International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.1 (2016), pp.125-134
http://dx.doi.org/10.14257/ijmue.2016.11.1.13

Frame Complexity-Based R-Q Model for Low Bit Rate Video Coding

Sanghyun Park[1]

[1]*Department of Multimedia Engineering, Sunchon National University, Suncheon, Korea*
[1]*shark@scnu.ac.kr*

Abstract

In this paper, we propose an adaptive intra-frame rate-quantization (R-Q) model for H.264/AVC video coding. The proposed model aims at estimating accurate output bit rate according to the quantization parameters (QP) for intra-coded frames in low bit rate networks such as video sensor networks. By taking frame complexity measure into consideration, the parameters of the proposed model can be adaptively adjusted to the visual contents. In addition, the method of Kalman filtering is used to estimate the model parameters for real-time video encoding applications. It is shown by experimental results that the proposed model predicts the output bit rate accurately and thus achieves better mismatch performance than that of the existing method.

***Keywords**: H.264/AVC, R-Q modeling, Intra-frame encoding, Rate control*

1. Introduction

In recent years, low-cost devices such as CMOS cameras that can ubiquitously capture video contents from the various environments have appeared in almost all small wireless mobile devices, such as smartphones and tablet PC's. Furthermore, recent developments in sensor networking have encouraged the use of video sensors in these networks [1]. The demands for monitoring and surveillance systems are also growing with the increasing interest in security in public spaces. These needs have fostered the development of Wireless Video Sensor Network (WVSN) [2-3]. WVSN follows the trends in low-power processing, wireless networking, and distributed sensing. WVSN has also developed as a new technology with a number of potential applications, ranging from security monitoring and mobile multimedia to emergency response [1].

Currently, H.264/AVC standard, which was jointly developed by ITU and MPEG, is most widely used in many video coding applications including WVSN applications. To be applied to a broad range of applications, H.264/AVC video encoding standard has many advanced features. Therefore, the performances of H.264/AVC in terms of quality, bit rate, and complexity are determined by a large number of encoding parameters for the advanced features [4-5].

In a resource constrained environment such as a WVSN, it is very imperative to choose the right configuration and setting parameters that lead to the optimal coding performance [4]. WVSN is usually mission-driven and application specific, so it must operate under a set of unique constraints. The bandwidth of wireless channels is a limited resource in the wireless communications. Therefore, among many requirements, network bandwidth is more critical than others [6].

In video coding, rate control aims to achieve good perceptual quality given the transmission bandwidth constraints. That is, rate control regulates the amount of the coded bits by adjusting quantization parameter (QP) while maximizing the video presentation quality. To achieve this, rate control algorithms employ the rate-quantization (R-Q) model, which represents the number of the coded bits as a function of the QP. In a

ISSN: 1975-0080 IJMUE

WVSN, the exact expectation for the intra-frame encoding is more important than the inter-frame encoding because the network bandwidth is limited and intra-frames consume more bits than inter-frames to achieve identical quality.

Li *et al.* have proposed an adaptive rate control framework for H.264/AVC in JVT-G012 [7]. JVT-G012 can provide an efficient rate control algorithm for inter-frames, but it cannot provide a proper method for intra-frames (I-frames). Usually, the I-frame of a group of picture (GOP) is encoded using the predetermined QP, which is called the initial QP. In many rate control algorithms, the initial QP for an I-frame depends on the average QP of P-frames in the previous GOP as JVT-G012 does. The potential problem is that this scheme cannot estimate the bit rate of an I-frame accurately. It is because the characteristics of the current GOP is not considered when encoding the current I-frame [8]. However, it is quite important to estimate the exact bit rate and control the quality of an I-frame to a suitable level for a fixed target bandwidth in a low bit rate network. Thus, an efficient R-Q model for intra-frame is highly desirable for h.264/AVC that must be adaptive to the frame complexity.

