

# A Rate Control Algorithm for Low Delay Video Coding

Yao Chunlian<sup>1,2,\*</sup>, Jiang Dong<sup>2</sup>, Guan Xinnan<sup>1</sup>, Wu Hongli<sup>3</sup> and Cao Qian<sup>1</sup>

<sup>1</sup>School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China;

<sup>2</sup>The Key Laboratory of Advanced Information Science and Network Technology of Beijing, Beijing Jiaotong University, Beijing;

<sup>3</sup>Institute of Advanced Information Technology of Beijing, Beijing

**Abstract:** For the standard definition video, the inherent delay produced by the existing coding standards is above 260ms. The delay is too big to apply in real-time video processing system. In order to reduce codec system delay, this paper gives a low delay coding structure. Firstly, the input video is divided into a number of slices for reducing the acquisition delay, and then the slice is coded as a basic unit. Secondly, In order to solve the problem of bit allocation with low delay coding process, this paper presents a constant bit rate control algorithm based on slices; frame and macro block line rate allocation strategy is given. The experimental results show that, compared with the frame/field codec system, new slice coding structure and rate control algorithm can keep the quality of the reconstructed image and effectively reduce the inherent codec system delay. The delay of the proposed algorithm is about 100ms and it can satisfy the real-time requirement.

**Keywords:** Coding, low delay, rate control, slice, video.

## 1. INTRODUCTION

In order to improve the compression performance, ITU and ISO organization has launched a series of video coding standards, such as MPEG-x series standards made by ISO and H.26x series standard made by ITU, and the newest HEVC standard. But from the view of application, these standards are divided into two categories [1-3]: H series is mainly for transmission application, and Mpeg series is mainly for storage video application. Although the formulation and application background are different, these coding standard adopts the same the basic coding framework, which uses frame as the basic coding unit, and DCT+motion compensation as basic inter-frame coding strategy [4-6].

With the 4G network and 3G network has been put into operation, real-time multimedia communication has become the core business of 3G and a new generation of mobile communication network. The real-time video communication has strict requirements on the network bandwidth, delay and packet loss rate. The above code standard can obtain good effect when there is not the real-time requirement. But wireless channel has high bit error rate and the limited available bandwidth, so it faces a great challenge for real-time video communication. To sum up, there are some contradictions: the contradiction between the high computational complexity and the limited computing resources; the contradiction between the huge amount of video data and the limited channel bandwidth; contradiction between error sensitivity of video data and high error rate of wireless channel

[7-10]. In this several prominent contradictions, the first two contradictions in essence is the relationship between a large amount of data and transmission channel. Therefore, for new generation based on mobile communication network, it is necessary to study the low delay coding algorithm.

According to the requirements of mobile multimedia on low bit rate, low resolution video coding, several video standards has mobile multimedia part, such as HEVC, and H.264. The H.264 standard has the optimization and improvement for the embedded platform and gives the international open source software X264, which improves the coding performance from the angles of code optimization, the chip optimization and parallel design. In china, the seventh part of AVS ("digital video and audio codec technology standardization working group") is called AVS-M. But from the test data, the AVS-M encoding time is about 60 times of X264; in terms of coding efficiency, the average PSNR (Peak signal noise ratio) of AVS-M is about 0.9dB lower than that of the X264, or in the same case of PSNR coding bit rate of AVS-M is 19% higher than that of the X264, it need to be further improved [11-15].

## 2. THE DELAY ANALYSIS OF TYPICAL CODING SYSTEM

Through the analysis of a typical codec system, the delay consists of image acquisition, compression, transmission buffer, data link transmission, the receive buffer, decompression, image format conversion and display time. In general, for CBR (Constant bit rate) type channel, the data transfer time is fixed in the coding and decoding process which is needn't be counted in the delay. In the real-time video processing system, receiving and decompression processing is in parallel, so receiving time can't take into account. Therefore,

in the analysis of delay, we can only consider the acquisition, coding, transmission buffer time.

In the coding standards, frame is the basic unit, then the total delay of the encoding and decoding system are analyzed as follows:

The frame is the basic acquisition unit, and acquisition time is calculated as follows:  $T_{capture} = 1000ms / N_{frames}$

In the real-time application, coding time is calculated as follows:

$$T_{coding} = 1000ms / N_{frames}$$

And the compression bit rate is stored in send buffer, the send buffer is used to solve the problem of transmission and variable bit rate streams. IF the buffer is greater, the delay is greater too.

