Enlarged Block Sizes and Motion Search Ranges for High Definition Video Coding

Yunyang Dai, Qi Zhang, Siwei Ma and C.-C. Jay Kuo

Ming Hsieh Department of Electrical Engineering and Signal and Image Processing Institute
University of Southern California, Los Angeles, CA 90089-2564

***Abstract*—A rate-distortion (R-D) model is derived to investigate the potential coding gain from the use of a larger macroblock size for high definition (HD) video coding in this work. Coding rates of using different block sizes for HD video coding are compared with the R-D model, and limitations of the current coding techniques that might impede the additional R-D gain of using enlarged MB for HD coding are explained. Experiments show that an average bit rate gain of 27 % across QP range can be achieved by using enlarged MB and extended motion search range.**

Keywords —Digital cinema, high definition video, macroblock size, rate distortion model, search range

I. Introduction

Digital cinema has revolutionized the movie industry due to its superior viewing quality and cost reduction advantages as compared with traditional film prints. According to Motion Picture Association's statistics, the number of worldwide digital cinema screens has been increased by 27 times from 1999 to 2005 [1]. Digital Cinema Initiative (DCI) was established in 2002 by major studios and filming corporations to ensure an open system architecture for digital cinema. The use of 4k/2k cameras specified in DCI [2] further reflects the increasing market demand for richer contents with increased fidelity and overall more involving experience.

However, one of the main challenges with these high definition (HD) video contents is the storage and bandwidth requirements for streaming over IP networks. To support the rising demand for higher resolution with better visual quality, H.264 incorporated the Fidelity Range Extension (FR-Ext) in its high profile to support 4k/2k contents [3]. Furthermore, the latest release of HD-DVD and Blue-ray disc both supports a much higher resolution (1080p). In this work, we will refer to any video contents with resolution lower than 1080p as SD contents and others HD contents.

Recent studies report that current coding schemes may not be efficient in the coding of HD video contents since the macroblock (MB) size used in H.264's motion compensated predictive coding does not scale well to the increased object size in HD contents. Hence, different research efforts have been conducted on the HD content compression with increased MB sizes. There have been results on the benefit of allowing MB size larger than 16x16. For example, extended MB modes such as 64x64 and 32x32 MB sizes were incorporated to H.264 high profile in [7]. However, results shown that the percentage of blocks with SKIP and direct mode decreases as the MB size increases. The overall rate-distortion (R-D) gain diminished as the bit rate increases. Although different heuristic approaches have suggested in the past that a larger MB size could bring more bit savings, most of the results reported so far do not show a significant performance gain by adopting a larger MB size.

We attempt to understand this issue from a theoretical viewpoint in this paper. We begin with an analytical R-D estimation for a motion compensated (MC) coder derived from high bit rate coding theory. Coding rates of using different block sizes for HD coding are compared by applying the R-D model and the potential benefit of larger block sizes used for coding flat areas is shown in Sec. II. Limitations of the current coding techniques that might impede additional R-D gain from using larger MB sizes for HD contents are explained in Sec. III, which is attributed to a limited motion search range. Based on experimental results in Sec. IV, we observe that a significant R-D gain for HD video can be obtained from increased MB sizes as well as an enlarged motion search range. Finally, concluding remarks are given in Sec. V.

II. Enlarged Block Sizes for HD Video Coding

For HD content encoding, the major emphasis is on high fidelity and the bit rate is in the range of around 10Mbps or above with a relatively small quantization step size. Hence, we will focus on the high bit rate coding scenario from now on. In this section, we examine a theoretical framework using entropy-based R-D estimation. The derived analytical models suggest that, as the content resolution scales up, a larger MB size should introduce an extra coding gain.

