Adaptive Down-Sampling Video Coding

Ren-Jie Wang[a], Ming-Chen Chien[ab] and Pao-Chi Chang*[a]
[a]Dept. of Communication Engineering, National Central Univ., Jhongli, Taiwan;
[b]Dept. of Electrical Engineering, Chin Min Institute of Technology, Miaoli, Taiwan

ABSTRACT

Down-sampling coding, which sub-samples the image and encodes the smaller sized images, is one of the solutions to raise the image quality at insufficiently high rates. In this work, we propose an Adaptive Down-Sampling (ADS) coding for H.264/AVC. The overall system distortion can be analyzed as the sum of the down-sampling distortion and the coding distortion. The down-sampling distortion is mainly the loss of the high frequency components that is highly dependent of the spatial difference. The coding distortion can be derived from the classical Rate-Distortion theory. For a given rate and a video sequence, the optimum down-sampling resolution-ratio can be derived by utilizing the optimum theory toward minimizing the system distortion based on the models of the two distortions. This optimal resolution-ratio is used in both down-sampling and up-sampling processes in ADS coding scheme. As a result, the rate-distortion performance of ADS coding is always higher than the fixed ratio coding or H.264/AVC by 2 to 4 dB at low to medium rates.

Keywords: Down-sampling, Video Coding, H.264, High-Definition Video

1. INTRODUCTION

Multimedia applications on various devices and heterogeneous networks become popular. The demand for High-Definition (HD) video is also increasing. HD video usually delivers much higher bitrates than traditional video. However, the available rates provided by the networks, especially wireless networks, are not always adequate for high quality video. In such a condition, down-sampling coding, which down-samples the image and encodes the smaller sized images, is one of the solutions to raise the image quality. In general, it yields better performance than the original full-size video coding at low to medium rates.

Literatures on down-sampling coding focus on the discussions for down-sampling or up-sampling procedures. In the procedure of up-sampling, super-resolution technique can be applied to the decoder of down-sampling coding to raise the rate-distortion performance [1]. However, the computational complexity is very high. Fast algorithms of super-resolution also exist [2], but the degraded R-D performance significantly reduces the advantage of super-resolution. To improve the coding performance, video frames can be classified into different types based on different visual characteristics, and then be processed with conditional super-resolution [3]. However, the coding efficiency improvement by adjusting up-sampling procedure is limited, and the improvement would be varied because of different video contents. We shall focus on the down-sampling procedure for better and stable performance. It is observed that the optimum down-sampling ratio depends on the bitrates [4]. Namely, the superior performance over the traditional coding only exists in a range of bitrates. A variable down-sampling ratio may extend the superior bitrates range. For still images, adaptive down-sampling ratio has been applied to JPEG image compression [5].

Sampling-rate selection schemes for video coding were proposed in recent years. An adaptive down-sampling mode decision in the encoder was proposed [6]. The modes including different directions and sizes can be determined by the features of block contents. This method also provides better RD performance than regular video coding. However, there exists blocking effect because of different distortion features in different blocks. Moreover, modifying coding loop loses the syntax conformation for video coding standard. Another related discussion for sampling-rate selection is on the resolution transcoding with up-sampling procedure [7]. The method estimates the bitrates for different resolution versions and then selects the largest resolution satisfying the bitrates constraint to encode. However, this strategy may not reach the minimum distortion and only five resolution versions can be used.

*pcchang@vaplab.ce.ncu.edu.tw; phone 886 3 4227151; fax 886 3 4229187; vaplab.ce.ncu.edu.tw

Multimedia on Mobile Devices 2010, edited by Reiner Creutzburg, David Akopian, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 7542, 75420P · © 2010 SPIE-IS&T · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.840257

In this work, we propose an Adaptive Down-Sampling (ADS) coding for H.264/AVC video coding. Furthermore, the optimum down-sampling resolution-ratio is derived based on the models of down-sampling distortion and coding distortion. The architecture and distortion analysis are discussed in Section 2. The derivation of optimum resolution-ratio is also shown in this section. Section 3 provides simulation results. Finally, Section 4 provides the conclusions.