Buffer delay is calculated as follows:  $T_{buffer} = \alpha * 1000ms / N_{frames}$

For different frame types, the size of the buffer is different also. In one GOP (group of pictures), there are three types of pictures. One is I frame, this frame is intra frame and it is a key frame in one GOP. The second frame type is P-frame, and P-frame is an inter-frame, it is coded by prediction method, and its prediction reference frame is the reconstructed image of I or P frame and the reference frame is only be forward frame. The third frame type is B-frame, it is Bi-predictive frames, and coded by prediction method, and its prediction reference frame is the reconstructed image of I or P frame, But the prediction reference frame can be forward frame and backward frame. And the compression bits is different of the three types frame. The output bits of I frame is the most than that of other two types frame. And B frame is the least.

The ratio between the output bits of I-frame and the output bits of P-frame is 8:3. So, usually set the buffer size is 3.5 ~ 4 times of the average frame bit rate.  $\alpha = 3.5 \sim 4$  if (frame = I)

For I frame, the output bit stream may fill 3/4 size of the buffer, so the send buffer delay is generally relatively large.

$$T_{delay} = T_{capture} + T_{coding} + T_{buffer}$$

In the PAL system,  $N_{frames} = 25$ , 25 frames per second,  $T_{delay}$  can be drawn as follows.

$$\begin{aligned} T_{delay} &= T_{capture} + T_{coding} + T_{buffer} \\ &= 40ms + 40ms + (3.5 \sim 4) \times 40ms \\ &= 220ms \sim 240ms \end{aligned}$$

If we do not consider the delay of channel, the maximum delay value is 240ms. 240ms is too great to use in real-time application.

For real-time application,  $T_{coding}$  can't be reduced, so it is very important to reduce the time of  $T_{capture}$ ,  $T_{buffer}$ . In this

paper, we mainly from the acquisition input to reduce the delay, and improve the coding rate control algorithm to reduce the count of bit stream will reduce the space requirements of the buffer, thereby reducing the buffer delay.

### 3. SLICE-BASED ACQUISITION STRUCTURE

In order to reduce the delay of existing coding system, this paper proposed a Slice-based acquisition structure, shown as shown in Fig. (1).

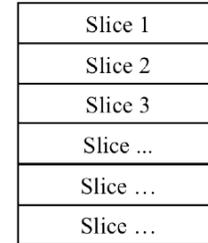


Fig. (1). The slice-based acquisition structure.

In this structure, a whole frame is divided into several slices, after collecting one slice, the slice is directly encoded.

If the image type is I-frame, the coding unit is a slice and not a frame as in traditional coding system. The size of each slice is calculated by the following formula.

$$N_{slice\_MB} = \alpha \times \beta \times N_{MB}$$

$$N_{\_MB\_slice} = N_{frame} / N_{slice\_MB}$$

$$N_{slice\_CU} = \alpha \times \beta \times N_{CU}$$

$$N_{\_CU\_slice} = N_{frame} / N_{slice\_CU}$$

Explanations of parameters are as follows:

For the video standard before HEVC, MB(macro-block) is the main coding unit.

$N_{slice\_MB}$  is the number of rows in one slice

$N_{MB}$  is the number of rows in a macro block, in general,  $N_{MB} = 16$ .

$\alpha$  is the frame resolution parameter, and it's a positive integer, when the resolution is SDTV(Standard definition TV),  $\alpha = 1$ , or the resolution is bigger than SDTV(Standard definition TV),  $\alpha \geq 2$ .

$\beta$  is the system delay requirement parameter, and it's a positive integer. When the system delay is bigger than 200ms,  $\beta \geq 2$ , or When the system delay is smaller than 150ms,  $\beta = 1$ .

$N_{frame}$  is the scanning line number of a frame

$N_{\_MB\_slice}$  is the number of slices in one frame. when the coding unit is macro-block.

For the video standard HEVC, CU(coding unit) is the main coding unit.

$N_{CU}$  is the number of rows in a coding unit, in general,  $N_{CU} = 64 \text{ or } 32$ .

$N_{slice\_CU}$  is the number of rows in a slice.

$N_{\_CU\_slice}$  is the number of slices when the coding unit is CU.