A. R-D Characteristics for HD Video

For a motion compensated coder, the displacement difference x can be modeled as a zero-mean Laplacian distribution. Its probability density function (PDF) is given by

$$p(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}\,|x|}{\sigma} \right). \tag{1}$$

Consider a high bit rate quantizer Q with quantization stepsize Δ. Shannon's theorem states that its entropy is a lower bound of average bits needed to encode each symbol in x as

$$H(\Delta) = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx = \log_2\left(\frac{\sqrt{2}e\sigma}{\Delta}\right). \quad (2)$$

Under the high bit rate assumption, the bit rate and distortion needed to encode the displacement difference R_d can be approximated [4], respectively, by

$$R_d(\Delta) = H(\Delta) - \frac{1}{2}\log_2(12D), \quad (3)$$

$$D(\Delta) = \frac{\Delta^2}{12}. \quad (4)$$

By substituting (4) into (3), we obtain the rate model as

$$R_d = H(\Delta) - \log_2\Delta = \log_2\left(\frac{\sqrt{2}\sigma e}{\Delta^2}\right). \quad (5)$$

By adding the header information requirement for each MB, we can derive a complete rate model as

$$R_{total} = R_{hdr} + R_d = R_{hdr} + \log_2\left(\frac{\sqrt{2}\sigma e}{\Delta^2}\right), \quad (6)$$

where R_{hdr} includes the information of quantization parameters, motion vectors, etc. It can be obtained by searching the lookup table (LUT) similar to the one used in H.264. In the next subsection, we will use the above R-D model to analyze the benefit of a larger MB size in HD content.

B. Impact of Increased Block size on HD Video Coding

We proceed to analyze the rate produced by different block modes to understand the impact of larger block sizes in higher resolution. We use blocks of size $N \times N$ (*N>16*) in HD content as the $N \times N$ mode and the same block decomposed into 4 smaller subblocks of equal size as the $N/2 \times N/2$ mode. Intuitively, for a flat block with a given Δ, the $N \times N$ mode will have a smaller entropy value than that from each subblock in the $N/2 \times N/2$ mode. For fair comparison, we will use smaller Δ; namely,

$$\Delta_{N\times N} = \gamma \cdot \Delta_{N/2\times N/2} ,$$

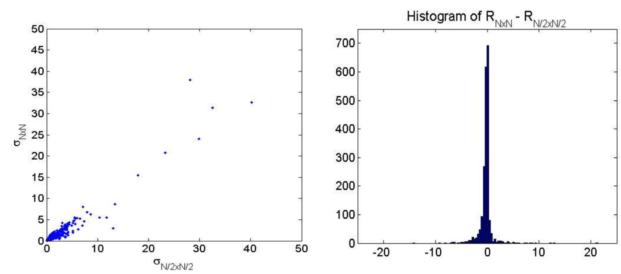
where $\gamma \in [0.8,1)$ for *N=32,* in the $N \times N$ block case to compensate for the reduced entropy. Empirical studies show that, after MC, a mode of larger block sizes in HD tends to produce a relatively smaller variance value as compared to the segmented mode in flat regions as shown in Fig.1, where the x-axis and y-axis provide variance values of the $N \times N$ and $N/2 \times N/2$ modes, respectively. The relationship can be mathematically expressed as:

$$\sigma_{(u,v)N\times N} = k \cdot \sigma_{(u,v),N/2\times N/2}, \quad (7)$$

with $k \in [0.2,0.5]$ for flat surfaces when N=32. If the header information is similar, by comparing only the bit rates calculated by (3), the bit rate difference

$$R_{N\times N} - R_{N/2\times N/2} = \sum_{u,v}\log_2\left(\frac{\sqrt{2}e\sigma_{(u,v),N\times N}}{\Delta^2_{N\times N}}\right) - \log_2\left(\frac{\sqrt{2}e\sigma_{(u,v),N/2\times N/2}}{\Delta^2_{N/2\times N/2}}\right) = \sum_{u,v}\log_2\left(\frac{k_{(u,v)}}{\gamma^2_{(u,v)}}\right), \quad (8)$$

usually generates a negative value which in explains the advantage of larger MB sizes for HD contents. The histogram of the bit rate difference is shown in Fig. 1 (b).