2. ADAPTIVE DOWN-SAMPLING CODING

2.1 Architecture of adaptive down-sampling coding

The architecture of the proposed ADS coding is shown in Figure 1. The original frame F_{high} is first down-sampled as DF_{low}. It is then encoded as conventional video coding with the bitrates R. At the decoder side, the video sequence is decoded as RDF_{low}. It is then up-sampled to the original resolution as $URDF_{high}$. The resolution-ratio a is defined as the area ratio of the original frame area to the down-sampled frame area. For instance, in the case of down-sampling from CIF to QCIF, a is set to 0.25. Figure 2 shows the performance comparison between down-sampling scheme (circle) and original H.264 (triangle). There exists a cross-over rate that the down-sampling scheme outperforms the original H.264 under the rate, but is worse if the rate is higher than the cross-over rate. We also observe that a small a results in high gain in PSNR but low cross-over rate. Therefore, it is extremely important to find a suitable a for a given rate.

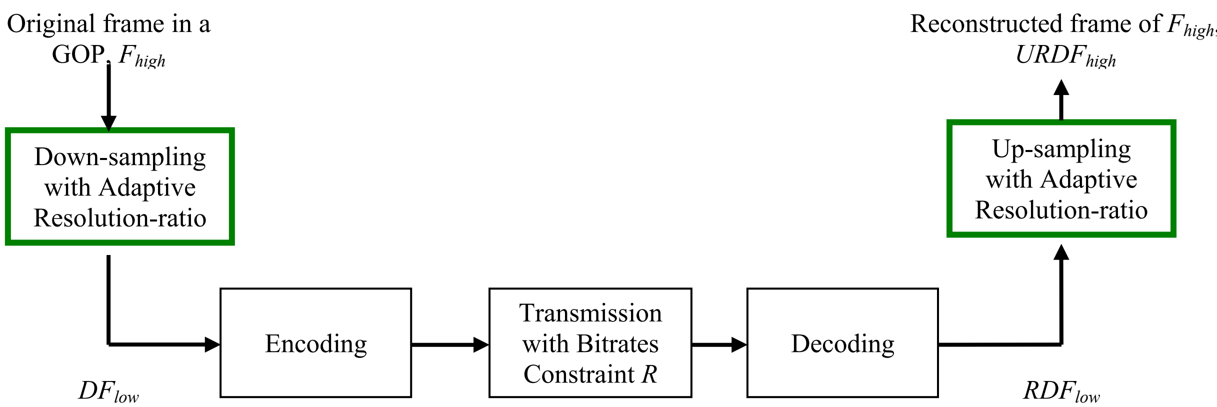

Figure 1. Down-sampling coding scheme with adaptive resolution-ratio.

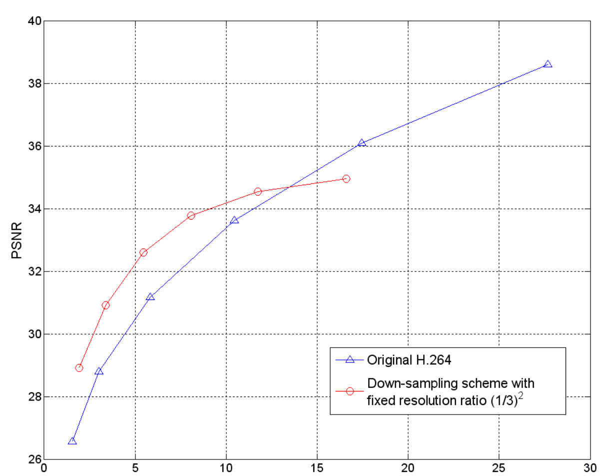

Figure 2. Performance comparison between down-sampling scheme and original H.264, Sequence Riverbed.

2.2 Optimization of resolution-ratio

In this system, two types of distortions are involved in the down-sampling coding scheme. Both are related to resolution-ratio under bitrate constraint. The first distortion is created from down-sampling, denoted as D_{down}. It is inverse proportional to the resolution-ratio and is highly dependent of characteristics of video sequences. The second distortion is generated from the quantization of the encoder, denoted as D_{coding}. It is proportional to the resolution-ratio and also depends on bitrates and characteristics of video sequences. Figure 3 shows the coding distortion at various rates as well as down-sampling distortion versus the resolution-ratio. It clearly shows that the two distortions have opposite trends as the resolution increases. Thus it is interesting to study the optimum resolution-ratio that minimizes the overall distortion.

