consuming process since it depends on the number of elements in $\mathbf{T}_s$. Thus, it is to worthwhile to design a time-reduction process based on the temporal characteristic between GOPs. To do this work, we define a candidate GOP set $\hat{\mathbf{T}}_s$ which includes an intra- as well as inter-frames as following:

$$\hat{\mathbf{T}}_s = \{I_{c,j}^s \mid 1 \le j \le N,\ 1 \le c \le D\} \tag{14}$$

where $I_{c,j}^s$ is a frame in the *c*th GOP, and *N* is the number of frames in each GOP.

To analyze the characteristic between GOPs in $\hat{\mathbf{T}}_s$, we have to mesure the similarity between frames. We choose the Spearman's Rank Correlation Coefficients (*RCC*) defined in [11] as the similarity between two frames. In [11], *RCC* is used as a measurement of correlation in nonparametric statistics when the data are in ordinal form. The ordinal signature can be computed very simply [12]. The ordinal signature is based on a luminance feature. An example for the ordinal signature is shown in **Fig. 6**. An image is partitioned into $B_x \times B_y$ equal-sized blocks, which makes the system independent of input image sizes, and the $B_x \times B_y$ subimage is calculated by taking the average value of each block. This array is converted to the ordinal signature.



(a)

90	118
64	182

(b)

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

(c)

(a) Image is partitioned into $B_x \times B_y$ equal-sized blocks (2×2 in this example).
(b) Average values of blocks. (c) Ordinal signature of (b).
Fig. 6. An example of ordinal signature.

Let $G[I_{c,j}^s]$ be an ordinal signature for $I_{c,j}^s \in \hat{\mathbf{T}}_s$. The value of *RCC* between the *j*th and (*j*+1)th frames is calculated as follows:

$$RCC(G[I^s_{c,j}], G[I^s_{c,j+1}]) = 1 - \frac{6 \times Diff(G[I^s_{c,j}], G[I^s_{c,j+1}])}{B \times (B^2 - 1)} \quad (15)$$

where $Diff(\cdot)$ is the sum of squared differences of each position between the ordinal signatures of $G[I^s_{c,j}]$ and $G[I^s_{c,j+1}]$, and B is the number of blocks in the frame, which is set to 4 ($B_x = B_y$ =2) in this paper. It is obvious that $-1 \le RCC(\cdot) \le 1$. In case of $RCC(\cdot) \cong 1$, we can say that two frames are highly correlated. **Fig. 7** shows the value of *RCC* between each two frames in the test sequence (Lecture). The x-axis indicates each $I^s_{c,j} \in \hat{\mathbf{T}}_\mathbf{s}$. Even though most of *RCC* values are very close to 1, there exist some valleys in the graph.

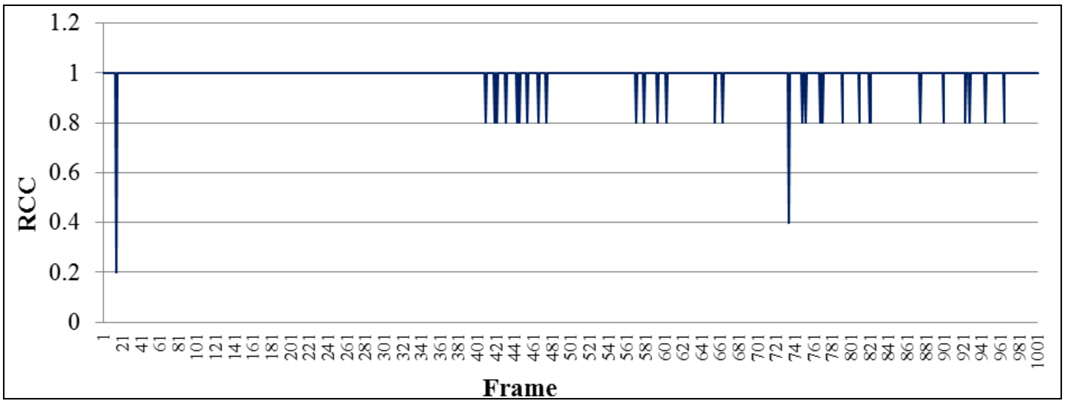


Fig. 7. The value of RCC between each two frames (Lecture).

