MACROBLOCK SAMPLING AND MODE RANKING FOR COMPLEXITY SCALABILITY IN MOBILE H.264 VIDEO CODING

Tiago A. da Fonseca and Ricardo L. de Queiroz

Universidade de Brasilia
Department of Electrical Engineering
Brasilia, Brazil
`tiago@image.unb.br, queiroz@ieee.org`

ABSTRACT

We propose a framework for complexity scalability in H.264. The prediction is constrained so that only a subset of prediction modes are tested. The test subset is found by ranking the most "popular" modes (those the are most often picked as best) and selecting the modes that maximize their expected occurrence frequency given a complexity constraint. Ranking is performed by selecting a small set of macroblocks (the sampling population), for which all modes are tested. The remaining macroblocks only test the available dominant modes. The modes in the sampling population of a frame are used to process the next frame. Results are shown to verify the performance of the proposed method, which reveals sizeable complexity savings at small penalties.

***Index Terms*—** H.264/AVC, mode decision, complexity scalability, mobile video coding.

1. INTRODUCTION

In H.264/AVC [1], the many small improvements over previous encoding methods led to enhanced coding efficiency for a wide range of applications including video telephony, video conferencing, digital TV, streaming video etc. The H.264/AVC coder has been very well described in the literature [2, 3, 4, 5]. As a hybrid DPCM encoder, the H.264 searches for the best possible prediction of the encoding signal in order to spend less bits in its compressed representation. When compressing digital video, temporal (Inter) and spatial (Intra) redundancies are explored. "Inter" prediction generates prediction from one or more previously encoded video frames using block-based motion estimation and compensation [6]. A prediction block can also be formed based on planar extrapolation of previously encoded and reconstructed neighbouring pixels ("Intra" prediction). The encoder may either select the prediction mode for each block that minimizes the difference between the predicted block and the block to be encoded, or to use rate-distortion (RD) optimization [7].

Table 1. Relative computational complexity for Mobile [CIF] in JM13.2 H.264/AVC *High Profile* coding and various motion estimation search-window sizes.

	Window Size (pixels)/QP			
Coding Stage	**32**/23	**32**/36	**64**/23	**64**/36
Inter Prediction	64.7%	73.2%	73.2%	70.3%
4×4 Intra Prediction	8.0%	6.1%	6.6%	4.9%
8×8 Intra Prediction	6.9%	5.1%	5.8%	4.1%
16×16 Intra Prediction	0.8%	0.8%	0.7%	0.7%
Other Stages	19.6%	14.8%	16.6%	12.1%
Total	100.0%			

When encoding video sequences in a mobile scenario, complexity is an issue and real-time video coding is challenging. H.264 optimal software encoder implemenations demand high computation due to the need to test all these prediction mode options. As they take most of the time during encoding, reducing the complexity of this stage seems to be the natural way to reduce overall complexity.

Table 1 shows complexity profiling numbers for the High-Profile in JM13.2[1]. We used `gprof`[2], turned RD Optimization on, set the number of reference frames to 5, and used *UMHexS* motion estimation [8]. These results are similar to those in [9] and confirm that the encoder spends great part of the execution time doing "Inter" prediction for all different modes.

2. COMPLEXITY REDUCTION TECHNIQUES

As shown in [9] and in Table 1, the encoder complexity distribution indicates that complexity reduction can be most efficiently achieved through simplifying the prediction module, particularly the motion estimation step.

Sub-optimal fast motion estimation techniques were proposed [8, 10] and incorporated in the H.264/AVC reference software. Through exploring the variety of macroblock partition options available to H.264/AVC, there are works [11, 12] that only carry motion estimation for the most probable partition. The early prediction of skipped macroblocks prior to

---

This work was supported by HP Brasil and by Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, CNPq, Brasil, under grant 473696/2007-0.

[1] JM Available: http://iphome.hhi.de/suehring/tml/

[2] http://www.gnu.org/software/binutils/

motion estimation [13, 14] is also used for complexity control. Intra-prediction tests can also be reduced by means of selecting the most probable best mode according to heuristics [15, 16]. Combinations of previously proposed complexity reduction techniques were implemented in H.264, achieving substantial computational savings without a complexity control mechanism [17]. An effective approach is to generalize the rate-distortion analysis to add a complexity optimization variable. This concept is well suited to the emerging field of wireless digital video communications, where energy and delay constraints are stringent [18, 19, 20].