(a) (b)

Figure 1. (a) The relationship between variances and (b) the histogram of bit rate differences for two block modes

Furthermore, in H.264, a significantly large percentage of blocks choose the SKIP mode so that no motion vector or residual information is to be transmitted. This contributes greatly to bit savings. Next, we investigate that majority of blocks can still be considered as the SKIP mode candidates by introducing larger MB size in HD contents,.

Previous work on DCT coefficients indicate that if the standard deviation of the DCT coefficient satisfies the following condition

$$3\sigma_F < 2.5\Delta, \quad (9)$$

99% of block coefficients will be quantized to zero [6]. Since the variance of the DC coefficient $\sigma_F^2(0,0)$ always has the largest magnitude. Thus, if DC coefficient is quantized to zero, then all other AC coefficients will be quantized to zero as well. $\sigma_F^2(u,v)$ can be approximated by the pixel input variance $\sigma_f^2(u,v)$ as:

$$\sigma_F^2 = \sigma_f^2[HRH^T]_{u,u}[HRH^T]_{v,v}, \quad (10)$$

where $R = \begin{bmatrix} 1 & \rho & \rho^2 & & \rho^7 \\ \rho & 1 & \rho & \cdots & \\ \rho^2 & \rho & 1 & & \\ & \vdots & & \ddots & \\ \rho^7 & & & & 1 \end{bmatrix}.$

For $\rho = 0.6$, we have

$$\sigma_F^2 = \sigma_f^2 \begin{bmatrix} 9.48 & 5.55 & 3.38 & & 0.80 \\ 5.55 & 3.25 & 1.98 & \cdots & \\ 3.38 & 1.98 & 1.20 & & \\ & \vdots & & \ddots & \\ 0.80 & & & & 0.07 \end{bmatrix}. \quad (11)$$

From (9)-(11), we can find the approximated condition for all DCT coefficients to be quantized to zero is

$$\sigma_f < \frac{2.5\Delta}{3\sqrt{[ARA^T]_{0,0}}}. \quad (12)$$

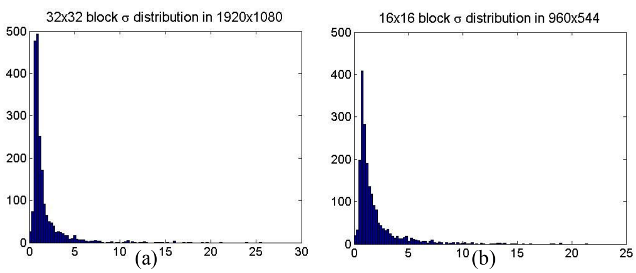


Figure 2. Block stastistics distribution for different resolutions in surface regions.

For targeted 10Mbps in HD video, we have $\Delta \approx 6$ (equivalent to $qp \approx 20$ in H.264 $\Delta = 2^{(qp-4)/6}$). We see from Fig. 2 that the larger block has 10 percents more blocks that meet the SKIP mode criterion as specified in (12).

III. Enlarged MV Search Range (SR) for HD Video

We justified the use of increased MB size for HD contents in Sec. II. However, this proposed coding gain seems to diminish by a large percentage in actual coding environment. We explore this issue by investigating the displacement difference produced by the current coding scheme. Given a displacement block after MC, the true displacement difference d can be expressed as a sum of the displacement difference after quantization $\hat{d}$ and its associated quantization error ε_Δ as $d = \hat{d} + \varepsilon_\Delta$. As quantization step size Δ increases, ε_Δ will increase, which reduces the information in the quantized displacement block/frame.

In the case of encoding HD contents, Δ is usually set to a smaller value to keep higher fidelity. Hence, ε_Δ is kept relatively low. Consequently, the remaining $\hat{d}$ tends to contain more information/texture compared to the low bit rate environment. Thus, it is important to find a more accurate motion prediction to reduce *d*. There exist two popular techniques: 1) exploiting extra temporal correlation from blocks from *t-i* or *t+i* frames, or 2) finding more spatial correlation from neighboring blocks in the same frame. The first technique proves to be extremely effective in the case of B frames. A significant coding gain can be achieved by considering more blocks along the time axis. However, further exploring the temporal correlation would not provide much additional coding gain for HD video. On the other hand, motion correlation in the spatial domain is reduced as the spatial resolution scales up. The current search range (SR) used for SD video might no longer be sufficient to cover the extended motion range of moving objects. The SR has to be increased adaptively in higher resolution to compensate for the extended motion distance and thus keep $\hat{d}$ small.