For the P frame or a B frame, P/B frames will have a slice with intra-frame coding method; the rest slices still adopt inter-frame coding method. So there is no great difference between I and P frames, each frame image compression bit rate is relatively smooth, send buffer can be set smaller than that of frame unit coding structure.

Here to analyze the interlaced scanning video, for PAL signal, 25 frames / sec, resolution is 720×576, in one slice, there are only odd lines or even lines.

$$N_{frame} = 576, \alpha = 1, \beta = 1, N_{slice\_MB} = N_{MB},$$

$$N_{\_MB\_slice} = N_{frame} / N_{slice\_MB} = 36$$

For example, Slice-Based system delay is calculated: dividing slice to macro block unit, including only odd or even numbered rows; a frame consists of several slices.

$$\text{the acquisition time is : } T_{capture} = 40ms / 36 = 1.11ms .$$

For slice-based codec, compression bits is relatively smooth, can set the buffer size is that of 1.25 ~ 1.5 frame, send buffer delay is  $T_{buffer} = 50 \sim 60ms$ . For real time coding,

$$T_{coding} = 40ms .$$

The whole system delay including decoding time and display time is

$$T_{delay} = T_{capture} + T_{coding} + T_{buffer}$$

$$= 1.11ms + 40ms + (1.25 \sim 1.5) \times 40ms$$

$$= 91.11ms \sim 101.11ms$$

$T_{delay}$  is below 101.11ms. Compared with frame-based coding method, slice-based reduced by 100 ms .

#### 4. RATE CONTROL ALGORITHM IN LOW LATENCY STRUCTURE

According to the above description, the capture video structure is based on slice-unit and coding unit is slice-unit too. The bit-control algorithm in standard coding is based on frame or macro-block, which is not suitable for slice-based structure.

In CBR (constant bit rate) system with slice-based acquisition structure, the bit allocation strategy in video coding will also make corresponding adjustment.

As described in formula (1), the GOP (a GOP has N frames image) consists of  $N_{gop} \times N_{line}$  macro block lines.

$$\begin{cases} B_{gop} = \sum_{j=0}^{N_{gop}} \sum_{k=0}^{N_{line}} R_{kline,j} \\ \text{Min} \left( \sum_{j=0}^{N_{gop}} \sum_{k=0}^{N_{line}} D_{kline,j} \right) \end{cases} \quad (1)$$

Explanations of parameters are as follows:

$R_{kline,j}$  and  $D_{kline,j}$  is coding bits and distortion of the number K macroblock line in the number J frame,  $B_{gop}$  is the GOP target bits

$N_{gop}$  is the frame number in GOP

$N_{line}$  is the macro block lines in one frame.

#### 4.1. The Bit Allocation Strategy For Macro Block Line

Hwangjun Song [16] proposed *HOD* definition as formula (2):

$$HOD(f_n, f_m) = \frac{\sum_{i \geq |TH|} hod(i)}{N_{pix}} \quad (2)$$

In formula (2),  $i$  is the pixel luminance value

$N_{pix}$  is pixel number in one frame,

$HOD(i)$  is the histogram statistics of the difference image between video frames  $(f_n, f_m)$ ,

$TH$  is threshold used to detect the similarity, it is a positive integer, in the test it is set to 10.

The current target bit rate of the K frame in GOP is

$$T_k = \left( 1 + \mu \times \frac{HOD_k - \mu_{hod}}{\mu_{hod}} \right) \cdot \frac{B_{gop}}{N_{gop}} \quad (3)$$

In formula (3),

$\mu_{hod}$  is the average *HOD* value of GOP

$HOD_k$  is the *HOD* value of the number K frame

$\mu$  is the weight coefficient, usually set to 1.

Obviously, in order to calculate the  $\mu_{hod}$ , we need to analyze the whole GOP. However, it is not to be accepted in the application of low latency.

For real-time application, formula (4) is adopted; the current *HOD* value is instead of the average *HOD* value.

$$T_{kline} = \left( 1 + \mu \times \frac{HOD_k - \frac{1}{k} \sum_{i=1}^k HOD_i}{\frac{1}{k} \sum_{i=1}^k HOD_i} \right) \times \left( \frac{\frac{B_{rem}}{N_{rem}} - R_{used}}{N_{remline}} \right) \quad (4)$$

In formula (4),

$T_{kline}$  and  $HOD_k$  represent the current target rate of the line k macro block and  $HOD$  value.  $B_{rem}$  and  $N_{rem}$  is the remaining bits before coding current frame in the entire GOP and the frame number.  $N_{remline}$  is the remaining macroblock line in the current frame

$R_{used}$  is the bits which has been generated.