In this paper, we develop an adaptive intra-frame R-Q model for H.264/AVC rate control in low bit rate networks such as WVSN. The output bit rate of an I-frame is usually correlated with its gradient-based frame complexity measure, so the proposed algorithm employs the gradient as a measure of frame complexity. The process of calculating gradient is computationally low-cost and does not need any pre-encoding, so gradient-based methods are proper for real-time applications. By considering the gradient of each I-frame, the proposed algorithm is capable of accurately estimating the bit rates as a function of the initial QP. Experimental results show that the proposed algorithm outperforms the existing method for estimating the output bit rate of I-frames.

The rest of this paper is organized as follows: Section 2 summarizes related studies on frame complexity-based R-Q models. The development of the proposed R-Q model that is adaptive to the scene contents is discussed in Section 3. Section 4 demonstrates the experimental results for the performance comparison. Finally, a conclusion is drawn in Section 5.

2. Frame Complexity-Based R-Q Model

When a frame is encoded using the intra-coding scheme, the output bit rate is affected by the many encoding parameters. In this paper, we focus on the quantization parameter (QP) and the frame complexity. Apparently, the output bit rate decreases as the QP value increases when encoding a frame. With the same QP value, encoding different frames will consume different numbers of bits. It is because the output bit rate of each intra-coded frame is affected by its visual characteristics. Obviously, the output bit rate increases as its visual complexity increases.

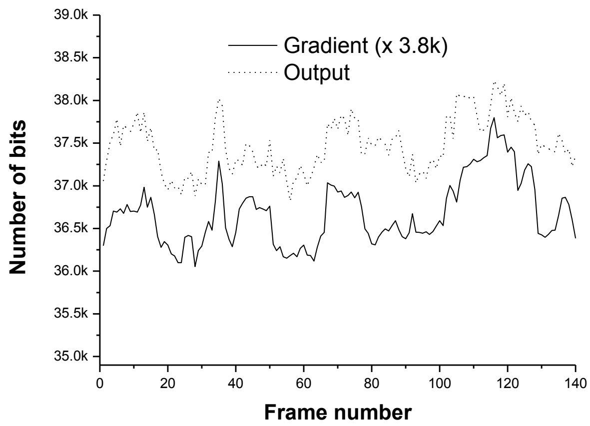


Figure 1. The Gradient per Pixel and the Number of bits after Encoding for different Frames of Akiyo Sequence

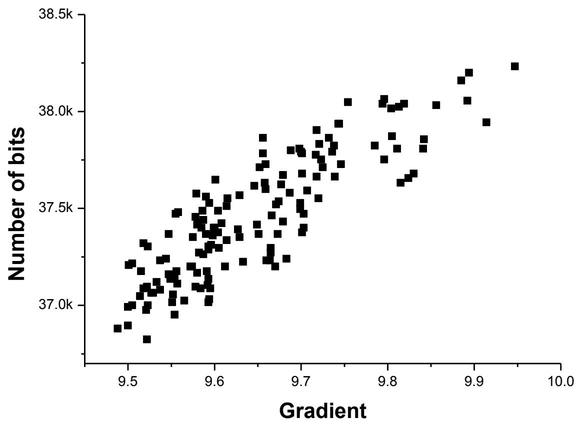


Figure 2. Scatter Plots of the Number of Bits versus the Gradient Value for Akiyo Sequence

There are many activity measures or complexity measures for still image coding. In [9], Kim *et al.* have suggested that the gradient-based method is more reliable for still image coding. The gradient-based method is also proper for the real-time encoding, so we use the gradient of a frame as the complexity measure of the frame. In the proposed R-Q model, the visual complexity of a frame is defined as the average gradient per pixel of the frame. The average gradient of a frame is calculated by

$$G = \frac{1}{M \cdot N}\left\{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(\left|I_{i,j} - I_{i+1,j}\right| + \left|I_{i,j} - I_{i,j+1}\right|\right)\right\} \tag{1}$$

Where M and N are the horizontal and vertical dimensions of the frame, respectively. $I_{i,j}$ denotes the luminance value of the pixel at the location of (i, j) . Figure 1 illustrates the average gradient value and the output bit rate for Akiyo sequence. To compare the characteristics of two curves conveniently, the average gradient values are multiplied by 3800. It is seen that the gradient value can represent the complexity of each frame since the shapes of two curves are quite similar. To further investigate the relationship between the gradient for the visual complexity and the output bit rate, we show a set of scatter plots of bit rate versus gradient value in Figure 2. We can see that there is a linear relationship between these two factors.