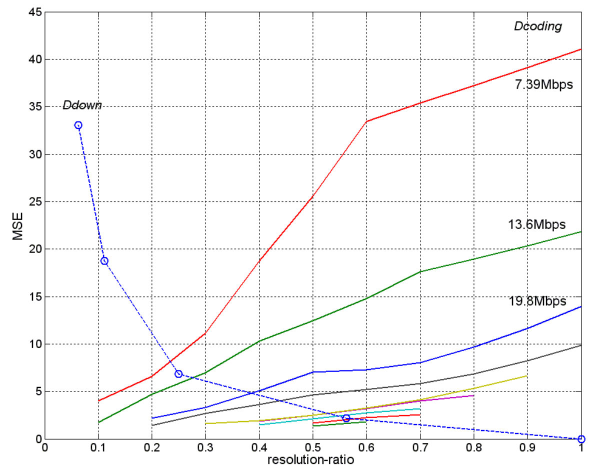

Figure 3. Coding distortion at various rates and down-sampling distortion versus the resolution-ratio (Riverbed).

Since the coding distortion is mainly quantization noise which has different cause from the down-sampling distortion that is mainly the loss of high frequency components, we propose an additive noise model as

$$D(R,a,seq) = D_{coding}(R,a,seq) + D_{down}(a,seq) \quad (1)$$

where D represents the average distortion of each pixel of video sequence for a given video sequence "*seq*". Figure 4 shows the experimental results of comparisons between the total distortion and the additive distortion model. They are very close at medium to high rates, different at low rate but the trend (i.e., location of the minimum MSE, Mean Square Error) is still similar.

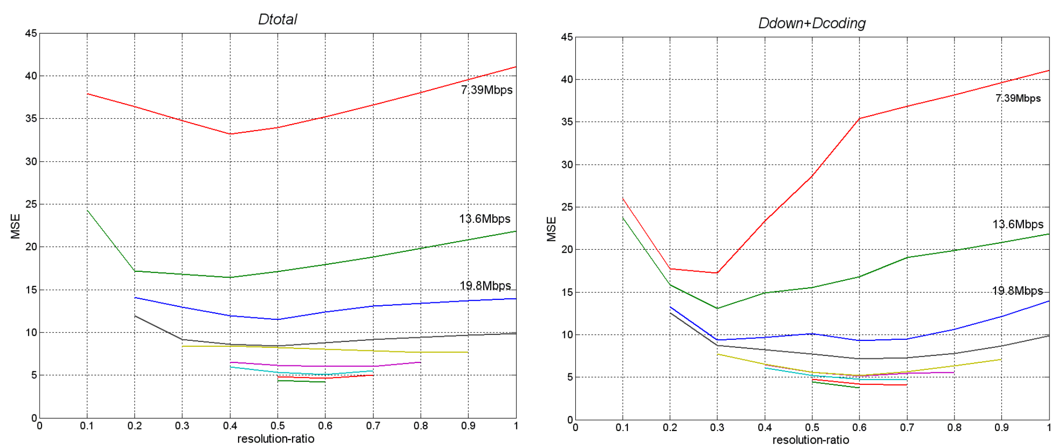

Figure 4. Comparisons between the total distortion and additive distortion model (Riverbed)

Because the down-sampling distortion D_{down} is mainly the loss of the high frequency components that is highly dependent of the spatial difference, a reasonable model for down-sampling distortion can be described as

$$D_{down} = \sigma_s^{\ 2}(\frac{1}{a} - 1) \tag{2}$$

where $\sigma_s^{\ 2}$ is the variance in pixel domain, i.e., spatial difference, of a frame in video sequences. Figure 5 shows the MSE verses the resolution-ratio a for various video sequences. Although the MSE of different sequences varies significantly, the distortion model, shown as the dash line, is actually very accurate.