Because the Intra slice coding efficiency is relatively low, in order to improve the quality of the code, the target rate of slice is corrected by  $T_{kline} \times \alpha$ , by the experiment test, we set  $\alpha = 3$ .

For the Intra and Inter slices, we use two model to calculate and update, and the model is similar as the standard model.

#### 4.2. The Rate Control Algorithm Description

Combined with the bit allocation method, the improved low delay rate control algorithm is described as follows:

Step1: initialization model. Intra and Inter rate distortion model and the buffer is initialized, the first Intra and Inter macroblock row quantization parameter QP is set to 15.

Step2: macro block line rate allocation

Step2.1: read the current buffer fullness, and calculate  $HOD_k$  value of the current macro block line k.

Step2.2: to determine the current macro block line coding type. If it is the first Intra and Inter macroblock row, jump directly to the Step4.

Step2.3: otherwise the calculate  $T_{kline}$  (the current target rate of the line k macro block)

$$T_{kline} = T_{kline} \times 0.95 + S \times 0.05$$

$S$  is the actual coding bits of the same type of macro block lines.

By the status of buffer, adjust the  $T_{kline}$  value:

$$T_{ksline} = T_{kline} \times \left( \frac{(B + 2 \times (Bs - B))}{(2 \times B + (Bs - B))} \right) B$$

is the size of the buffer

$Bs$  is the actual filling degree of buffer

In order to avoid buffer overflow, adjust the  $T_{kline}$  value as follows:

$$T_{kline} = \begin{cases} \text{Max}(B_{gop}/30, 0.9 \times Bs - B), T_{kline} + B > 0.9 \times Bs \\ Rp + 0.1 \times Bs - B, T_{kline} + B - Rp < 0.1 \times Bs \end{cases}$$

$B_{gop}$  is the GOP target bits

$R_p$  is output bits value when coding the current macroblock row.

Step3: calculate the quantization parameter QP

Step3.1: If  $T_{kline} < 0$ , QP is directly set to  $\text{PreAverageQP} + 2$  and then jump to Step4

Step3.2: otherwise, calculate QP value using the two order model as standard method.

In order to maintain the continuity of video quality, the variation range of QP is limited by the following formula.

$$\text{QP} = \text{MAX}(\text{MIN}(\text{Qp}, \text{PreAverageQP} + 2), \text{PreAverageQP} - 2)$$

$\text{PreAverageQP}$  is the average QP of the previous frame. The formula is only to restrict the QP value of current frame; it shouldn't be differ too much with the previous frame.

Step4: code and update the bit-rate control model, after coding the current macro block, then update the two order Rate-distortion model by the standard rate control algorithm.

If all macro-block rows (namely all the slices) in the current frame have been coded, we first update  $B_{rem}$  and  $N_{rem}$  parameters, and then process the next frame, or jump to Step2.

#### 5. LOW LATENCY SYSTEM PERFORMANCE ANALYSIS

Table 1 gives test results of the low delay coding system, including the average Y PSNR (peak signal noise rate), and the output bits of every frame. In Table 1, Original is the standard test result and the proposed is the test result of the paper.

We choose three CIF sequence, basket is a moving fast video and flower is a colourful video and piano is a relative static video.

