

RESEARCH

Open Access

Bitrate control using a heuristic spatial resolution adjustment for a real-time H.264/AVC encoder

Chae Eun Rhee¹, Jin-Sung Kim^{2*} and Hyuk-Jae Lee¹**Abstract**

Conventional bitrate control algorithms that change only the quantization parameter (QP) often suffer from quality degradation when the target bitrate is very low. Therefore, rate control algorithms that adjust spatial resolution in addition to QP control have recently been proposed, but their computations are too complex to be processed in real time. This research proposes a very simple, but effective, rate control algorithm that employs spatial resolution control as well as the existing QP-based bitrate control. The spatial resolution ratio for the best peak signal-to-noise ratio (PSNR) is calculated using a simple estimation model which defines the relationship between the PSNR and the spatial resolution at very low bitrate compression. In the proposed bitrate control algorithm, two scalability tools for adjusting the QP and the spatial resolution ratio are used sequentially to reach the target PSNR and the control decision is made for a group of pictures. Experimental results show that the proposed bitrate control algorithm approximates an optimal solution and yields a better subjective quality as well as objective quality at various bitrates compared to the conventional QP-based bitrate control algorithm. The decision of the control parameters requires very small computational complexity and is made in a completely automatic manner so that the proposed algorithm is well suited for real-time applications.

Keywords: H.264/AVC, bitrate control, down-sampling, spatial resolution control

1. Introduction

H.264/AVC standard is widely used in video streaming, video communication, and various mobile video applications due to its high compression efficiency through the use of many advanced tools. In recent video applications, the video resolution tends to be larger and, thus, the bandwidth requirement for video transmission also increases. As a result, the importance of the bitrate control is growing as it is necessary to regulate the bitrate of a video stream in order to achieve the target bitrate. In bitrate control algorithms, the target bits are allocated at the frame-level or the macroblock-level by considering the fullness of the output buffer and the encoding complexity. While achieving the allocated bit budget, rate control attempts to make quality compromises in ways that would have a minimum degradation on perceived through controlling quantization parameter (QP) values. When the QP is small, much of the

information is preserved and when the QP is large, much of the information is discarded to reduce the bitrate at the cost of an increased distortion.

To achieve a target bitrate in real constraint situations, such as within a specific channel bandwidth or within defined encoder and decoder buffer sizes, QP values must vary dynamically based on the complexity of the input video and the current bitrate. QP values are determined using a rate-quantization (R-Q) model where the generated bitrate is modeled as a function of the QP and the complexity of the residual signal such as the mean absolute difference (MAD). In H.264/AVC, QPs are used not only in rate control, but also in rate distortion optimization (RDO). However, the famous “chicken-egg” dilemma complicates the selection of the QP value. The function for the RDO needs a pre-determined QP for its lambda factor. However, the QP can be determined by the MAD, which is only available after the RDO is obtained. To resolve that dilemma, the MAD is predicted from the previous frame using a linear function [1]. The estimated MAD from the previous frame is often different from the actual MAD in the

* Correspondence: jinsungk@sunmoon.ac.kr

²Department of Electronic Engineering of Sun Moon University, Asan 336-708, Korea

Full list of author information is available at the end of the article

current frame. Thus, an inadequate QP may be selected. For improved coding efficiency, enhanced distortion models [2-6] are developed and rate-distortion (R-D) model is optimized or contents-aware bit allocation is proposed [5-10]. A number of recent approaches incorporate characteristics of the human visual system (HVS) into bitrate control. Moreover, there are several reports on region-of-interest (ROI)-based bit allocation [11-13]; such approaches can potentially improve the perceived visual quality of images. In addition to ROI-based methods, a new R-D model, along with frame skipping and bit-allocation schemes, using various perceptual metrics that are based on the characteristics of HVS, is proposed [14-19]. However, the conventional bitrate control algorithms based on QP control often suffer from perceptible image quality degradations, such as blocking, ringing, or texture-deviation artifacts, when the target bitrate is very low. Furthermore, when the target bitrate is not satisfied, even with the maximum QP value, the conventional rate control cannot avoid a sudden frame drop which results in video quality degradation.

This article proposes a simple but effective bitrate control algorithm that applies spatial resolution controls to the conventional QP-based bitrate control. In the proposed algorithm, a new model that represents the relationship between the spatial resolution and the peak signal-to-noise ratio (PSNR) for low bitrate coding is proposed. By using the proposed model, the spatial resolution that gives the highest PSNR at a given bitrate is estimated. In the proposed model, the computational complexity is very low as it only requires a small number of parameters and the value of parameters is obtained in a heuristic manner. The two scalability tools, the QP and the spatial resolution, are processed sequentially to achieve the target PSNR. The proposed spatial resolution adjustment which is applied in the group of pictures (GOP)-level is used as a coarse-grain bitrate control. Inside a GOP, QP is changed by a conventional bitrate control to meet the allocated bits in a fine-grain manner. Thus, the target bitrate is satisfied with a combination of two control methods. To estimate the perceptual quality of encoded video sequences with reduced spatial resolution, Video Quality Metric (VQM) software [20] is used to measure the subjective quality in addition to the PSNR which measures the objective quality. The VQM computes the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and color distortion, and combines them into a single metric. Experimental results show that proposed bitrate control scheme outperforms the conventional QP-based bitrate control algorithm at a variety of bitrates. At a low bitrate, the PSNR and VQM values with the proposed spatial resolution control scheme are improved up to

1.85 and 5.15 dB, respectively, when compared to that with the conventional QP-only control. There is only a small difference between the real optimal spatial resolution and the spatial resolution obtained using the proposed scheme.

This article is organized as follows. In Section 2, background is presented and Section 3 explains the proposed bitrate control algorithm. Experimental results are presented in Section 4 and conclusions are given in Section 5.

2. Background

2.1. Previous study on spatial and temporal resolution controls

To improve flexibility in bit allocation, some rate control algorithms adjust frame rate or spatial resolution. When the frame rate decreases, additional bits can be allocated to each frame and frame image quality can be improved. However, frame skipping should be done very carefully because motion artifacts such as flickering or motion jerkiness may degrade subjective video quality. In [21,22], the decision for the frame skip is based on buffer fullness and the spatial and temporal quality of the video. In [23], the similarity between successive frames measured by the PSNR is used to skip frames adaptively. To optimize coding performance, frames to be skipped are determined based on an R-D model [24,25], but these works cannot be used in real-time applications. In [26-28], motion artifacts are reduced by adjusting the frame rate gradually based on the motion activity of the previous sub-GOP which is expressed by the histogram of difference image, thereby preserving motion smoothness. Even though a number of previous works have contributed to frame rate controls, the effect of providing bitrate control through frame rate adjustment is somewhat limited because, to avoid motion artifacts, the frame rate cannot be reduced below a certain value. In addition, temporal scalability may not be very effective for increasing subjective quality by temporal to spatial bitrate exchange. This is because the quality degradation due to dropping a frame is easily perceived, especially in low frame rate communications such as that used in two-way multimedia communication for mobile devices [29].

Spatial resolution control is another approach used for bit allocation when the target bitrate is very low. The dynamic resolution conversion (DRC) mode, which is supported in the advanced simple profile of MPEG-4, enables the video object plane to be encoded with reduced spatial resolution [30]. Similarly, a reduced resolution update (RRU) coding tool is adopted in Annex Q of the H.263 standard [31]. The RRU reduces the bitrate by coding the prediction error residuals at a reduced spatial resolution. However, the DRC and RRU

techniques are not included in the H.264/AVC standard. Meanwhile, a number of previous studies have reported a relationship between down-sampling and video quality at a low bitrate. In [32], the optimum down-sampling ratio is determined according to the bitrate. In [33], it is reported that a down-sampled video, prior to compression and later up-sampled, visually outperforms that video compressed directly at high resolution with the same number of bits, when the target bitrate is very low. With this observation, a method to find the optimal down-sampling rate is suggested. These schemes are exploited by JPEG image compression standard not by video compression standards [29]. In [34,35], discrete cosine transform (DCT) coefficients are decimated prior to quantization to reduce spatial resolution, but modifying the coding loop loses the conformation to the syntax compatibility for the video coding standard. In video transcoders, spatial resolution control has been an important factor in meeting a different target bitrate [36]. However, many previous works have focused on simplifying the computations in transcoding. In [37], the linear R-Q model is proposed to select the proper frame size, but it is not applicable to practical applications. Recently, in [38], the overall distortion is analyzed and the optimal spatial resolution is derived for a given bitrate. In [39], the spatial resolution ratio is appropriately selected, according to picture quality, bit rate, and power consumption. As shown in the above-mentioned works, in order to accomplish bitrate control by spatial resolution control, it is very important to select the optimal spatial resolution. When the selected spatial resolution is greater than the optimal one, the image quality can be degraded by using an excessively high QP value, while the quality can be degraded by aliasing artifacts when the selected spatial resolution is less than the optimal one. Nonetheless, it is difficult to find the optimal resolution because complex estimation models, which define the relationships among picture quality, bitrate, spatial resolution, and power consumption, are used in those previous works. Moreover, parameters used in the previous methods depend on the characteristics of the video content and the specific coding methods; thus, they cannot be

calculated in real time. Therefore, the previous works cannot be applied to “on the fly”, real-time rate control.