The R-Q models using the gradient information have been proposed. In [10], Jing *et al.* have proposed the R-Q model using the linear relationship between the output bit rate and the gradient of a frame. In this model, the output bit rate is formulated as a function of the quantization step size (Q_{step}) instead of the QP. In H.264/AVC, the relationship between QP and Q_{step} is formulated as

$$Q_{step} = 2^{\{(QP-4)/6\}}. \tag{2}$$

The model for the bit rate is formulated as follows:

$$R(Q_{step}) = G \cdot f(Q_{step}) \tag{3}$$

$$f(Q_{step}) = a \cdot (Q_{step})^{b} \tag{4}$$

Where R and G are the output bit rate and the average gradient per pixel calculated from (1), and a and b are model parameters, respectively. The ways to determine the values of the parameters are as follows: The value of parameter b is fixed to a constant value of -0.80 regardless of video contents based on experimental results. The value of parameter a is recursively determined using

$$a_{k+1} = \alpha \cdot a_k + (1-\alpha) \cdot \frac{R_k}{G_k \cdot (Q_{step})_k^{b}} \tag{5}$$

$$a_0 = \frac{R_0}{G_0 \cdot (Q_{step})_0^{b}} \tag{6}$$

Where R_k, G_k, and $(Q_{step})_k$ are actual output bit rate, the average gradient value per pixel, and the Q_{step} for the kth intra-frame, respectively. Experimental results show that this model have achieved significant improvements over the other existing models such as the traditional Cauch density-based models [10].

3. Proposed Intra-Frame R-Q Model

In H.264/AVC video coding, the QP is used for the Lagrangian multiplier method for the Rate Distortion Optimization (RDO). Therefore, we propose the output bit rate model as a function of the QP while Jing *et al.*'s model is as a function of the Q_{step}. Figure 3 gives an example of the relationship between the output bit rate and the QP using the 50th frames of Akiyo and Bridge-Far sequences. As shown in Figure 3, the logarithm of the bit rate decreases in proportion to the QP value. It is also seen in the Figure 1 that the output bit rate of a frame increases in proportion to the average gradient of the frame. To compensate the variation of the output bit rate, we use the bit rate that is divided by the gradient value. Thus, in the proposed scheme, the bit rate model is formulated by

$$\ln(R/G) = c + d \cdot QP \tag{7}$$

Where c and d are model parameters, respectively.

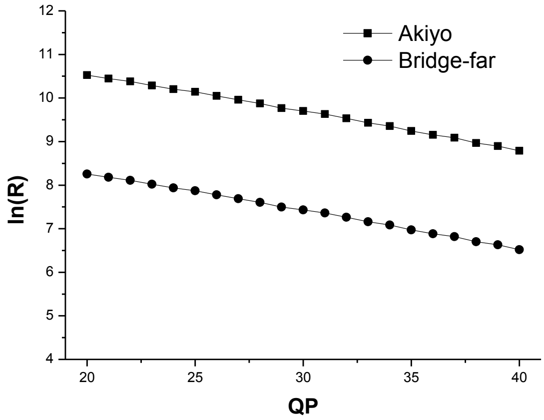


Figure 3. Relationship between the Bit Rate and the QP for the 50th Frames of Akiyo and Bridge-Far Sequences