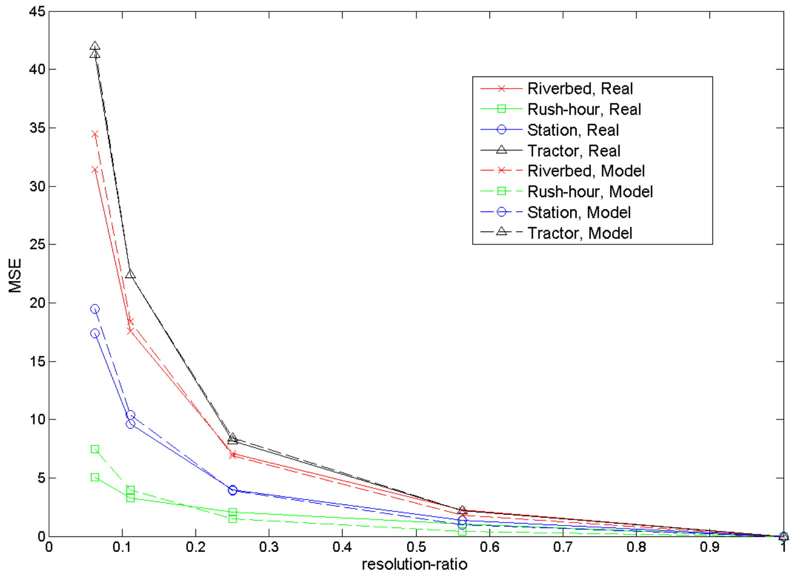

Figure 5. Down-sampling distortion modeling

The coding distortion can be derived from the classical Rate-Distortion (RD) theory [8]. The R-D model, which includes the transform, quantization, inverse quantization, and inverse transform, for an MB has been proposed [9] as

$$D_{ref} = \sigma_t^{\ 2} \cdot 2^{-\gamma R} \tag{3}$$

where $\sigma_t^{\ 2}$ is the temporal difference of frames in video sequence and γR is the bits used to represent a pixel on average.

In this work, the proposed resolution-rate-distortion model that takes resolution-ratio into consideration can be described based on (3) as

$$D_{coding} = \sigma_t^{\ 2} \cdot 2^{-\gamma R / a} \tag{4}$$

The coding distortion increases with a. For small a, the number of effective pixels per frame after down-sampling decreases, and the average bitrates allocated to each pixel increases. Thus the coding distortion may decrease because of the relatively high bitrates/pixel. Figure 6 shows the MSE versus resolution-ratio and the distortion model at high rates and low rates, respectively. In either case, the mismatch between the actual MSE and the model is very limited.

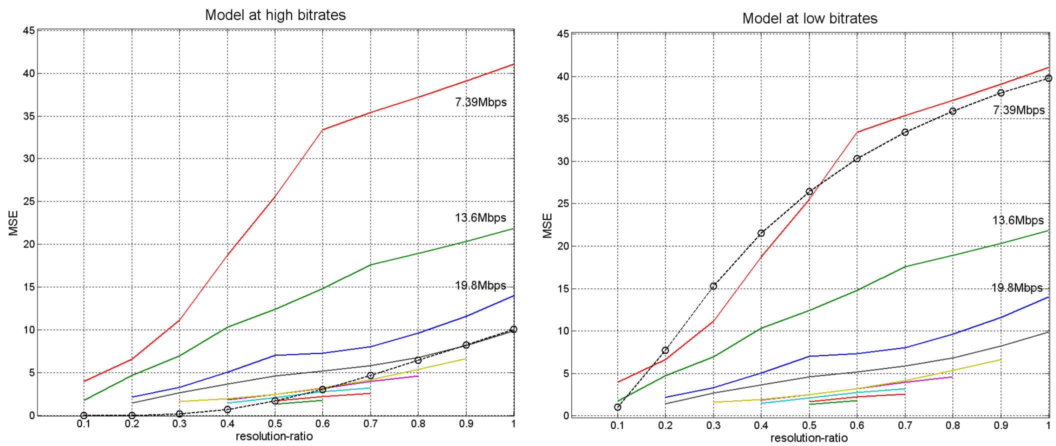

Figure 6. Coding Distortion Model at Low and High bitrates (Riverbed)