2.2. Comparison between spatial and temporal resolution controls

Rate control plays a critical role in the video encoder. However, the rate control algorithm is not standardized because it is independent of the decoder. For an enhanced rate control, additional control factors, such as frame rate and spatial resolution, can be used. The HVS is less sensitive to temporal details and more sensitive to spatial details, if a video is stationary. However, for video with high motion, the opposite is true. Empirically, a subjective quality degradation caused by a frame drop is more serious than that caused by spatial down-sampling at low bitrate. In Table 1, high definition (HD) video sequences are encoded at the target bitrates of 400 and 250 kbps and the spatial and temporal resolution controls are compared. When the target bitrate cannot be satisfied, even with a maximum QP value, a dyadic spatial resolution reduction (denoted by s_drop) or a dyadic temporal resolution reduction (denoted by t_drop) is applied. To implement t_drop , every alternate frame is dropped; subsequently, the undropped frames are repeated twice to replace the dropped frames. The VQM value of the Y component is measured to evaluate the subjective quality where lower VQM values represent a better subjective quality. A sequence-to-sequence comparison is made. To get the VQM of t_drop , the sequence which consists of the dropped and repeated frames as well as the undropped frames is compared with the original sequence. As shown in Table 1, s_drop always outperforms t_drop . This experiment shows that increasing the bitrate per frame by lowering the frame rate for a spatial quality does not lead to a higher subjective quality than that from a method using spatial scalability for any sequence.

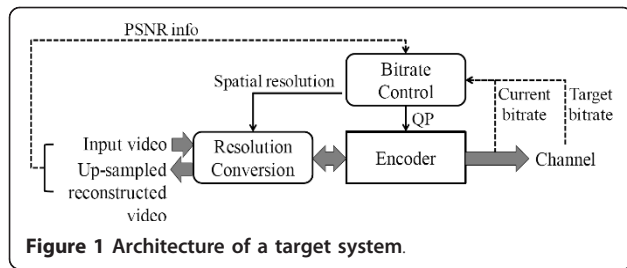
3. The proposed bitrate control with spatial scalability

3.1. Architecture of the target system

Figure 1 shows a block diagram of the target system used in this research. The input video is encoded by an

Table 1 Comparison of VQM values between spatial and temporal drops at low target bitrates

| | Target bitrate (kbps) | Sequences | | | | |
|---------------------|-----------------------|-----------|-----------------|-----------|-----------|---------|
| | | Bluesky | Pedestrian area | Station 2 | Sunflower | Tractor |
| VQM (Y) | | | | | | |
| t_drop | 400 | 6.94 | 4.75 | 2.48 | 3.78 | 4.29 |
| | 250 | 11.33 | 8.02 | 4.78 | 7.9 | 6.86 |
| s_drop | 400 | 3.03 | 2.51 | 2.01 | 2.16 | 3.25 |
| | 250 | 3.46 | 2.69 | 2.44 | 2.4 | 3.18 |
| $t_drop - s_drop$ | 400 | 3.49 | 2.06 | 0.04 | 1.38 | 1.12 |
| | 250 | 7.84 | 5.96 | 4.74 | 6.52 | 5.75 |



H.264/AVC encoder and the reconstructed frames are stored for the next frames. It is assumed that both the encoder and decoder sides include modules to adjust the spatial resolution of the input and displaying videos, respectively. Based on the target bitrate and other encoding results, such as the current bitrate and PSNR information, the bitrate control module determines the proper QP value and the spatial resolution ratios for the encoder and the resolution conversion module, respectively. For implementation of the proposed bitrate control module in Figure 1, the bitrate control algorithm in the JM 13.2 reference software is used for the control of the QP value, whereas a new algorithm, described next, is proposed for spatial resolution control. In this article, the spatial resolution for the best video quality is determined by considering the PSNR which indicates objective quality. PSNR values are obtained from the difference between the original video and the up-sampled reconstructed video. The reason why the PSNR is chosen over VQM is that the VQM calculation is computationally expensive and needs the buffering of a few frames whereas the PSNR computation is quite simple.

3.2. Spatial resolution control

The reconstructed and up-sampled video includes two kinds of distortion. One is generated from the encoding process and the other is caused by spatial up/down-sampling. When spatial resolution control is used for bitrate control, it is important to find a resolution ratio that shows the best quality video under a given target bitrate. According to [39], experiments have shown that the PSNR degradation due to down-sampling and up-sampling operations increases approximately in proportion to the bitrate and the extent of the reduction in spatial resolution. Let $PSNR_{coding_down}$ denote the PSNR of a video that has been down-sampled and encoded. Then, $PSNR_{coding_down}$ is formulated as

$$PSNR_{coding_down} = q_1 \cdot \log R + q_2 - q_3 \cdot \left(\frac{1}{sa} - 1 \right) \cdot R \quad (1)$$

where q_1 , q_2 , and q_3 are constants which depend on the video content and R is the bitrate of the encoded

stream. The term sa , referred to here as the spatial resolution ratio, represents the ratio of the down-sampled frame area to the original frame area. When sa is smaller than 1, the frame is down-sampled. Equation (1) describes the relationship between sa and PSNR and thus is used for calculating the optimal spatial resolution. However, the parameters such as q_1 , q_2 , and q_3 in Equation (1) depend on the video content and cannot be known prior to encoding. Thus, this optimal solution cannot be applied to real-time systems on the fly.

Figure 2 shows the PSNR of HD-sized video sequences, *Station2*, at various spatial resolution ratios and at three bitrates: 600 kbps, 1 Mbps, and 2 Mbps. In Figure 2, the solid curves show the PSNR values obtained by simulation. Each graph has a peak PSNR value at a certain sa . The initial PSNR without spatial down-sampling is denoted by $PSNR_{full}$ (when $sa = 1$). The peak PSNR of each graph is denoted by $PSNR_{peak}$ and the sa that gives $PSNR_{peak}$ is denoted by sa_{peak} . In the graph for the 600 kbps bitrate in Figure 2, $PSNR_{full}$ is marked with a circle while $PSNR_{peak}$ and sa_{peak} are marked with a triangle and a rectangle, respectively. If the spatial resolution ratio is adjusted to sa_{peak} , then the highest PSNR is achieved for a given bitrate.

In order to reduce the complexity of calculating sa_{peak} , a method for finding sa_{peak} based on a simplified model obtained from Figure 2 is proposed. In Figure 2, dotted lines connect $PSNR_{full}$ and $PSNR_{peak}$. Within the range of sa from 1 to sa_{peak} , the dotted lines are very close to the measured data. Based on this proximity, the equation for $PSNR_{coding_down}$ in Equation (1) is reformulated as a simplified model for low bitrate control as shown in Equation (2) in which α and β are positive and α represents the slope in the modeled graphs. If the information for α , $PSNR_{full}$, and $PSNR_{peak}$ are given, the sa_{peak} can be estimated using the linear model in

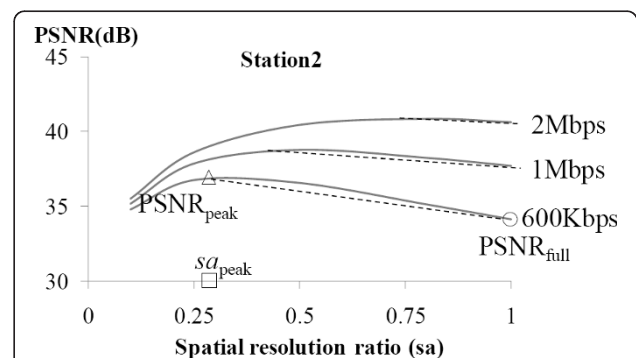


Figure 2 PSNR depending on the spatial resolution ratio at various low bitrates obtained with HD-sized Station2 sequence.