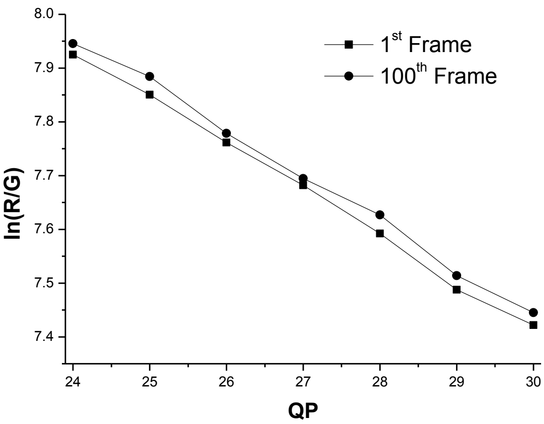


Figure 4. Relationship between the Bit Rate Divided by the Gradient and the QP for the 1st and the 100th Frames of Akiyo Sequence

The frames in a sequence have similar visual contents, so the parameters of each frame vary slightly. Because of this reason, Jing *et al.* have fixed one of the two parameters [10]. Figure 4 shows the R-Q relationships of the 1st frame and the 100th frame of Akiyo sequence. It is seen that two graphs are similar but slightly different. We also investigate the differences between video sequences. Figure 5 shows the R-Q relationships of Akiyo, Bridge-far, and Grandma sequences where the 150 frames of each video sequence are encoded with the randomly selected QP's. As shown in Figure 5, the slopes and the intercepts of graphs are slightly different each other. If these differences are ignored, the accuracy of the estimation becomes somewhat lower. Thus, in the proposed scheme, both of the two parameters (c and d) are recursively estimated. In order to estimate the parameters recursively, we employ the method of Kalman filtering [11]. To apply Kalman filtering, we first define a state vector. In the proposed scheme, the state is defined as a vector of the values of two parameters and then the linear stochastic difference equation of the state is defined as

$$x_k = \begin{pmatrix} c_k \\ d_k \end{pmatrix} \tag{8}$$

$$x_{k+1} = x_k + w_k \tag{9}$$

$$E(w_k \cdot w_j^T) = Q_k \delta_{k-j} \tag{10}$$

Where c_k and d_k are model parameters of the kth frame, and w_k and Q_k are the process noise and the covariance matrix of w_k, respectively. In this system model, the process noise w_k represents the differences of frames in a sequence.

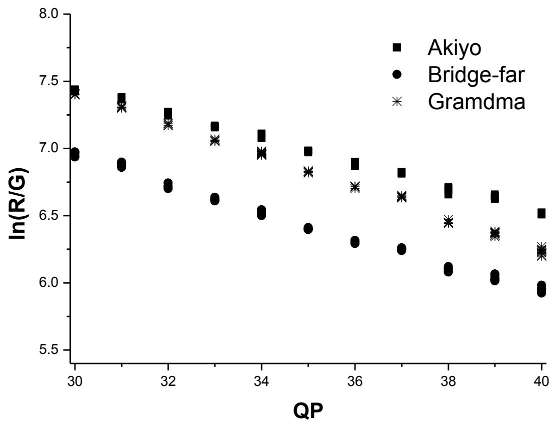


Figure 5. Relationship between the Bit Rate Divided by the Gradient and the QP for Akiyo, Bridge-far, and Grandma Sequences

In WVSN, frames are encoded in real-time, so measurements are made sequentially. In a Kalman filter-based approach, the state of a system is estimated from recursively obtained measurements of some quantity. After encoding, we obtain the output bit rate and the gradient of the frame as measurements, so a noisy measurement vector y_k is defined by

$$y_k = \ln(R_k / G_k) \tag{11}$$

Where R_k and G_k are the output bit rate and the gradient of the kth frame, respectively. With the state x_k and the measurement y_k, the measurement equation is defined as follows:

$$y_k = H_k \cdot x_k + v_k \tag{12}$$

$$H_k = (1 \;\; QP_k) \tag{13}$$

$$E(v_k \cdot v_k^T) = R_k \delta_{k-j} \tag{14}$$

Where QP_k is the QP value for the kth frame, and v_k and Q_k are the observation noise and the covariance matrix of v_k, respectively. In the proposed scheme, it is assumed that both w_k and v_k are white Gaussian noise.