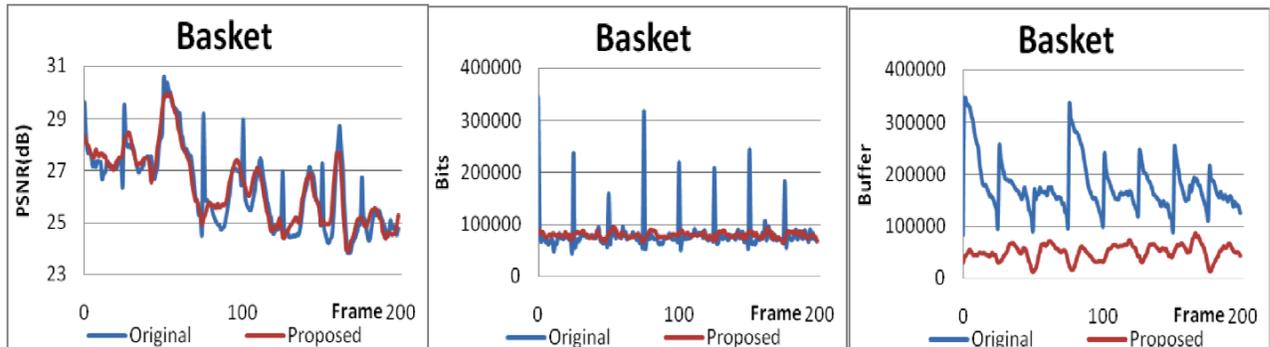
From Table 1, we can know that for basket and piano sequence, the proposed method can get better PSNR value and less bit rate than those of standard, so it can reduce the buffer delay. But for flower sequence, the proposed method get a little more bit than Original, get worse PSNR value, the reason is that flower sequence is more colourful and detail, if we adopt more intra-frame coding method, will generate bitrates.

From Figs. (2a-c), we can get the bit rate, buffer state and PSNR value of per frame in each sequence.

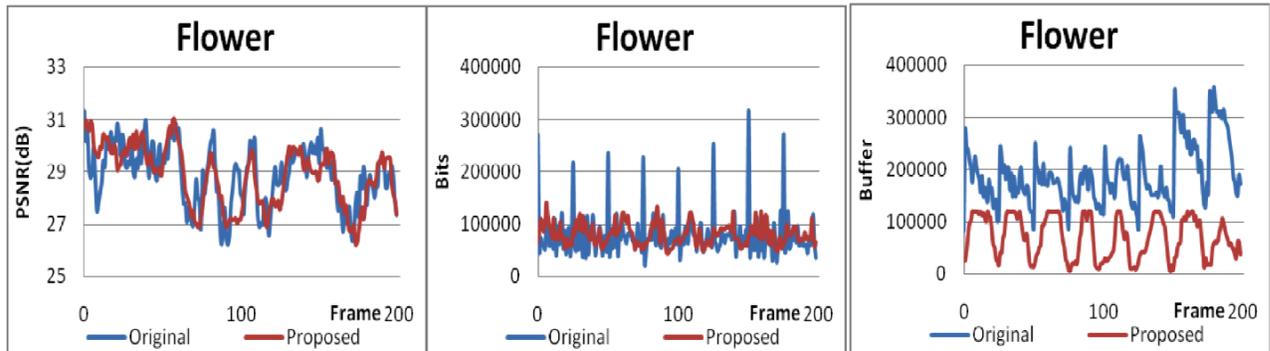
Compared with the two methods, the standard codec keep high buffer fullness, high rate fluctuations and high quality changes, and the low delay coding structure and rate control algorithms has less subjective quality fluctuation, little change of average rate, and get the same PSNR value as the standard method. But the smaller buffer structure and low delay coding rate distortion model is not accurate, the rate control precision declined. The results show that the low latency code achieves a good balance between the delay and image quality.

Table 1. The performance test result.

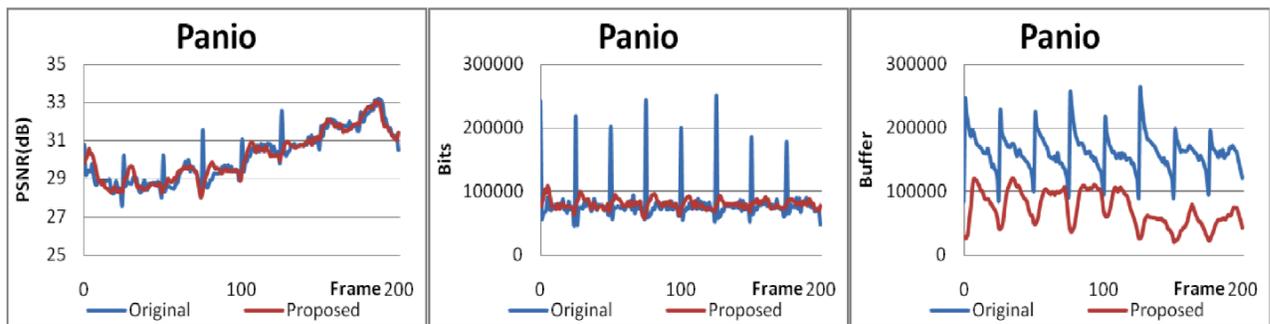
Sequence	Average Y PSNR(dB)		Bit-Rate (Kbps)	
	Original	Proposed	Original	Proposed
Basket	26.349	26.389	2000.480	1993.205
Flower	28.886	28.876	2002.698	2007.888
Piano	30.166	30.158	2000.652	1997.392



(a) The PSNR value, bit rate and buffer status of Basket sequence.