Equation (2).

$$\text{PSNR}_{\text{coding_down_lowR}} = -\alpha \cdot sa + \beta. \quad (2)$$

The slope of the model, α , can be estimated using the PSNR and sa values obtained from encoded frames. Let $\text{PSNR}_{\text{prev}}$ and sa_{prev} denote the PSNR and sa of the previous GOP, respectively, and let $\text{PSNR}_{\text{full}}$ be obtained when the first GOP is encoded with $sa = 1$. Based on the linear model in Equation (2), $\text{PSNR}_{\text{prev}}$ and $\text{PSNR}_{\text{full}}$ are expressed as given in Equations (3) and (4), respectively.

$$\text{PSNR}_{\text{prev}} = -\alpha \cdot sa_{\text{prev}} + \beta \quad (3)$$

$$\text{PSNR}_{\text{full}} = -\alpha \cdot 1 + \beta \quad (4)$$

If (4) is subtracted from (3), an estimate of α , denoted by α_{est} , is obtained by (5). The slope α_{est} obtained from (5) is the α value delayed by one GOP. It is based on the assumption that spatial resolution of the current frame can be applied to the next GOP, due to similarity between successive GOPs.

$$\alpha_{\text{est}} = \frac{(\text{PSNR}_{\text{prev}} - \text{PSNR}_{\text{full}})}{1 - sa_{\text{prev}}} \quad (5)$$

$\text{PSNR}_{\text{peak}}$ cannot be simply determined because it varies with both the image content and the target bitrate. In order to estimate $\text{PSNR}_{\text{peak}}$, $\text{PSNR}_{\text{peak}}/\text{PSNR}_{\text{full}}$ ratios of 12 video sequences are measured under various target bitrates and at various values of slope α . The used videos are as follows: *Akiyo* and *Coast Guard*, with CIF (352×288) resolution, *City*, *Crew*, and *Ice*, with 4CIF (704×576) resolution, *Aspen*, *Factory*, *Old Town Cross*, *Parkrun*, and *Pedestrian Area* with HD (1280×720) resolution and *West Windy Easy* and *Touchdown Pass*, both with full HD (1920×1080) resolution are used in the evaluation. A sample of such results is shown in Figure 3 which shows that $\text{PSNR}_{\text{peak}}/\text{PSNR}_{\text{full}}$ ratio is approximately proportional to the slope α regardless of video sequence types and bitrates. From this observation, $\text{PSNR}_{\text{peak}}$ is calculated as given in (6), with the coefficients being chosen experimentally.

$$\text{PSNR}_{\text{peak}} = (0.03 \times \text{PSNR}_{\text{full}} \times (\alpha - 0.5)) + \text{PSNR}_{\text{full}} \quad (6)$$

Based on the linear model in (2), sa_{peak} is calculated within the range from 0.1 to 1 as

$$sa_{\text{peak}} = 1 - \frac{(\text{PSNR}_{\text{peak}} - \text{PSNR}_{\text{full}})}{\alpha} \quad (0.1 \leq sa_{\text{peak}} \leq 1) \quad (7)$$

where α and $\text{PSNR}_{\text{peak}}$ are obtained from (5) and (6), respectively.

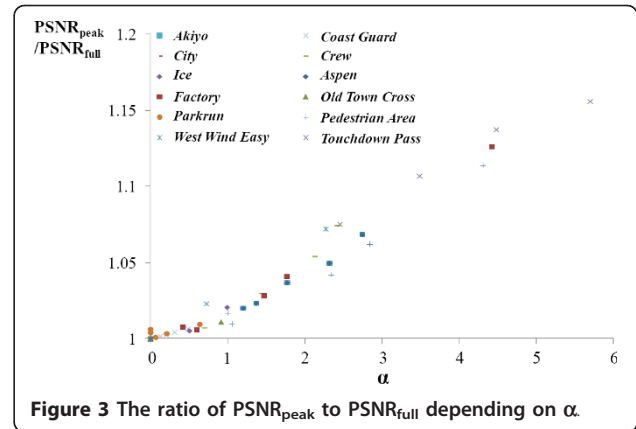


Figure 3 The ratio of $\text{PSNR}_{\text{peak}}$ to $\text{PSNR}_{\text{full}}$ depending on α .

Slope α cannot be estimated before the second GOP because $\text{PSNR}_{\text{prev}}$ and sa_{prev} in (5) are not given yet. In this case, the initial value of α is set to a small value from 0.1 to 0.3. This initial α is referred to as α_{init} . If $\text{PSNR}_{\text{peak}}$ in (6) is substituted in (7), sa_{peak} is in inverse proportional to α . Thus, if a small α_{init} is used in (7), sa_{peak} value is relatively large and thus, the PSNR degradation caused by an excessive down-sampling operation can be avoided. Once $\text{PSNR}_{\text{prev}}$ and sa_{prev} are obtained from the results for the second GOP encoded with α_{init} , then α_{est} calculated from (5) is used as follows:

$$\alpha = \begin{cases} \alpha_{\text{init}} & \text{for the second GOP} \\ \alpha_{\text{est}}, & \text{after the second GOP} \end{cases} \quad (8)$$

The target spatial resolution ratio, sa_{target} , is adjusted only when the video quality is lower than an acceptable level. Let $\text{PSNR}_{\text{target}}$ denote the PSNR for the acceptable image quality as required by users or applications. It is assumed that the reconstructed image is visually indistinguishable from the original one if the PSNR is greater than from 35 to 40 dB [40-43]. Thus, $\text{PSNR}_{\text{target}}$ is set to 40 dB in this article. If $\text{PSNR}_{\text{full}}$ is larger than $\text{PSNR}_{\text{target}}$, no spatial resolution adjustment is necessary, that is, the target spatial resolution ratio is set to 1. Otherwise, sa_{target} is adjusted to the sa_{peak} as in (7). Thus, sa_{target} is given by

$$sa_{\text{target}} = \begin{cases} 1, & \text{if } (\text{PSNR}_{\text{full}} \geq \text{PSNR}_{\text{target}}) \\ sa_{\text{peak}}, & \text{else} \end{cases} \quad (9)$$

3.3. The proposed bitrate control algorithm

Figure 4 shows the proposed bitrate control algorithm in which the QP and the spatial resolution ratio are determined sequentially in order to reach the $\text{PSNR}_{\text{target}}$. The QP value is decided frame-by-frame, whereas the spatial resolution ratios are determined for each GOP. At initialization, the $\text{PSNR}_{\text{target}}$ is defined, and the algorithm