At this moment, homogeneity of *RCC* values in a GOP is tested by Eq. (16) which is an average of *RCC* values in the *c*th GOP:

$$\omega_c = \frac{1}{N-1} \sum_{j=1}^{N-1} RCC(G[I^s_{c,j}], G[I^s_{c,j+1}]) \quad (16)$$

where N is the GOP size. We can say that neighboring frames in that GOP are highly correlated in case of $\omega_c = 1$. We plot the values of R_c defined in Eq. (12) and ω_c in **Fig. 8**. It shows that two values are inversely correlated. In the region of $\omega_c = 1$, the number of bits for each GOP is very similar. However, we can not say that two GOPs are similar even though they have the same value of ω. There may exist that they have different ordinal signature values. If two GOPs have different values with the same ω, they may have different number of bits for encoding them. In that case, two GOPs must be treated as candidate GOPs in order to estimate the bitrate for the sequence. In case of $\omega_c \neq 1$, those GOPs are also treated as candidate GOPs since they have totally different R_c values. Let's call those candidate GOPs as key-GOPs. To formulize this process, we define λ_c as follows.

$$\lambda_c = \begin{cases} 1, & \text{if } \omega_c = \omega_{c+1} = 1 \text{ and } G[I^s_{c,1}] = G[I^s_{c+1,1}] \\ 0, & \text{else} \end{cases} \quad (17)$$

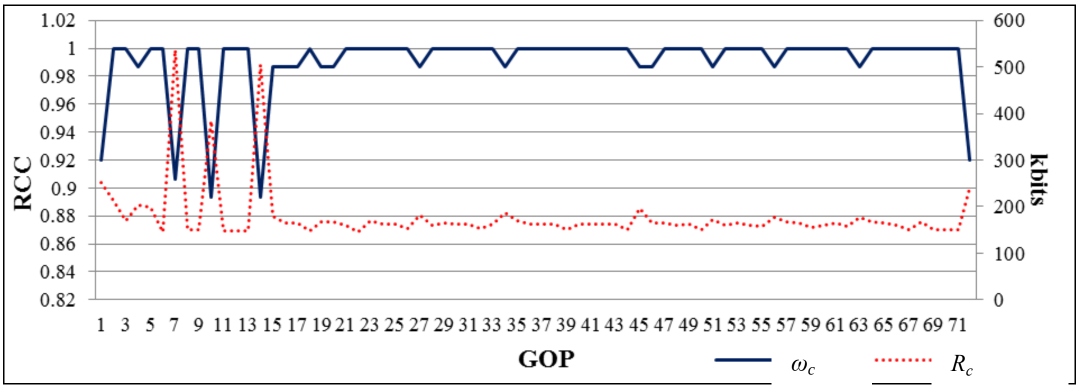


Fig. 8. The relationship between the actual number of encoded bits and average of *RCC*.

If all λ_c's are equal to 1 in the region of ω_c = 1 whose interval is *a*, that is, [*c*, *c*+*a*], a key-GOP for that region is the c'th GOP which has the maximum value of FC_{intra} in the interval of [*c*, *c*+*a*]. Thus, we can claim that the computational complexity can be significantly reduced since the size of a key-GOP set (denoted as $\mathbf{H}_s$) is generally much smaller than that of $\hat{\mathbf{T}}_s$.

5. Experimental Results

The performance of the proposed method was evaluated on several IPTV contents. According to a subjective test, those IPTV contents are classified into two classes: low-complexity and high-complexity videos. These are Standard Definition (SD) resolution video contents, which are categorized into 4 genres: lecture, religion and documentary, drama and animation, and music video and sports. A total of 30 video contents in **Table 1** are used as test sequences.

Table 1. Test sequences

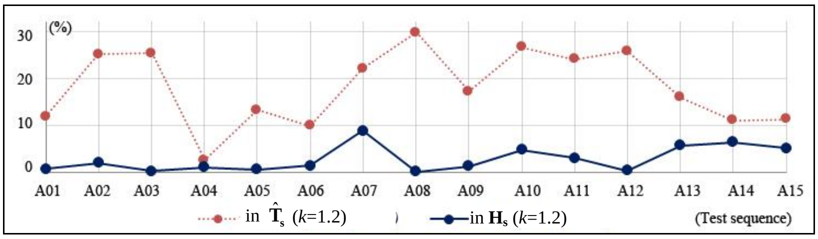
Class	Genre	Test Sequence
Low-complexity video	Lecture	A01 – A15
High-complexity video	Religion & Documentary	B01 – B06
	Drama & Animation	C01 – C04
	Music video & Sports	D01 –D05