From the previous estimate and the new measurement, the Kalman filter equations recursively determine the system state $\hat{x}$ as follows:

$$P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1} \tag{15}$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{16}$$

$$\hat{x}_k^- = \hat{x}_{k-1}^+ \tag{17}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k (y_k - H_k \hat{x}_k^-) \tag{18}$$

$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \tag{19}$$

Where P is the error covariance matrix, and K is the Kalman gain. The method of Kalman filtering has two important phases. At the time k, to calculate a new estimate of the state $\hat{x}_k$, the time update is done using the previous estimate $\hat{x}_{k-1}$, and then the measurement update is processed with the new measurement y_k. A compelling property of Kalman filtering is that it is capable of producing estimates in real-time.

Comparing the proposed R-Q model with Jing *et al.*'s model, the main difference is the way used to estimate the model parameters. In the proposed scheme, the both model parameters are adaptively determined using the method of Kalman filtering, which is known to be an optimal estimator. In Jing *et al.*'s method, the parameter a is determined by the simple adaptive equation and the parameter b is fixed regardless of video sequences. The parameter b is related to the slope in the proposed model. As shown in Figure 5, each sequence has a unique value of the slope. As a result, the proposed model is more adaptive to the changes of visual characteristics.

4. Experimental Results

Numerous experiments have been conducted to evaluate the performance of the proposed R-Q model, which has been implemented with the latest version of the JVT reference software, JM18.3 using baseline profile. The results achieved by the proposed model are compared with those achieved using Jing *et al.*'s model [10].

The same encoding parameters are used for both algorithms in order to ensure that the comparison is fair. For the experiments, rate-distortion optimization is enabled, and CAVLC is used for entropy coding. An "IPPPP…" GOP structure with a GOP size of 2 is used. That is, one in every two frames is encoded using intra-coding mode. The network bandwidth is assumed to be 80kbps. The simulation was conducted with the first 300 frames of six QCIF test sequences of Akiyo, Bridge-Far, Coastguard, Grandma, Hall, and Salesman.

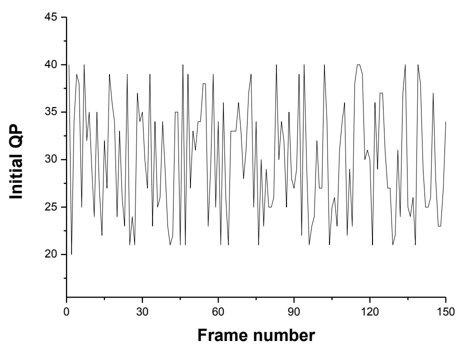


Figure 6. Randomly Selected QP Values for 150 Frames

In our experiments, the first GOP of a sequence is encoded with the initial QP value of 30, which is determined by the way of JVT-G012. The initial QP values of the following GOP's are randomly determined, and the values are between 20 and 40. Figure 6 shows the randomly selected QP values by frames. In order to measure the accuracy of each model, we calculate the average value of the frame bit rate mismatch as

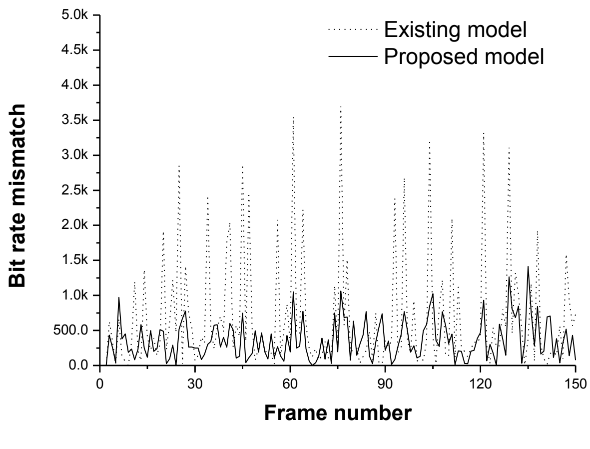
$$M = \frac{1}{N_f} \sum_{i=1}^{N_f} \left| R_{est,i} - R_{act,i} \right| \tag{20}$$

Where N_f is the total number of encoded I-frames and $R_{est,i}$ and $R_{act,i}$ are the estimated bit rate by the R-Q model and the actual bit rate after encoding for the ith frame, respectively.