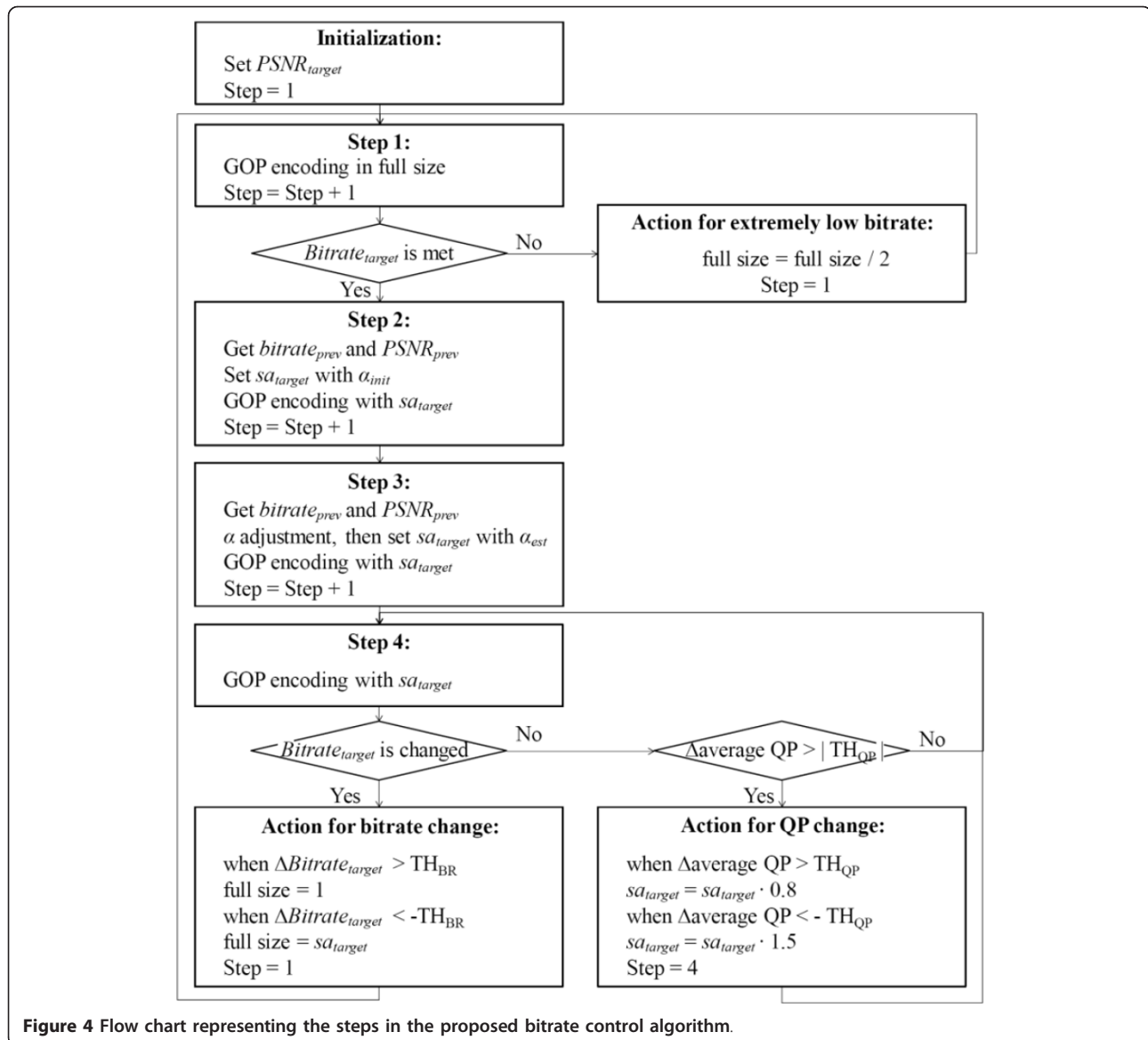


Figure 4 Flow chart representing the steps in the proposed bitrate control algorithm.

step is started at Step 1. In the first GOP of Step 1, the spatial resolution ratio is not changed but only the QP is controlled to meet the target bitrate like a conventional QP-based rate control algorithm. If the target bitrate denoted by $\text{bitrate}_{\text{target}}$ in Figure 4 cannot be satisfied in Step 1, the spatial resolution for the next GOP is simply down-sampled by a factor of 2, compared to that of the current GOP because meeting the $\text{bitrate}_{\text{target}}$ is paramount. The encoding for the next GOP is started from Step 1 with a half-reduced spatial resolution. If the generated bitrate of the GOP in Step 1 meets the $\text{bitrate}_{\text{target}}$ in Figure 4, Step 1 proceeds to Step 2. In Step 2, the sa_{peak} and $\text{PSNR}_{\text{peak}}$ for the GOP to be encoded are calculated from (6) and (7), where the PSNR obtained from Step 1 and α_{init} are used. The

sa_{target} is then determined from (9). In Step 3, the PSNR obtained from Step 2 is evaluated and used as $\text{PSNR}_{\text{prev}}$ in (5) to adjust the proper α . Using the adjusted α denoted by α_{est} , the sa_{peak} , $\text{PSNR}_{\text{peak}}$, and sa_{target} for the GOP in Step 3 are calculated from (6), (7), and (9), respectively. The sa_{target} , determined once in Step 3, is used continually for the subsequent GOPs in Step 4. As long as the R-D characteristics of successive frames are similar, encoding with a sa_{target} works well. To cope with varying R-D characteristics, actions for bitrate change and QP change are described in Figure 4. If $\text{bitrate}_{\text{target}}$ is changed, the relation between the spatial resolution and the PSNR becomes different, thus the slope α_{est} value needs to be refreshed through Steps 1, 2, and 3. Before going back to Step 1, the full size is set

to 1 which is a full resolution with no reduction of spatial resolution when the increase of $\text{bitrate}_{\text{target}}$ is greater than TH_{BR} . If the decrease of $\text{bitrate}_{\text{target}}$ is greater than TH_{BR} , the full size is set to the current sa_{target} . A new sa_{target} for the decreased target bitrate will be determined as a value less than the current sa_{target} . In this article, TH_{BR} is calculated by using $0.02 \times \text{frame rate} \times \text{original spatial resolution}$. Even though the $\text{bitrate}_{\text{target}}$ is the same, the motion characteristics of video can be changed. If the motion is faster, the average QP value of the recent frames becomes higher than that of the previous ones, and vice versa. Therefore, sa_{target} is adjusted in a fine-grain manner by the change in the average QP. As shown in Figure 4, the whole flow to decide the proper spatial resolution is processed automatically and does not depend on advanced information about characteristics of the video content and the specific coding methods. Therefore, the proposed algorithm can easily be applied to real-time applications.

4. Experimental results

The proposed bitrate control scheme is implemented and integrated into the JM 13.2 reference software which adopts the QP-based bitrate control. To resize the spatial resolution, the up/down-sampling algorithms recommended in the SVC are used. For down-sampling, the algorithm based on the Sine-windowed Sinc-function is applied where a set of seven filters is used to support the extended range of the spatial scaling ratio. For up-sampling, the SVC normative up-sampling algorithm is applied which is based on a set of 6-taps filters derived from the Lanczos-3 filter. In this experiment, five HD video sequences, *Pedestrian Area*, *Tractor*, *Station2*, *Sunflower*, and *Blue Sky*, two full HD video sequences, *Speed Bag* and *Life* and two 4CIF videos, *Harbor* and *Soccer*, are used. The chosen length of a

GOP is 30 frames and 150 frames are encoded. The GOP structure is IPPP.

Table 2 shows the average PSNR and VQM values of Y component for the fourth and the fifth GOPs. In the second column, 'Conventional' represents that the nine video sequences are encoded in the full size without spatial resolution control. When the target bitrate cannot be satisfied even with the maximum QP value, the frame rate is decreased by a half. The frame drop is realized by a factor of 2 in the encoder side which encodes all the macroblocks as the SKIP mode in the dropped frame. Thus, in the decoder side, frames are displayed at 30 fps with the dropped frame by a repetition of the previous frame. 'Proposed' and 'Optimal' represent, respectively, the proposed spatial resolution control and the optimal spatial resolution values. The optimal values are obtained experimentally by changing the resolution ratios from 0.1 to 1 and by measuring the PSNR values. Unlike a full size encoding, the rate control schemes represented by 'Proposed' and 'Optimal' do not use a frame-drop to meet the target bitrate. Note that the optimal resolution is a theoretical upper-bound of the spatial control and cannot be calculated on the fly. Experiments are conducted using these three rate control schemes with various bitrates: 250, 400, 600, 800, and 1000 kbps for HD, 400, 800, 1500, and 2000 kbps for full HD and 150, 300, 400, and 1000 kbps for 4CIF. In the result of HD with the high target bitrates of 1000 kbps, PSNR and VQM enhancements by the proposed rate control are 0.82 and 0.12 dB, respectively. As the target bitrate decreases, the VQM enhancement by the proposed rate control is increased by 5.15. Note that the PSNR difference between the proposed rate control and the conventional QP control is not large at the target bitrates of 400 and 250 kbps, whereas the PSNR enhancement by the proposed algorithm is 1.85 dB at