3. COMPLEXITY CONTROL MODEL

The H.264/AVC prediction stage is rather complex due to the many tests on the various prediction modes available to each macroblock. However, when compressing video sequences, we verify that the prediction modes applied to encode the signals become concentrated in small classes. The frequency profile of selected prediction modes for "Pedestrian Area" sequence at different resolutions, ranging from QCIF (176×144 pixels) to 1080p (1920×1080 pixels), is shown in Fig. 1. Each graph color represents a prediction mode [1] and $P \leq 8 \times 8$ incorporates all motion compensated prediction modes where the block size is smaller or equal to 8×8-pixels. We can observe for this sequence that as we increase the resolution, prediction modes tend to polarize themselves around larger macroblock partitions. Profiles were also built for many other sequences [21] and the general behavior suggests that some computational effort could be saved when encoding video sequences by avoiding less frequent macroblock prediction modes.

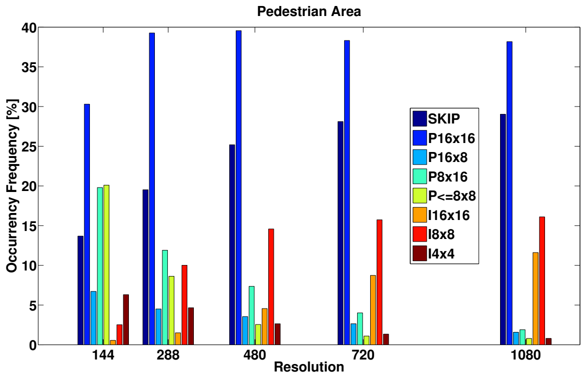


Fig. 1. Prediction modes occurrency frequency × resolution for Pedestrian Area sequence.

Complexity is scaled based on the suppression of less frequent prediction modes using the framework shown in Fig. 2. We begin by randomly selecting few macroblocks to preview the frequency distribuition of the prediction modes of the next frame. Then, we select the dominant modes, i.e. the modes the ocurred the most in the sampling set. This dominant mode set should ensure we meet the desired complexity C and is only used for the next frame. The algorithm follows: Let each frame have N macroblocks and the set of dominant modes ($\mathbf{D}$) initially comprise all modes. For the n-th P- or B-frame:

1. Randomly select a set ($\mathbf{S}$) of N_S macroblocks from the n-th frame. The remaining $N - N_S$ macroblocks form the complement set $\mathbf{S}'$.
2. Test all prediction modes for macroblocks in set $\mathbf{S}$, in order to pick the best mode.
3. Test only dominant modes in $\mathbf{D}$ for macroblocks in $\mathbf{S}'$.
4. Rank best modes in set $\mathbf{S}$. While the accumulated complexity is within a desired complexity budget C, assign the top modes to $\mathbf{D}$.
5. Set $n \leftarrow n + 1$ and repeat.

Even though the best mode frequency distribution is not stationary, tests have shown that this is a good approximation for adjacent frames [21]. Errors in finding the dominant modes will be reflected by a small degradation on the encoder RD performance.

The sampling population size, N_S, is found rather empirically. The smaller N_S, the larger the savings, but the worse the performance. In [21], we found that $N_S = N/10$ is satisfactory in many cases and we used this value in all simulations.

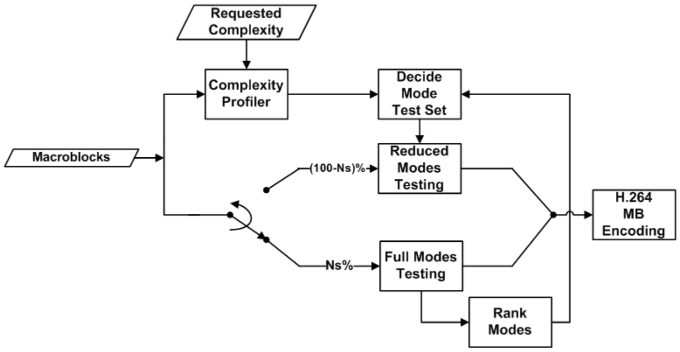


Fig. 2. Complexity control framework.

In the ranking step of the above algorithm, each mode is associated with a frequency of occurrence and a complexity (estimated execution time). We exhaustively search for the set of prediction modes, whose added complexities fall below a complexity budget C, that maximizes the combined frequencies.

4. EXPERIMENTAL RESULTS

The proposed method was implemented in the H.264 reference software JM13.2, using *UMHexS* motion estimation. Complexity was measured in net execution time. Fig. 3 presents RD perfomance curves for "Mobile" CIF sequence, for different complexity settings: full (100% of encoding complexity), 80% of encoding complexity, 60% of encoding complexity and so on. We observe that the modified codec performance is very close to the reference verification model in such a way that we can not readly perceive the visual diferences.