(b) The PSNR value, bit rate and buffer status of Flower sequence.



(c) The PSNR value, bit rate and buffer status of piano sequence.

Fig. (2). The PSNR value, bit rate and buffer status of three sequences.

**CONCLUSION**

In the existing codec framework, analysis showed that the codec has large inherent delay and can't meet the real-time application. So, in the paper propose a low delay coding structure to solve real-time application requirement problem. Analysis shows that, slice-based acquisition structure can effectively reduce the inherent coding system delay. Com-

pared with frame codec, it can reduce about 100ms. In order to match low latency code structure, a new rate control algorithm is proposed, macro block line rate allocation is given, and the algorithm can consider the code bit, buffer fullness etc. The experimental results show that, compared with the frame /field based codec, new structure and rate control algorithm can keep the quality of the reconstructed image and

produce average rate, effectively reduce the inherent codec system delay. It can satisfy some application scenarios low delay requirements.

### CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

### ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (No.61103124) and the Key Laboratory of Advanced Information Science and Network Technology of Beijing (No.XDXX1301).

### REFERENCES

- [1] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol.13, no.7. pp. 560-576, 2003
- [2] Zheng Xiang, Ye Zhiyuan, Zhou Bingfeng. "An overview of the core technology of JVT draft," Journal of software. 15(1), pp.58-68, 2004
- [3] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.
- [4] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC)", IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1669-1684, Dec. 2012.
- [5] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC Complexity and Implementation Analysis," IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, December 2012.
- [6] Sadka, A.H., "compressed video communications". 2002, England: John Wiley&Sons.
- [7] Yu-Han Chen, Tung-Chien Chen, Chuan-Yung Tsai, Sung-Fang Tsai, and Liang-Gee Chen, "Algorithm and Architecture Design of Power-Oriented H.264/AVC Baseline Profile Encoder for Portable Devices," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no.8, pp. 1118-1128, Aug. 2009.
- [8] ZhiYang,Hua Cai and Jiang Li, "A framework for fine-granular computational- complexity scalable motion estimation," IEEE International symposium on Circuits and Systems (IS-CAS), pp.5473-5476, 2005,
- [9] Parkvall S., Englund E., Lundevall, M.,etc."Evolving 3G mobile systems: broadband and broadcast services in WCDMA," IEEE Communications Magazine ,2006, 44(2), pp.68-74.
- [10] Yu, K.M., *et al.*, "practical video codec for mobile devices," IEEE International Conference on Multimedia & Expo, pp.509-512, 2003,
- [11] Etoh M.,Yoshimura T., "Wireless video applications in 3G and beyond," IEEE Wireless Communications, 2005,12(4), pp.66-72.
- [12] Huang Tiejun, Gao Wen "AVS draft", "2005.10
- [13] Bruno Macchiavello, Debargha Mukherjee, and Ricardo L. de Queiroz, "Iterative Side-Information Generation in a Mixed Resolution Wyner-Ziv Framework," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no.10, pp. 1409-1423, Oct. 2009.
- [14] Thung-Hiung Tsai, Jin-Jang Leou. A Rate Control Scheme for H.264 Video Transmission. ICME. 2004,1(8):112-116
- [15] Nejat Kamaci, Yucel Altunbasak, Russell Mersereau. Frame Bit Allocation for the H.264/AVC Video Coder Via Cauchy-Density-Based Rate and Distortion Models[C].In: IEEE Transactions On Circuits And Systems For Video Technology, 2005, 15:994-1006
- [16] Hwangjun Song, C.C Jay Kuo. A Region-Based H.263+ Codec and Its Rate Control for Low VBR Video, IEEE Trans on Multimedia, 2004, 6(3):489-500.

Received: September 16, 2014

Revised: December 23, 2014

Accepted: December 31, 2014

© Chunlian *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.