Fig.3 gives displacement frame differences extracted from the H.264 encoder with different SRs. With SR=32, the displacement frame still show a clear outline of the cars in the original YUV. By increasing SR to 96, most of the details are removed and the overall pixel intensity range of the displacement difference is reduced by almost half. This example shows that better MC is achieved with larger SR for HD contents.

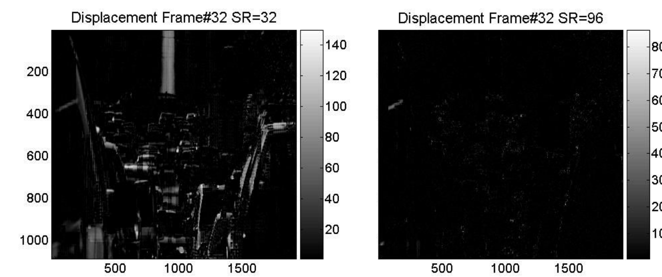


Figure 3. Displacement frames generated from sequence rush hour HD (1920x1080) with different search ranges.

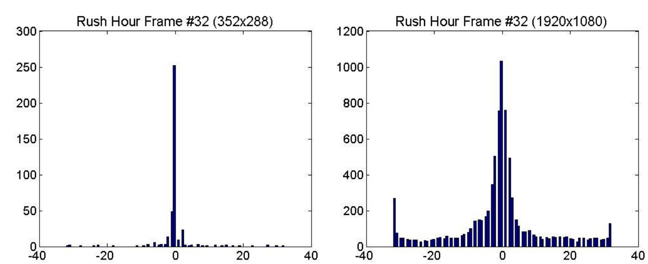


Figure 4. MV distributions in vertical direction with SR=32 for the rush hour sequence with different resolutions.

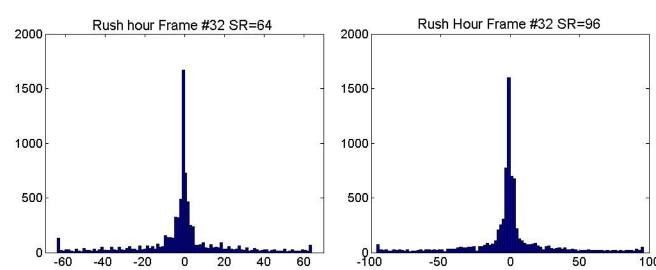


Figure 5. Comparsion of MV distributions in vertical direction with two search ranges for the rush hour sequence.

Furthermore, we see from Fig.4 that the motion vector distribution differs significantly in HD contents. The two peaks on tails indicate that suboptimal matches stop on the boundary of the SR. It is shown in Fig.5 that, as SR increases, the two tail peaks are gradually smoothed out. In other words, by increasing the search range in the HD content, the scattered motion correlation can be better exploited.

IV. EXPERIMENTAL RESULTS

We implemented a block-based coder similar to H.264 with an increased block size of 32x32 in our experiment. The displacement difference block of the Luma channel (no Chroma channel) after motion estimation was encoded by a 4x4 DCT transform and a uniform scalar quantizer, followed by an entropy coder. Motion vectors were coded separately. A Lagrangian cost function was used to select the best mode based on the rate distortion model in Section II.A. Rate and distortion statistics were first collected based on macroblock size no larger than 16x16 (N=16). The 32x32 macroblock size (N=32) was then added to study the coder behavior. We further combined the increased macroblock size with an enlarged search range of 96 (SR=96) for performance comparison.

Table 1 shows the percentage of the best mode using 32x32 block size with two different SR values. We see that with an enlarged SR values (from 32 to 96), the chance of choosing 32x32 blocks to be the best mode is nearly doubled. The importance of enlarged SR with respect to the large block is obvious.