Table 2 Comparison of the PSNR and VQM among the conventional control, the proposed spatial resolution and the optimal spatial resolution at various target bitrates

| Measurement tools | Rate control methods | Target bitrate (kbps) | | | | | | | | | | | | |
|-------------------|-------------------------------|-----------------------|-------|-------|-------|-------|-------------|-------|-------|-------|-------|-----------|-------|-------|
| | | 1280 × 720 | | | | | 1920 × 1072 | | | | | 704 × 576 | | |
| | | 250 | 400 | 600 | 800 | 1000 | 400 | 800 | 1500 | 2000 | 150 | 300 | 400 | 1000 |
| PSNR (Y) | Conventional | 29.14 | 30.88 | 30.49 | 32.53 | 33.75 | 19.56 | 26.99 | 29.75 | 30.90 | 23.11 | 27.43 | 27.20 | 30.25 |
| | Proposed | 30.14 | 31.08 | 32.35 | 33.74 | 34.57 | 31.70 | 33.55 | 35.36 | 36.11 | 25.40 | 27.20 | 27.97 | 30.70 |
| | Optimal | 30.36 | 31.54 | 32.74 | 33.88 | 34.68 | 32.07 | 33.90 | 35.75 | 36.57 | 25.78 | 27.45 | 27.99 | 31.06 |
| | Difference(Full, Proposed) | 1 | 0.2 | 1.85 | 1.21 | 0.82 | 12.15 | 6.56 | 5.60 | 5.20 | 2.29 | -0.24 | 0.77 | 0.45 |
| | Difference(Optimal, Proposed) | 0.22 | 0.46 | 0.39 | 0.14 | 0.12 | 0.37 | 0.35 | 0.39 | 0.46 | 0.39 | 0.25 | 0.02 | 0.37 |
| VQM (Y) | Conventional | 7.78 | 4.45 | 2.65 | 2.04 | 1.8 | 9.61 | 5.46 | 2.82 | 2.45 | 7.32 | 4.99 | 2.94 | 2.16 |
| | Proposed | 2.63 | 2.36 | 2.13 | 1.81 | 1.68 | 2.82 | 2.20 | 1.78 | 1.59 | 3.50 | 2.89 | 2.76 | 2.02 |
| | Optimal | 2.65 | 2.35 | 2.07 | 1.84 | 1.68 | 2.63 | 2.11 | 1.65 | 1.50 | 3.46 | 2.82 | 2.75 | 2.00 |
| | Difference(Full, Proposed) | 5.15 | 2.09 | 0.51 | 0.23 | 0.12 | 6.79 | 3.26 | 1.04 | 0.86 | 3.82 | 2.10 | 0.18 | 0.14 |
| | Difference(Optimal, Proposed) | -0.02 | 0.01 | 0.06 | -0.03 | 0 | 0.19 | 0.09 | 0.13 | 0.10 | 0.04 | 0.07 | 0.02 | 0.02 |

the target bitrate of 600 kbps. This is because the frame rate is decreased in the ultra-low bitrate in the case of the conventional algorithm. Thus, at the target bitrates of 400 and 250 kbps, the PSNR of each frame encoded with the conventional algorithm is a little enhanced because the allocated bits per frame are increased. In this experiment, the optimal spatial resolution is chosen because it has the highest PSNR value than the other resolutions. The VQM value with the optimal spatial resolution could be a little worse than the one calculated from the proposed bitrate control like the cases at bitrate 250 and 800 kbps because of the difference between the two calculation methods, PSNR and VQM. However, the optimal PSNR values are always higher than the results from the conventional or the proposed bitrate control. For full HD and 4CIF as well as HD, the PSNR and VQM of the proposed spatial resolution are very close to that of the optimal one as shown in Table 2.

Figure 5 presents the R-D performance for three video sequences when using the proposed spatial resolution control. Both objective (PSNR) and subjective (VQM) results are presented to show the performance of the proposed algorithm. The gray solid curve represents the results of encoding with optimal spatial resolution whereas the black solid curve represents the results when the proposed algorithm is applied. The conventional rate control is represented by the black dotted curve. In Figure 5a-c, at the target bitrate of 600 kbps, the proposed algorithm enhances PSNR by 2.2, 3.1, and 3.4 dB, respectively, compared to results using the conventional rate control. Furthermore, the differences

between 'Optimal' and 'Proposed' are negligible. As the target bitrate increases, PSNR improvement decreases. When the target bitrate is extremely low, such as 250 and 400 kbps, the allocated bits per pixel (bpp) are just 0.009 and 0.014 bpp, respectively. A conventional QP-based bitrate control cannot meet the target bitrate, even when using the maximum QP value. In the 'Conventional' method of Figure 5, frames are dropped to meet the target bitrate. Thus, video sequences are encoded at 15 and 7.5 fps for 400 and 250 kbps, respectively, whereas the frame rate is 30 fps in 600, 800, and 1000 kbps. For three video sequences, *Sunflower*, *Pedestrian Area*, and *Tractor* at 400 kbps, frame skips work for PSNR enhancement, to a limited amount, because the allocated bits per pixel are increased. However, additional frame skips, used to meet the target bitrate of 250 kbps, do not help the quality enhancement of each frame for the *Sunflower* and *Tractor* videos. Because the temporal correlation between frames becomes low, it results in low compression efficiency. The PSNR of the *Pedestrian Area* video is increased a little more at 250 kbps. In general, the results depend on the characteristic of the video sequence. In Figure 5d-f, VQMs clearly show that the proposed algorithm produces a significant improvement compared to the conventional rate control. The proposed rate control algorithm maintains a similar VQM quality from 1000 to 250 kbps, while the VQM of the conventional rate control increases drastically as the target bitrate decreases.

In Table 3, the average values of *sa*, as determined in the experiments summarized in Table 2, are shown. As the target bitrate increases, the 'Optimal' *sa* also

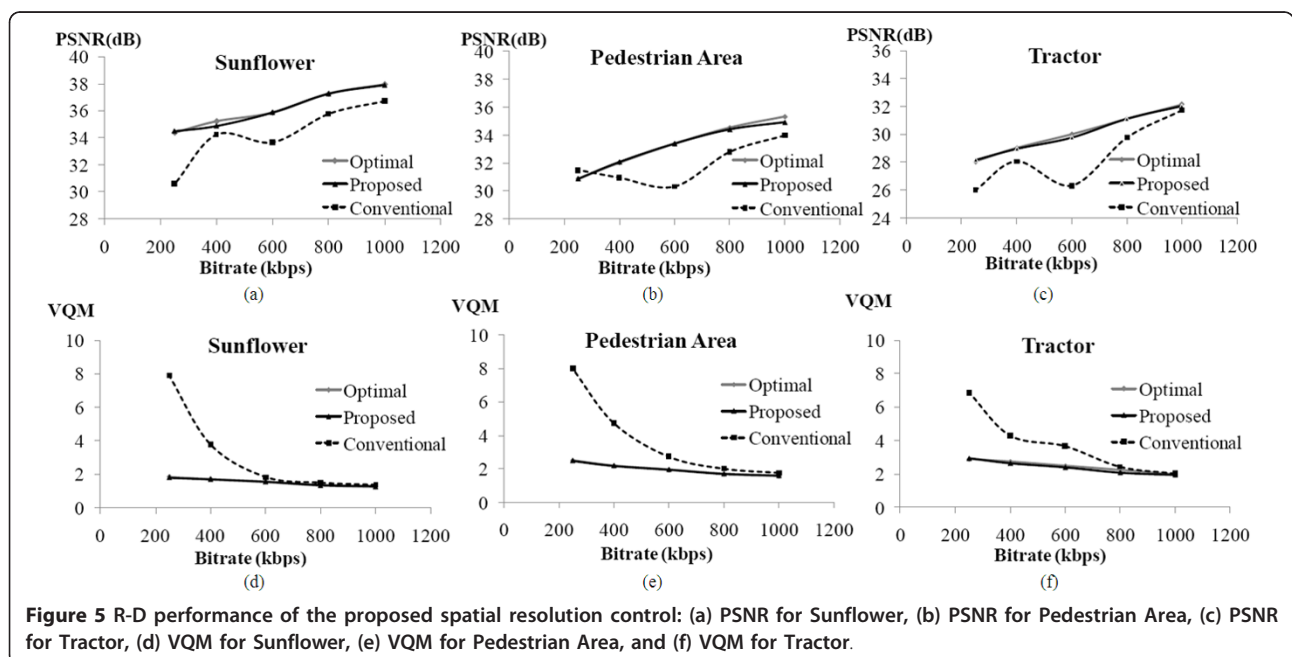


Figure 5 R-D performance of the proposed spatial resolution control: (a) PSNR for Sunflower, (b) PSNR for Pedestrian Area, (c) PSNR for Tractor, (d) VQM for Sunflower, (e) VQM for Pedestrian Area, and (f) VQM for Tractor.