The overall distortion can be shown in (5) by combining (2) and (4) as

$$D = \sigma_t^2 \cdot 2^{-\gamma R/a} + \sigma_s^2 (\frac{1}{a} - 1) \tag{5}$$

For a given rate and a video sequence, the resolution-ratio can be adjusted to get the optimum value by letting the first order differentiation be zero

$$\frac{dD}{da} = \sigma_t^2 \cdot 2^{-\gamma R/a} \cdot \ln 2 \cdot -\gamma R \cdot \frac{-1}{a^2} + \sigma_s^2 \cdot \frac{-1}{a^2} = 0 \tag{6}$$

and the optimum *a* can be describe as

$$a_{opt} = \frac{-\gamma R}{\log_2(\frac{\sigma_s^2}{\sigma_t^2} \cdot \frac{1}{\ln 2} \cdot \frac{1}{\gamma R})} \tag{7}$$

Although (7) looks complicated, the relationship between the optimum resolution-ratio and the bitrates is actually nearly linear, as shown in Figure 7. This optimal resolution-ratio a_{opt} is used in both down-sampling and up-sampling processes in ADS coding scheme.

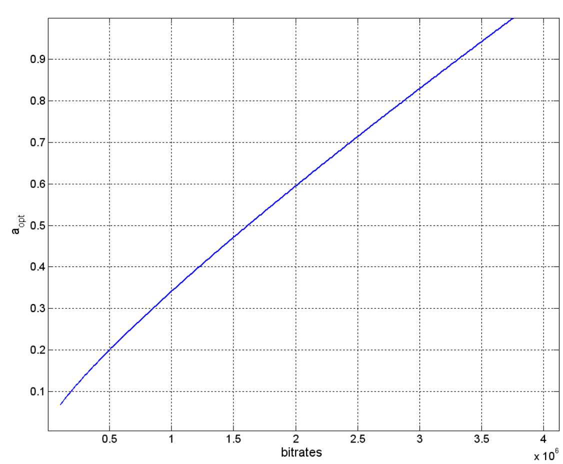

Figure 7. Optimum resolution-ratio at various rates

3. SIMULATIONS

We implemented the proposed algorithm on JM12.4 H.264/AVC codec and assessed the performance on sample HD video sequences. The first 25 frames of the test sequences including “Riverbed”, “Station”, and “Rush-hour” at 25 fps frame rate are used in the experiments as shown in Figure 8. Each sequence is coded with the structure IPPP and GOP size 25. Both the RDO and fast motion search algorithm UMHexagons are enabled as listed in Table 1. The bi-cubic interpolation is chosen for both down-sampling and up-sampling filters.

The PSNR comparison of the down-sampling schemes and the original H.264 is presented in Figure 9. It is shown that the down-sampling video coding with fixed resolution-ratio 0.25 outperforms the regular H.264 video coding at the bitrates only before the cross-over rate 28Mbps. However, the fixed ratio down-sampling scheme performs poorly at high bitrates. On the other hand, the proposed algorithm that utilizes the optimum resolution-ratio performs better than the original H.264 as well as the fixed ratio down-sampling scheme at all bitrates. Moreover, the proposed algorithm achieves up to 2dB gain in PSNR over regular H.264 coding at low to medium rates. Among the three test sequences, “Riverbed” processed by two methods are presented in Figure 10. The upper-right corners are shown for easy comparisons of the details. As the figure shows, our proposed scheme is able to improve the visual quality significantly by eliminating the blocking effect that is crucial to HD video.

Table 1 Encoding parameter list in JM12.4

Profile	Baseline
Resolution	HD(1920x1056)
Test sequence	Riverbed, Station, Rush-hour
FramesToBeEncoded	first 25 frames
FrameRate	25 frame per second
RDO	On（high-complexity）
Fast motion estimation	UMHexagons
Number of reference frame	1 frame
Search range for HD	±128
Block mode	all mode
Subpixel Motion Estimation	On
GOP	25(IPPP~~)