TABLE I. PERCENTAGE OF BLOCK MODES USING 32X32 BLOCK SIZE

	Rush Hour		Blue Sky	
QP	**SR=32**	**SR=96**	**SR=32**	**SR=96**
12	5.02%	9.62%	5.27%	11.54%
16	7.50%	15.66%	7.75%	16.71%
20	12.75%	34.25%	13.09%	36.53%
24	19.49%	48.06%	22.30%	55.02%
28	28.04%	62.03%	30.49%	66.15%
32	39.12%	75.22%	40.34%	80.21%
36	46.18%	81.99%	49.12%	84.32%
40	53.27%	89.07%	62.97%	91.02%

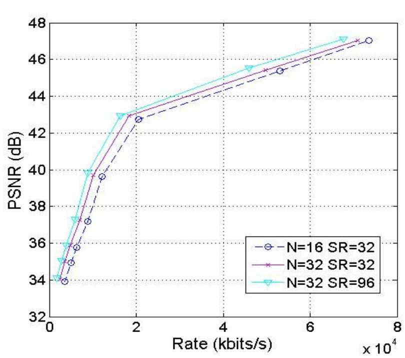


Figure 6. The rate distortion curve for blue sky 1920x1080 @30Hz

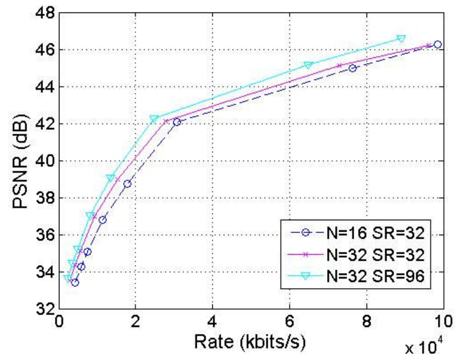


Figure 7. The rate distortion curve for rush hour 1920x1080 @30Hz

The empirical R-D curves for different QP settings are shown in Fig.6 and Fig.7. We see that the R-D tradeoff improves with an increased macroblock size (N) and SR. The saving becomes more obvious as the bit rate (or the PSNR) becomes higher. With extended SR and a larger macroblock size, we can achieve an average of 27% bit saving across the QP range.

V. CONCLUSION AND FUTURE WORK

The R-D model for HD video coding was derived to show that the R-D gain achieved by introducing a larger block size in flat regions. It was also shown that there could exist more blocks with an increased size in HD to be coded with efficient SKIP or Direct mode to achieve bit savings furthermore. The impact of enlarged motion SR in HD contents was studied and confirmed by experimental results. Our study is however still preliminary. In addition to the block sizes and search ranges, we would like to search for other means to improve HD video coding performance in the near future.

REFERENCES

[1] U.S. Entertainment Industry: 2005 MPA Market Statistics: http://www.mpaa.org/05%20Economic%20Review.pdf

[2] Digital Cinema System Specification V1.0, July, 2005 http://www.dcimovies.com/DCI_Digital_Cinema_System_Spec_v1.pdf

[3] G.J.Sullivan, P.Topiwala and A.Luthra, The H.264/AVC advanced video coding standard: overview and introduction to the fidelity range extensions," SPIE Conf. Applications of Digital Image, Processing XXVII, Vol. 5558, pp. 454-474, August 2004

[4] Stephane Mallat and Federic Falzon, Analysis of low bit rate image transform coding, IEEE Trans. on Signal Processing, Vol. 46, No.4, April 1998

[5] Hsueh-Ming and Jiann-Jone Chen, Source model for transform video coder and its application – part I: fundermental theory, IEEE Trans. Circuits Syst. for Video Technol., Vol. 7, No.2, April 1997

[6] Ming-Ting Sun and I-Ming Pao, Statistical computation of discrete transform in video coders, Journal of Visual Communication and Image Representation, June 1998

[7] S. Naito, A. Matsumura and A.Koike, Efficient coding scheme for super high definition video based on extending H.264 high profile, Proc. of SPIE-IS&T Electronic Imaging, 2005