Table 3 Comparison of the SA between the proposed spatial resolution and the optimal spatial resolution at various target bitrates

| Rate control methods | 1280 × 720 | | | | | 1920 × 1072 | | | | 704 × 576 | | | |
|--------------------------------|-----------------------|------|------|------|------|-------------|------|------|------|-----------|------|------|------|
| | Target bitrate (kbps) | | | | | | | | | | | | |
| | 250 | 400 | 600 | 800 | 1000 | 400 | 800 | 1500 | 2000 | 150 | 300 | 400 | 1000 |
| SA | | | | | | | | | | | | | |
| Proposed | 0.06 | 0.10 | 0.33 | 0.36 | 0.41 | 0.19 | 0.36 | 0.73 | 0.73 | 0.08 | 0.23 | 0.51 | 0.51 |
| Optimal | 0.04 | 0.10 | 0.26 | 0.26 | 0.36 | 0.16 | 0.34 | 0.64 | 0.82 | 0.12 | 0.26 | 0.44 | 0.58 |
| Difference (optimal, proposed) | 0.02 | 0.00 | 0.07 | 0.10 | 0.05 | 0.02 | 0.02 | 0.09 | 0.09 | 0.04 | 0.03 | 0.07 | 0.08 |

increases. The values of *sa* determined by the 'Proposed' algorithm are very similar to those from the 'Optimal' one. In this experiment, the difference between 'Optimal' and 'Proposed' *sa* values is, on average, just 0.05.

In Figure 6, the values of *sa* for each GOP are presented for three HD video sequences, *Sunflower*, *Pedestrian Area*, and *Tractor*, which are encoded with various bitrates. In Figure 6a-c, the vertical axis shows *sa* while the horizontal axis shows the GOP number. Step 1 of Figure 4 starts from GOP 1. For 400 and 250 kbps, the extremely low bitrate control is already carried out to meet the target bitrate, as explained in Figure 4. Thus, the full sizes in GOP 1 for 400 and 250 kbps are 0.5 and 0.25, respectively. Through the proposed spatial resolution adjustment from GOP 1 to GOP 3, the *sa* is determined and stabilized. In Figure 6d-f, the *sa* for each GOP is compared to the optimal one. The vertical axis in Figure 6d-f shows the difference between the

proposed and optimal *sa* values. To determine the difference, the optimal *sa* is subtracted from the proposed one. As shown in these graphs, the differences are very small.

Figures 7 and 8 show the 76th frame of the *Sunflower* sequence and the 135th frame of the *Pedestrian Area* sequence, respectively. The result of applying the conventional bitrate control, which uses only QP change, is shown in Figures 7a and 8a. As observed in these figures, much of the details in the frame is destroyed or decreased. In Figures 7b and 8b, QP and spatial resolution are controlled by the proposed rate control and, subjectively, the quality is better than the quality in Figures 7a and 8a.

Figure 9 shows the performance of the proposed algorithm when the motion characteristics are changed during encoding operations. For that figure, 300 frames of the 4CIF *Soccer* video sequence are encoded. The length

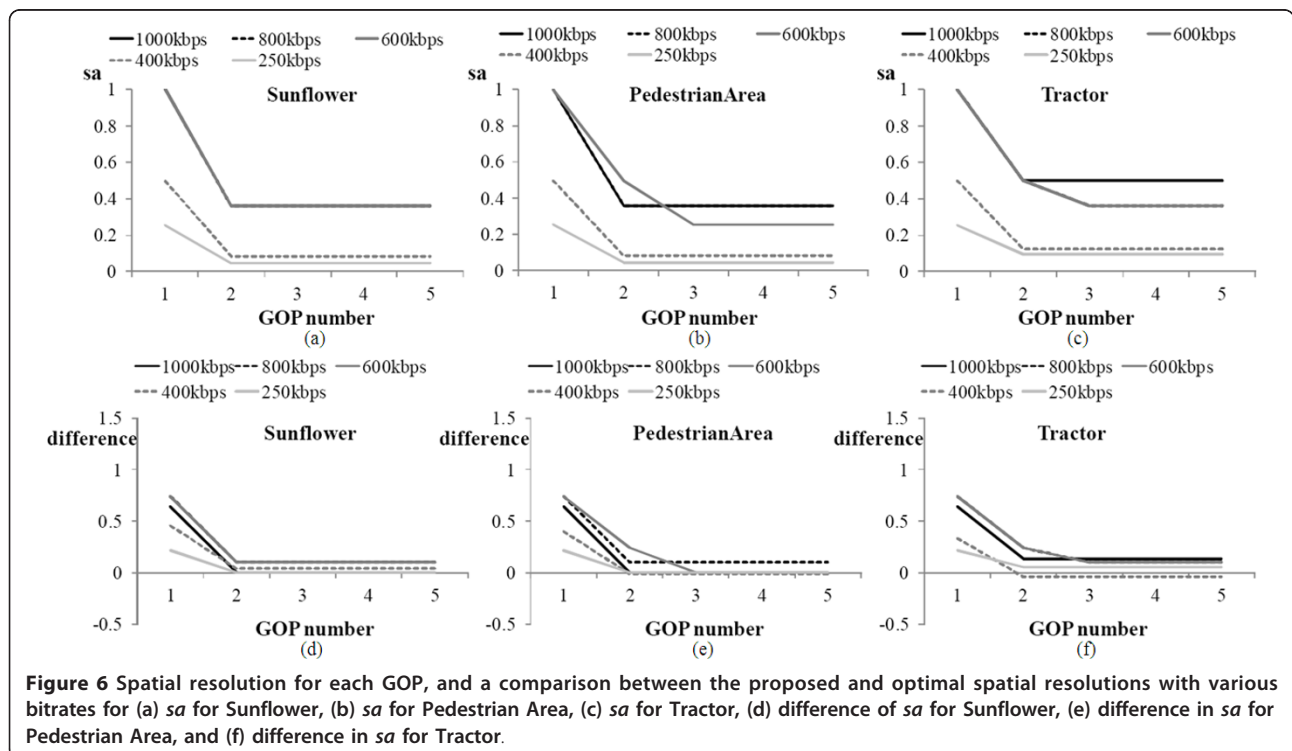


Figure 6 Spatial resolution for each GOP, and a comparison between the proposed and optimal spatial resolutions with various bitrates for (a) *sa* for Sunflower, (b) *sa* for Pedestrian Area, (c) *sa* for Tractor, (d) difference of *sa* for Sunflower, (e) difference in *sa* for Pedestrian Area, and (f) difference in *sa* for Tractor.



Figure 7 76th frame image of Sunflower sequence with 600 kbps: (a) conventional rate control, (b) proposed rate control.

of the GOP is 30. The values of sa and PSNR for each GOP are presented in Figure 9a, b, respectively. In that figure, the dotted graphs represent the results obtained from the proposed algorithm while the solid graphs represent the optimal values. In the 300 frames of *Soccer*, the first half has a slower motion than the second half. Until GOP 5, the sa value is determined to be 0.26 by the proposed algorithm. After that, the motion becomes faster, and consequently, the QP values are increased. Therefore, the sa value for the second half of the sequence is adjusted downward. In Figure 9a, sa is set to 0.19 for GOPs 6, 7, 8, and 9 and 0.13 for GOP

10. From GOP 3 to GOP 10, the differences between 'Proposed' and 'Optimal' values are just 0.02, on average. In Figure 9b, the PSNR values for 'Proposed' and 'Optimal' cases are compared and the differences are negligible.