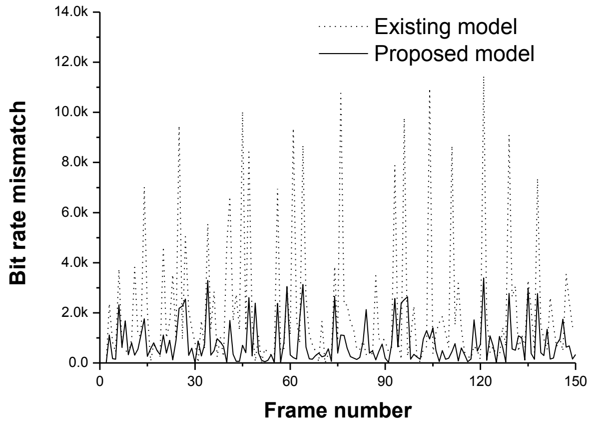
The major issue for the R-Q modeling is the accuracy of the estimated bit rate. Thus, the average mismatch value of each sequence is listed in Table 1 in order to compare the performances of the proposed model and Jing *et al.*'s model. Reduction rate in the Table 1 implies the percentage of the bit rate mismatch of the proposed model over that of Jing *et al.*'s model. The proposed scheme shows better performance than Jing *et al.*'s model in terms of the average value of the bit rate mismatch. On average, the mismatch of our model is about 50.2% of that of Jing *et al.*'s model. In the case of Grandma sequence, we have reduced the estimation error up to 28.5% of that of the existing method.

Table 1. Comparison of The Average Bit Rate Mismatch

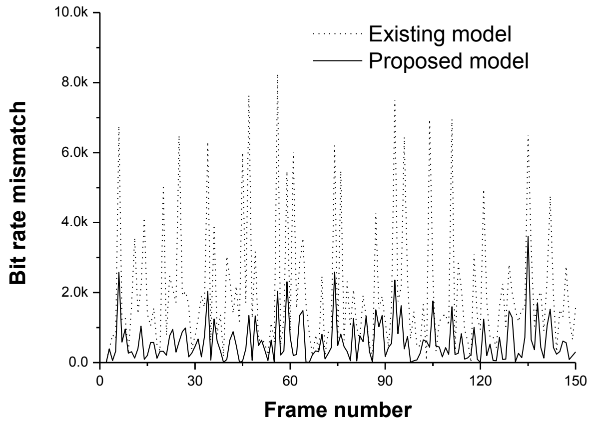
Video Sequence	Average mismatch (bits)		Reduction Rate (%)
	Jing *et al.* [10]	Proposed	
Akiyo	691.1	357.6	51.8
Bridge-far	2218.8	800.6	36.1
Coastguard	2364.9	1967.3	83.2
Grandma	2076.9	592.2	28.5
Hall	909.8	500.1	55.0
Salesman	2387.0	1114.4	46.7



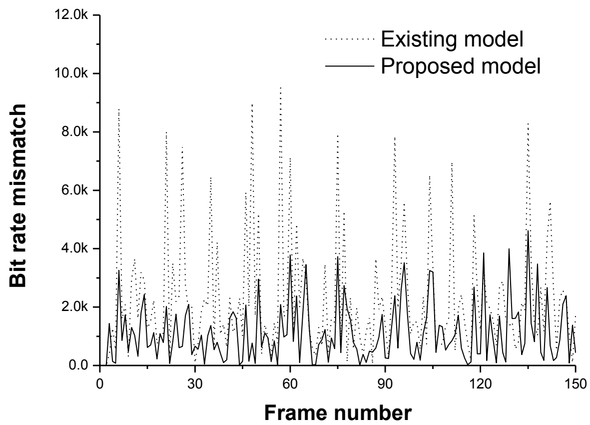
(a) Akiyo



(b) Bridge-Far



(c) Grandmother



(d) Salesman

Figure 7. Comparison of the Bit Rate Mismatch by Frames for Four Sequences

The frame-to-frame mismatch results of four sequences are shown in Figure 7, where it is shown that better results are obtained by the proposed model than Jing *et al.*' model. This is due to the fact that our scheme can efficiently adjust the model parameters based on the frame complexity measures. Comparing with Jing *et al.*'s model, no abrupt fluctuation of bit rate mismatch is observed in our scheme. Therefore, our R-Q model is more robust and efficient for H.264/AVC intra-frame rate estimation.