All the experiments are conducted on Windows XP operating system, with Intel Core (TM) i7 950 CPU 3.07GHz and 2GB RAM. A transcoder is implemented using libmpeg2 open source as MPEG-2 decoder and x.264 open source as H.264 encoder [13][14]. The GOP size *N* is 15 and set as IPPP. The target PSNR is set to 40dB. The simulated results encoded by the conventional scheme with a fixed bit rate of 2.5Mbps can be compared in terms of bit rate and quality to those encoded by the proposed scheme with the estimated bit rate. When evaluating the bit rate reduction, we calculate ΔR defined as:

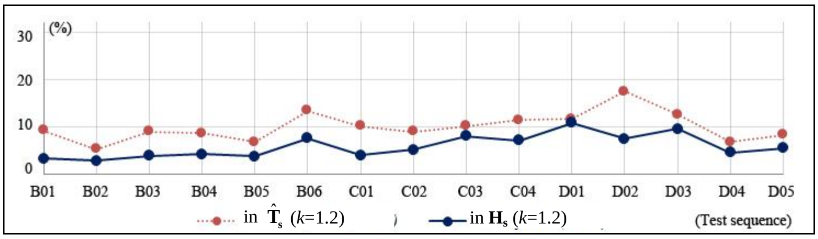
$$\Delta R = \frac{R_{\text{pro\_scheme}} - R_{\text{con\_scheme}}}{R_{\text{con\_scheme}}} \times 100 \qquad (18)$$

where R_{pro_scheme} and R_{con_scheme} indicate the estimated bit rate with the proposed scheme and the fixed bit rate with the conventional scheme, respectively.

The bit rate estimation for a sequence is calculated by encoding the frames. Instead of the total frames, the bit rate is estimated using the candidate frames. The computational complexity depends on the number of candidate frames. It is related to the time required to estimate the bit rate. **Fig. 9** shows the percentage for the number of candidate frames in $\hat{\mathbf{T}}_s$ and $\mathbf{H}_s$ when k=1.2. It means the percentatge for the number of encoded frames in the total frames. For $\mathbf{H}_s$, on average, the percentage of the number of candidate frames is 2.7% in low-complexity video and 5.8% in high-complexity video, respectively. It shows that $\mathbf{H}_s$ in low-complexity video compared to the number of candidate frames of $\hat{\mathbf{T}}_s$ is reduced by 15.4% on average, and high-complexity video is reduced by 4.2% on average. Thus, the proposed scheme based on the key-GOP extraction method can reduce the processing time because the number of candidate frames of the $\mathbf{H}_s$ is in proportion to the number of encoded frames in order to estimate the bit rate.



(a) Low-complexity video



(b) High-complexity video

Fig. 9. The percentage the numbers of candidate frames in $\hat{\mathbf{T}}_s$ and $\mathbf{H}_s$ when k=1.2.

Table 2 shows the results of bit rate reduction compared to the conventional scheme based on a fixed bit rate. The estimated bit rates based on the key-GOP extraction provide the same quality according to the class of test sequence, in which low-complexity video is estimated with a low bit rate and high-complexity video is estimated with a high bit rate of 2.5Mbps. IPTV services are not delivered over higher bit rate than the fixed bit rate because the network bandwidth is limited. Therefore, the bit rate reduction in the high-complexity video of **Table 2** is zero when the bit rate is estimated using the candidate frames in the high-complexity video of **Fig. 9**. The high-complexity video is finally delivered at the fixed bit rate. For this reason, there are no changes of the saving storage and quality in the high-complexity video. Thus, the results show the best performance with low-complexity video.

Table 2. Bit rate reduction ratio relative to fixed bit rate (2.5Mbps)

Class	Test sequence	R_{pro_scheme} (kbps)	ΔR (%)	Class	Test sequence	R_{pro_scheme} (kbps)	ΔR (%)
Low-complexity video	A01	1306	-49	High-complexity video	B01	2560	0
	A02	1045	-59		B02	2560	0
	A03	1102	-57		B03	2560	0
	A04	1082	-58		B04	2560	0
	A05	736	-71		B05	2560	0
	A06	968	-62		B06	2560	0
	A07	884	-65		C01	2560	0
	A08	566	-78		C02	2560	0
	A09	894	-65		C03	2560	0
	A10	795	-69		C04	2560	0
	A11	1096	-57		D01	2560	0
	A12	1419	-45		D02	2560	0
	A13	1317	-49		D03	2560	0
	A14	950	-63		D04	2560	0
	A15	1569	-39		D05	2560	0

Reducing the bit rate also reduces the need for storage space as shown in **Fig. 10**. On average, 56% of file size for fixed bit rate can be saved, which means the saving ratio of storage space is in proportion to the bit rate reduction ratio.