In this experiment, the performance of the proposed algorithm is tested when the target bitrate is changed in the middle of the encoding process. The 4CIF *Harbor* video sequence is used as an example. In Figure 10a, b, the target bitrate for the first five GOPs is 300 kbps, whereas the target bitrate for the last five GOPs is 1500 kbps. In Figure 10a, the optimal spatial resolutions are

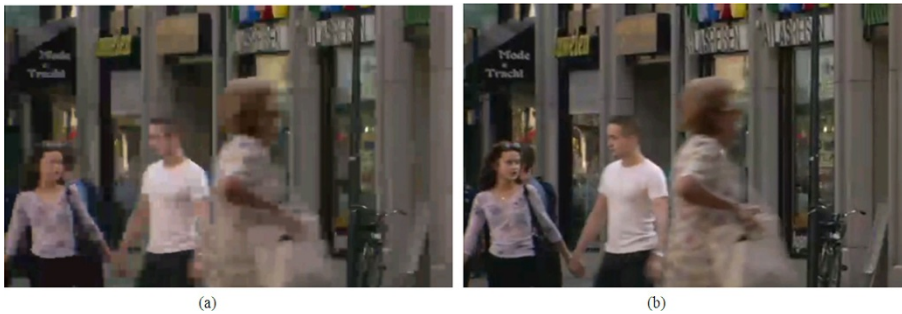


Figure 8 135th frame image of Pedestrian Area sequence with 600 kbps: (a) conventional rate control; (b) proposed rate control.

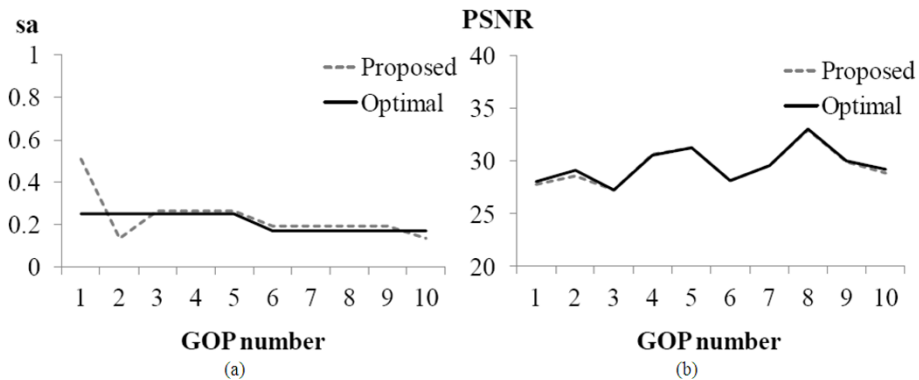


Figure 9 Performance of the proposed spatial resolution control when the motion characteristics are changed for *Soccer* in terms of (a) sa and (b) PSNR.

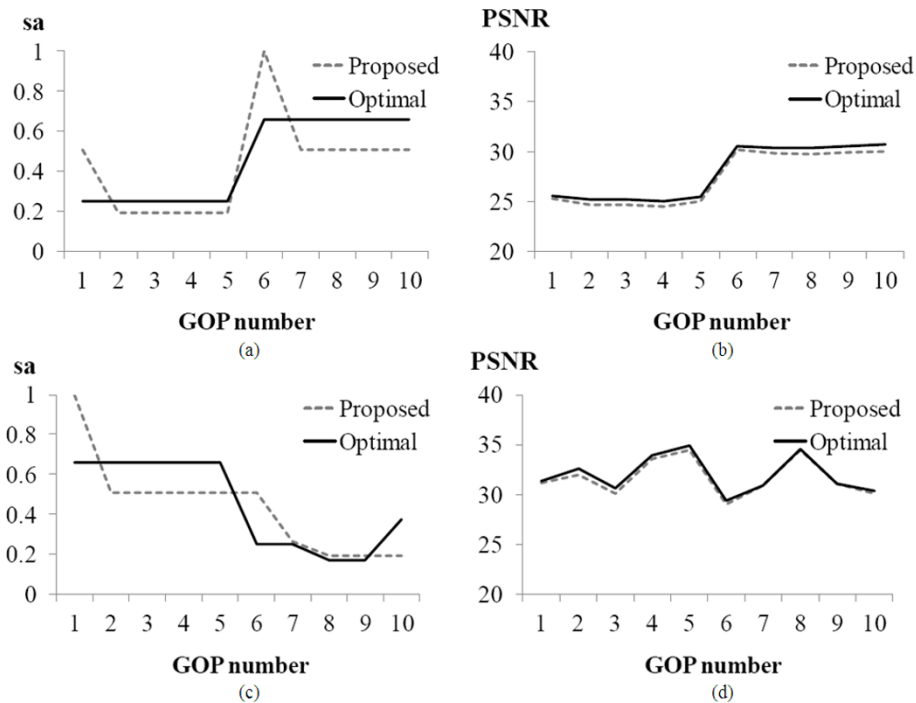


Figure 10 Performance of the proposed spatial resolution control when the target bitrate is changed for *Harbor* in terms of (a) *sa* change when the target bitrate is changed from 300 to 1500 kbps, (b) PSNR when the target bitrate is changed from 300 to 1500 kbps, (c) *sa* change when the target bitrate is changed from 1000 to 500 kbps, and (d) PSNR when the target bitrate is changed from 1000 to 500 kbps.

0.25 and 0.66 for the first and last half, respectively. In the proposed algorithm, *sa* is chosen initially as 0.5 because the 300 kbps target bitrate is extremely low; subsequently, the *sa* is determined to be 0.2 by the proposed algorithm. From GOP 5 onward, the target bitrate is increased to 1500 kbps. Thus, *sa* is set to 1 in GOP 5 and the process to determine the proper spatial resolution is restarted. On the basis of the bitrate change action explained in Figure 4, *sa* for the 1500 kbps target bitrate is determined to be 0.5. In Figure 10b, the PSNR values for the 'Proposed' and 'Optimal' cases are compared and the difference is 0.5 dB, on average. In Figure 10c, d, the target bitrate is reduced from 1000 to 500 kbps. For the first five GOPs, *sa* is decreased to 0.5 for 1000 kbps by the proposed algorithm. For the following five GOPs, *sa* is reduced because the target bitrate is reduced to 500 kbps. As shown in Figure 10c, the shape of the dotted graph obtained from the proposed algorithm closely follows the trend of the optimal graph. The difference between the two PSNR values in Figure 10d is just 0.28 dB.

5. Conclusion

The main contribution of this article is a real-time bitrate control algorithm using spatial down-sampling for the low bitrate encoding. The previous resolution

control schemes are too complex to be processed at run time. In this article, a simple estimation model which defines the relationship between the PSNR and the spatial resolution ratio is presented for low bitrate compression. This estimation model is used to find the resolution ratio for acceptable quality on the fly for real-time systems. Two scalability tools for the QP and spatial resolution ratio are determined sequentially to reach the target PSNR. Experimental results show that the proposed bitrate control algorithm is close to the optimal solution and yields the better PSNR and VQM quality at various bitrates compared to the conventional QP-based bitrate control algorithm.

Acknowledgements

This study was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (2011-0027502).

Author details

¹Inter-university Semiconductor Research Center (ISRC), Department of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea ²Department of Electronic Engineering of Sun Moon University, Asan 336-708, Korea

Competing interests

The authors declare that they have no competing interests.