Riverbed	Station	Rush-hour

Figure 8. Three types of different test sequences

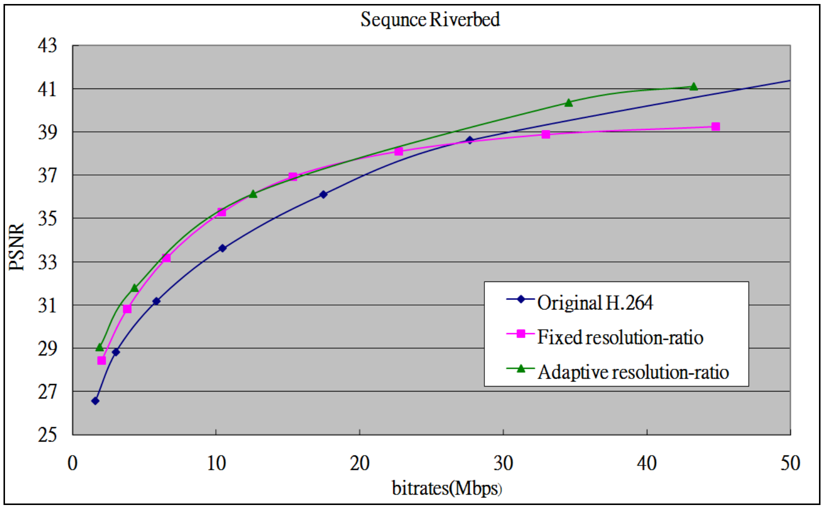

Figure 9. Rate-distortion performance of coding scheme with adaptive resolution-ratio

Figure 10. Subjective comparison of upper-right corner of Riverbed at 1.7Mbps

4. CONCLUTIONS

We have proposed a down-sampling coding scheme with adaptive resolution-ratio that performs efficiently for high-definition video. For a given rate and a video sequence, the optimal resolution-ratio can be determined toward minimizing the system distortion. This optimal resolution-ratio is used in both down-sampling and up-sampling processes in ADS coding scheme. Compared with fixed resolution-ratio, the proposed scheme has better RD performance at low as well as high bitrates. Compared with original H.264 (without down-sampling process), the proposed scheme has 2dB~4dB gain in PSNR at medium and low bitrates. This system is suitable for heterogeneous networks or variable bitrate environments.

REFERENCES

[1] R. Molina, A.K. Katsaggelos, L.D. Alvarez, and J. Mateos, "Towards a new video compression scheme using super-resolution," Proc. SPIE-IS&T Electronic Imaging, Visual Communications and Image Processing 6077, 0601-0613 (2006).

[2] G. M. Callicó, R. P. Llopis, S. López, J. F. López, A. Núñez, R. Sethuraman, R. Sarmiento, "Low-cost super-resolution algorithms implementation over a HW/SW video compression platform," EURASIP Journal on Applied Signal Processing archive, Papers 2006, 237 (2006).

[3] D. Barreto, L.D. Alvarez, R. Molina, A. K. Katsagelos, G.M. Callico, "Region-based super-resolution for compression," Springer Science +Business media on Multidim syst sign process, Papers 18, 59-81 (2007).

[4] C. A. Segall, M. Elad, P. Milanfar, R. Webb and C. Fogg, "Improved High-Definition Video by Encoding at an Intermediate Resolution," Proc. of the SPIE Conference on Visual Communications and Image Processing 5308, 1007-1018 (2004).

[5] A. Bruckstein, M. Elad and R. Kimmel, "Down Scaling for Better Transform Compression," IEEE Trans. On Image Processing, Papers 12(9), 1132-44 (2003).

[6] V. A. Nguyen, Y. P. Tan, and W.S. Lin, "Adaptive Downsampling/Upsampling for Better Video Compression at Low Bit Rate," in Proc. IEEE International Symposium on Circuits and Systems (ISCAS), 1624-1627 (2008).

[7] H.Y. Shu and L.P. Chau, "The Realization of Arbitrary Downsizing Video Transcoding," IEEE Transactions on Circuits and Systems for Video Technology, Papers 16(4), 540- 546 (2006).

[8] T. Berger, [Rate Distortion Theory], Englewood Cliffs, NJ: Prentice-Hall, (1984).

[9] Z. H. He, Y. F. Liang, L. L. Chen, I. Ahmad, and D. P. Wu, "Power-Rate-Distortion analysis for wireless video communication under energy constraints," IEEE Trans. Circuits Syst. Video Technol., Papers 15(5), 645-658 (2005).