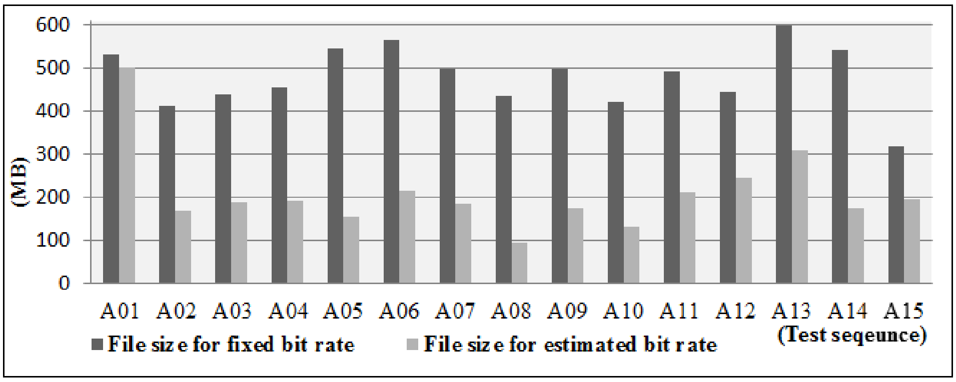


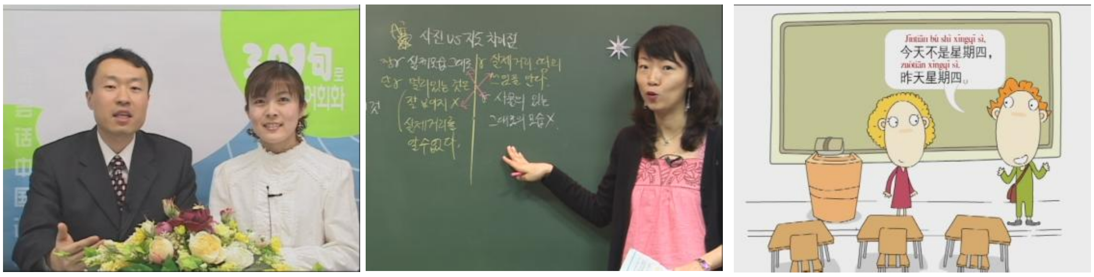
Fig. 10. Saving storage relative to fixed bit rate.

The video quality was evaluated using the encoded test sequence with the estimated bit rate, as is shown in **Table 2**. Compared with the fixed bit rate, **Table 3** shows the performance of PSNR. In low-complexity video, each PSNR value decreased by 2.31dB on average; however, all the PSNR results exceeded 40dB. The estimated bit rate causes negligible degradation in subjective quality compared to video content encoded using the fixed bit rate as shown in **Fig. 11**.

Table 3. Performance comparison for PSNR

Test sequence	PSNR (dB)		Test sequence	PSNR (dB)	
	R_{con_scheme}	R_{pro_scheme}		R_{con_scheme}	R_{pro_scheme}
A01	46.5	46.37	A09	46.57	43.65

A02	45.97	43.8	A10	46.87	44.06
A03	46.24	44.05	A11	46.37	44.21
A04	46.95	44.72	A12	46.2	44.93
A05	47.16	44.36	A13	43.99	42.13
A06	46.71	44.37	A14	46.1	43.69
A07	49.64	44.78	A15	45.65	44.42
A08	47.13	43.78	Average PSNR drop		-2.31



(a) Fixed bit rate

(b) Estimated bit rate

Fig. 11. Comparison of subjective quality on fixed bit rate and estimated bit rate in low-complexity videos.

6. Conclusions

This paper has presented a new bit rate estimation scheme using a model based on a given target quality of video streaming services such as IPTV services. This video bit rate estimation scheme applies the video content-adaptive bit rate to each item of video content, instead of applying a fixed bit rate to all content. Using the key-GOP extraction method, the proposed scheme can estimate the bit rate using several frames without multiple full encoding. On average, the frames are encoded by using about 2.7% of the test sequence’s full frames in low-complexity video. Therefore, we minimize the time required to estimate the bit rate. The proposed scheme can reduce the bit rate by up to 78%, compared to that of video content encoded using a fixed bit rate, in low-complexity video. Video content encoded using the estimated bit rate shows negligible subjective quality degradation compared to video content encoded using a fixed bit rate.