Received: 19 July 2011 Accepted: 23 April 2012 Published: 23 April 2012

References

- ZG Li, F Pan, KP Lim, S Rahardja, Adaptive rate control for H.264. *Proc IEEE Int Conf Image Processing Singapore*. **2**, 745–748 (2004)
- N Kamaci, Y Altunbasak, RM Mersereau, Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models. *IEEE Trans Circ Syst Video Technol.* **15**(8), 994–1006 (2005)
- S Ma, W Gao, Y Lu, Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans Circ Syst Video Technol.* **15**(12), 1533–1544 (2005)
- Z He, Y-K Kim, SK Mitra, Low-delay rate control for DCT video coding via p-domain source modeling. *IEEE Trans Circ Syst Video Technol.* **11**(8), 928–940 (2001). doi:10.1109/76.937431
- W Yuan, S Lin, Y Zhang, W Yuan, H Luo, Optimum bit allocation and rate control for H.264/AVC. *IEEE Trans Circ Syst Video Technol.* **16**(6), 705–715 (2006)
- D-K Kwon, M-Y Shen, C-C Jay Kuo, Rate control for H.264 video with enhanced rate and distortion models. *IEEE Trans Circ Syst Video Technol.* **17**(5), 517–529 (2007)
- C An, TQ Nguyen, Iterative rate-distortion optimization of H.264 with constant bit rate constraint. *IEEE Trans Image Process.* **17**(9), 1605–1615 (2008)
- G Sullivan, T Wiegand, K-P Lim, Joint model reference encoding methods and decoding concealment methods. Section 2.6 rate control JVT-1049 (2003)
- S Ma, W Gao, F Wu, Y Lu, Rate control for JVT video coding scheme with HRD considerations. *Proc IEEE ICIP Spain*. **3**, 793–796 (2003)
- HT Yu, F Pan, ZP Lin, A new bit estimation scheme for H.264 rate control. in *Proc IEEE Int Symp Consumer Electronics*, UK 396–399 (2004)
- X Yang, W Lin, Z Lu, X Lin, S Rahardja, E Ong, S Yao, Rate control for videophone using local perceptual cues. *IEEE Trans Circ Syst Video Technol.* **15**(4), 496–507 (2005)
- Y Liu, ZG Li, YC Soh, Region-of-interest based resource allocation for conversational video communication of H.264/AVC. *IEEE Trans Circ Syst Video Technol.* **18**(1), 134–139 (2008)
- H Li, Z Wang, H Cui, K Tang, An improved ROI-based rate control algorithm for H.264/AVC. in *Proc Int Conf Signal Processing China*. **2**, 16–20 (2006)
- M Hrarti, H Saadane, M Larabi, A Tamtaoui, D Aboutajdine, A macroblock-based perceptually adaptive bit allocation for H264 rate control. in *Proc Int Symposium on IV Communications and Mobile Network*, Morocco 1–4 (2010)
- C-M Huang, C-W Lin, A novel 4-D perceptual quantization modeling for H.264 bit-rate control. *IEEE Trans Multimed.* **9**(6), 1113–1124 (2007)
- Q Meng, Q Meng, Improved macroblock-level rate control algorithm with visual properties. in *Proc Int Workshop Intelligent Systems and Applications*, China 1–5 (2010)
- Z Cui, X Zhu, SSIM-based content adaptive frame skipping for low bit rate H.264 video coding. in *Proc Int Conf Communication Technology*, China 484–487 (2010)
- T-S Ou, Y-H Huang, HH Chen, SSIM-based perceptual rate control for video coding. *IEEE Trans Circ Syst Video Technol.* **21**(5), 682–691 (2011)
- R Jin, J Chen, The coding rate control of consistent perceptual video quality in H.264 ROI. in *Proc Int Symposium Computer Network and Multimedia Technology*, China 1–4 (2009)
- S Wolf, M Pinson, VQM software and measurement techniques. in *National Telecommunications and Information Administration Report* (2002)
- F Pan, X Lin, S Rahardja, KP Lim, ZG Li, DJ Wu, S Wu, Proactive frame-skipping decision scheme for variable frame rate video coding. in *Proc Int Conf Multimedia and Expo Taiwan*. **3**, 1903–1906 (2004)
- F Pan, ZP Lin, X Lin, S Rahardja, W Juwono, F Slamet, Adaptive frame skipping based on spatio-temporal complexity for low bit-rate video coding. *J Vis Commun Image R.* **17**(3), 554–563 (2006). doi:10.1016/j.jvcir.2005.07.006
- J Jun, S Lee, Z He, M Lee, ES Jang, Adaptive key frame selection for efficient video coding. *LNCS*. **4872**, 853–866 (2007)
- S Liu, CJ Kuo, Joint temporal-spatial bit allocation for video coding with dependency. *IEEE Trans Circ Syst Video Technol.* **15**(1), 15–26 (2005)
- A Vetro, Y Wang, HF Sun, Rate-distortion optimized video coding considering frameskip. in *Proc Int Conf Image Processing Greece*. **3**, 534–537 (2001)
- J Kim, Y-G Kim, H Song, T-Y Kuo, YJ Chung, C-CJ Kuo, TCP-friendly Internet video streaming employing variable frame-rate encoding and interpolation. *IEEE Trans Circ Syst Video Technol.* **10**(7), 1164–1177 (2000). doi:10.1109/76.875520
- H Song, C-CJ Kuo, Rate control for low-bit-rate video via variable-encoding frame rates. *IEEE Trans Circ Syst Video Technol.* **11**(4), 512–521 (2001). doi:10.1109/76.915357
- T Thaipanich, P-H Wu, C-CJ Kuo, Low complexity algorithm for robust video frame rate up-conversion (FRUC) technique. *IEEE Trans Consum Electron.* **55**(1), 220–228 (2009)
- AHAM Jackson, R McEwan, J Mullin, Impact of video frame rate on communicative behavior in two and four party groups. *Proc ACM Conf Comput Supported Cooperative Work*, Philadelphia, PA 11–20 (2000)
- Information technology - Coding of Audio-visual Objects - Part 2: Visual. in *International Organization for Standardization* (2000). ISO/IEC 14496-2:1999/Amd.1:2000(E)
- G Cote, B Erol, M Gallant, F Kossentini, H.263+: video coding at low bit rates. *IEEE Trans Circ Syst Video Technol.* **8**(7), 849–866 (1998). doi:10.1109/76.735381
- CA Segall, M Elad, P Milanfar, R Webb, C Fogg, Improved high-definition video by encoding at an intermediate resolution. in *Proc Conf Visual Communications and Image Processing USA*. **5308**, 1007–1018 (2004)
- AM Bruckstein, M Elad, R Kimmel, Down-scaling for better transform compression. *IEEE Trans Image Process.* **12**(9), 1132–1145 (2003). doi:10.1109/TIP.2003.816023
- HA Ilgin, LF Chaparro, Low bit rate video coding using DCT based fast decimation/interpolation and embedded zero tree coding. *IEEE Trans Circ Syst Video Technol.* **17**(7), 833–844 (2007)
- VA Nguyen, YP Tan, WS Lin, Adaptive downsampling/upsampling for better video compression at low bit rate. in *Proc of Int Symposium on Circuits and Systems*, USA 1624–1627 (2008)
- HY Shu, LP Chau, The realization of arbitrary downsizing video transcoding. *IEEE Trans Circ Syst Video Technol.* **16**(4), 540–546 (2006)
- Y-P Tan, Y Liang, H Sun, On the methods and performances of rational downsizing video transcoding. *Signal Process Image Commun.* **19**, 47–65 (2004). doi:10.1016/j.image.2003.08.017
- R-J Wang, M-C Chien, P-C Chang, Adaptive down-sampling video coding. *Proc SPIE Multimedia on Mobile Devices*. **7542**, 1–8 (2010)
- H Lee, Y Lee, J Lee, D Lee, H Shin, Design of a mobile video streaming system using adaptive spatial resolution control. *IEEE Trans Consum Electron.* **55**(3), 1682–1689 (2009)
- R Sirhindi, S Murtaza, M Afzal, Improved data hiding technique for shares in extended visual secret sharing schemes. *LNCS Inf Commun Secur.* **5308**, 376–386 (2008). doi:10.1007/978-3-540-88625-9_25
- S Samuel, WT Penzhorn, Digital watermarking for copyright protection. *Proc Conf AFRICON*. **2**, 953–957 (2004)
- SA Kasmani, A Naghsh-Nilchi, A new robust digital image watermarking technique based on joint DWT-DCT transformation. in *Proc Int Conf on Convergence and Hybrid Information Technology*, Korea 539–544 (2008)
- R Liu, G Wang, P Wang, W Huang, An image authentication scheme based on sliding window. in *Proc Conf Control and Decision*, China 2937–2940 (2008)

doi:10.1186/1687-6180-2012-87

Cite this article as: Rhee et al.: Bitrate control using a heuristic spatial resolution adjustment for a real-time H.264/AVC encoder. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:87.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com