5. Conclusions

H.264/AVC video coding standard has many advanced features, so it has also been applied to the video sensor networks. In the network where the bandwidth is limited such as video sensor networks, the control for intra-frames is significant. In this paper, an adaptive fame complexity-based bit rate model for H.264/AVC video coding has been proposed. The proposed algorithm takes the characteristics of each video sequence into consideration by using the gradient of each frame. In addition, the proposed model captures the diversities of each frame and each sequence using the method of Kalman filtering, so it can precisely estimate the output bit rate as a function of the QP. The proposed model does not require any pre-encoding to calculate the model parameters. Thus, it is suitable for H.264/AVC video coding in real-time. Experimental results show that the proposed scheme achieves better performance than that of the existing method.

References

[1] X. Liu, X. Gao, D. Zhao, J. Zhou, G. Zhai and W. Gao, "Model-based low bit-rate video coding for resource-deficient wireless visual communication", Neurocomputing, vol. 162, **(2015)**, pp. 180-190.
[2] I. F. Akyildiz, T. Melodia and K. R. Chowdhury, "A survey on wireless multimedia sensor networks", Computer Networks, vol. 51, no. 4, **(2007)**, pp. 921–960.
[3] S. Soro and W. Heinzelman, "A survey of visual sensor networks", Advances in Multimedia, vol. 2009, **(2009)**, pp. 1-21.
[4] B. A. B Sarif, M. T. Pourazad, P. Nasiopoulos and V. C. M Leung, "Encoding and communication energy consumption trade-off in H.264/AVC based video sensor network", Proceedings of 2013 IEEE 14th International Symposium and Workshops on World of Wireless, Mobile and Multimedia Networks, Madrid, Spain, June 4-7, **(2013)**.
[5] I. E. Richardson, "The H.264 Advanced Video Compression Standard", John Wiley & Sons, Ltd, West Sussex, **(2010)**.
[6] B. A. B Sarif, P. Nasiopoulos, V. C. M Leung and M. T. Pourazad, "Analysis of Power Consumption of H.264/AVC-based Video Sensor Networks through Modeling the Encoding Complexity and Bitrate", Proceedings of The 8th International Conference on Digital Society, Barcelona, Spain, March 23, **(2014)**.
[7] Z. Li, F. Pan and K. Pang, "Adaptive basic unit layer rate control for JVT", Proceedings of Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Pattaya, Thailand, March 7-14, **(2003)**.
[8] B. Yan and M. Wang, "Adaptive distortion-based intra-rate estimation for H.264/AVC rate control", IEEE Signal Processing Letters, vol. 16, no. 3, **(2009)**, pp. 145-148.
[9] W. Kim, J. Yi and S. Kim, "A bit allocation method based on picture activity for still image coding", IEEE Transactions on Image Processing, vol. 8, no. 7, **(1999)**, pp. 974-977.
[10] X. Jing, L. Chau and W. Siu, "Frame Complexity-Based Rate-Quantization Model for H.264/AVC Intra-frame Rate Control", IEEE Signal Processing Letters, vol. 15, **(2008)**, pp. 373-376.
[11] D. Simon, "Optimal State Estimation", A John Wiley & Sons, Inc., Hoboken, **(2006)**.

Author

Sanghyun Park, He received the B.S., M.S., and Ph.D. degrees in electronics engineering from Korea University, Seoul, Korea in 1995, 1997, and 2002, respectively. In 2004, he joined the Department of Multimedia Engineering at Sunchon National University, Suncheon, Korea, where he is currently a professor. His research interests include image processing, pattern recognition, and multimedia communication.