References

[1] S.O. Hwang, “Content and Service Protection for IPTV,” *IEEE Transaction on Broadcasting*, vol. 55, no. 2, pp. 425-436, Jun. 2009. Article (CrossRef Link)

[2] S. Park, S.H. Jeong, "Mobile IPTV: Approaches, Challenges, Standards and QoS Support," *IEEE Internet Computing*, vol. 13, no. 3, pp. 22-31, May–Jun. 2008. Article (CrossRef Link)
[3] T. Wiegand, G.J. Sullivan. G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transaction on Circuits and System for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003. Article (CrossRef Link)
[4] A. Joch, F. Kossentini, H. Schwarz, T. Wiegand, G.J. Sullivan, "Performance Comparison of Video Coding Standards using Lagrangian Coder Control," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. II-501-504, Sep. 2002. Article (CrossRef Link)
[5] ITU-R Recommendation BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU, 2002. Article (CrossRef Link)
[6] W.J. Kim, J.W. Yi, S.D. Kim, "A Bit Allocation Method based on Picture Activity for Still Image Coding," *IEEE Transaction on Image Processing*, vol. 8, no. 7, pp. 974-977, 1999. Article (CrossRef Link)
[7] X. Jing, L.-P. Chau, "A Novel Intra-rate Estimation Method for H.264 rate Control," in *Proc. IEEE Int. Symposium on Circuits and Systems*, May 2006. Article (CrossRef Link)
[8] Y. Zhou, Y. Sun, Z. Feng, S. Sun, "New Rate-distortion Modeling and Efficient Rate Control for H.264/AVC Video Coding," *Signal Processing: Image Communication*, vol. 24, no. 5, pp. 345-356, May 2009. Article (CrossRef Link)
[9] Y. Liu, Z.G. Li, Y.C. Soh, "A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard," *IEEE Transaction on Circuits and System for Video Technology*, vol. 17, no. 1, pp. 68-78, Jan. 2007. Article (CrossRef Link)
[10] Q. Tang, H. Mansour, P. Nasiopoulos, R. Ward, "Bit-rate Estimation for Bit-rate Reduction H.264/AVC Video Transcoding in Wireless Networks," in *Proc. of IEEE International Symposium on Wireless Pervasive Computing*, pp. 464-467, May 2008. Article (CrossRef Link)
[11] J.L. Myers, A.D. Well, "Research Design and Statistical Analysis," 2nd Edition, Lawrence Erlbaum.
[12] D.N. Bhat. S.K. Nayar, "Ordinal Measures for Image Correspondence," *IEEE Transaction on Pattern Analysis*, Apr. 1998. Article (CrossRef Link)
[13] http://ffmpeg.org/index.html
[14] http://www.videolan.org/developers/x264.html

Hye Jeong Cho received the B.S. degree in 2004 from the Department of Internet Information Engineering, Hanyang Women's College, Seoul, Korea. She is currently pursuing the joint M.S. and Ph.D. degree in electronic engineering, Kwangwoon University, Seoul, Korea. Her research interests include video coding, face recognition, and IPTV video streaming services.

Dae-Young Noh received both the B.S. and the M.S. degrees in electronic engineering from Kwangwoon University, Seoul, Korea, in 2001 and 2003, respectively. He is currently pursuing the Ph.D. degree in electronics engineering, Kwangwoon University, Seoul, Korea. His research interests include video compression, video coding, and image processing system.

Seong Hwan Jang was born Seoul, Korea, in 1964. He received both the B.S. and the M.S degrees in electronics engineering from Kyunghee University, Seoul, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from Georgia Tech, Atlanta, in 2005. In 1990, he joined Korea Telecom Research Center, Korea, as a member of research staff. His research interests include multimedia communication, video coding, IPTV services, and video streaming services.

Jae Cheol Kwon received the B.S. in Electronic Engineering from Hanyang University, M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology in 1988 and 2003 respectively, in Korea. Up to now from 1988, he has worked in KT as a principal research engineer of Central R&D Lab.. His research interests include networked video, video quality assessment and audiovisual systems.

Seoung-Jun Oh was born Seoul, Korea, in 1957. He received both the B.S. and the M.S. degrees in electronic engineering from Seoul National University, Seoul, in 1980 and 1982, respectively, and the Ph.D. degree in electrical and computer engineering from Syracuse University, New York, in 1988. In 1988, he joined ETRI, Daejeon, Korea, as a senior research member. Later he was promoted to a Program of Multimedia Research Session, ETRI. Since 1992, he has been a professor of Department of Electronic Engineering, Kwangwoon University, Seoul, Korea. He has been a chairman of SC29-Korea since 2001. His research interests include image and video processing, video compression, video coding, and motion estimation.