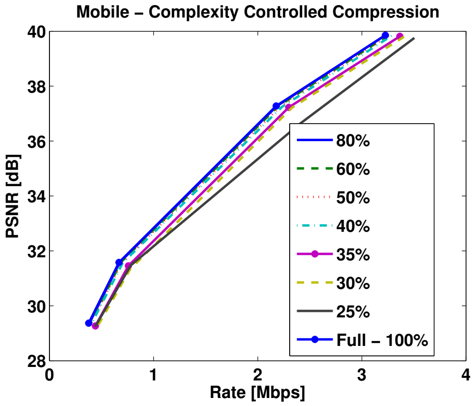


Fig. 3. Different complexity settings rate-distortion curves for sequence "Mobile [CIF]".

A more detailed approach is to plot the average difference between the performance curves for different sampling population sizes, as shown in Figs. 4 and 5. The average PSNR and bitrate differences between RD curves were evaluated as described in [22]. There is a small quality loss when predicting through dominant modes due to eventual mismatches. The rate loss is small (below 5%) if the complexity savings remains bellow 50% for CIF sequences. For QCIF sequences, the rate increase is around 10%. Complexity savings and compression performance are traded-off and losses can be reduced when the sequence is such that resolution is increased or when the prediction modes polarize themselves around a small mode subset.

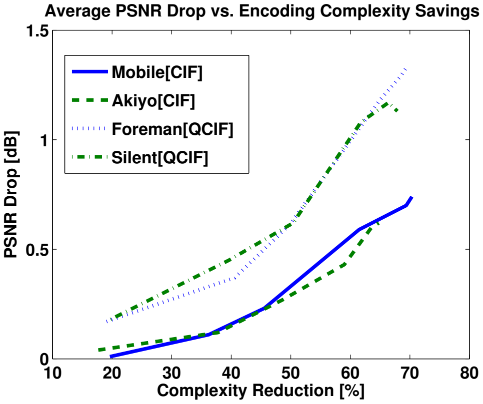


Fig. 4. Average PSNR drop *vs.* complexity reduction for four video sequences: "Mobile [CIF]", "Akyio [CIF]", "Foreman [QCIF]" and "Silent [QCIF]".

In order to measure the effect of resolution, sequences were encoded at different resolutions, from QCIF to 1080p. The results are presented in Fig. 6, where SD stands for 720×480-pixels frame size. There are more savings for higher resolution sequences than for lower ones, but the trend is not continuous.

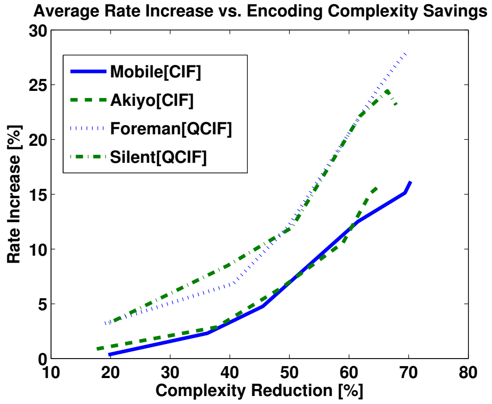
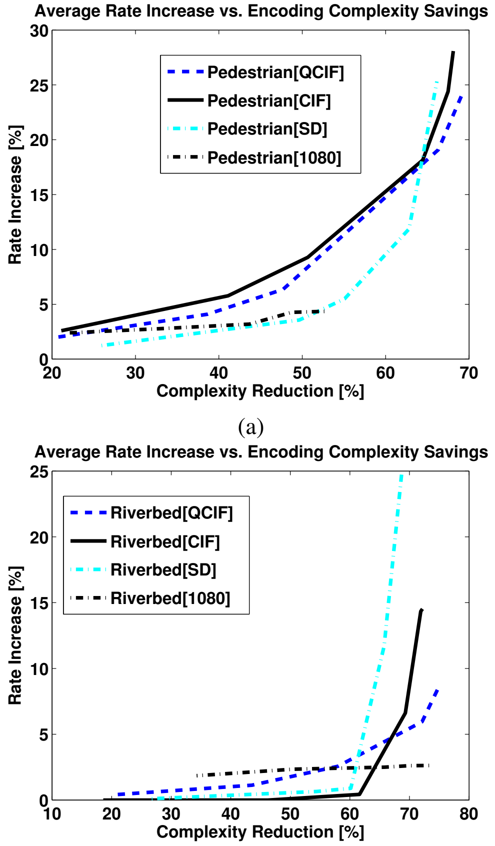


Fig. 5. Average rate increase *vs.* complexity reduction for four video sequences: "Mobile [CIF]", "Akyio [CIF]", "Foreman [QCIF]" and "Silent [QCIF]".



(a)

(b)

Fig. 6. Average rate increase *vs.* complexity reduction for two video sequences at different resolutions: (a) "Pedestrian Area" and (b) "Riverbed".

5. CONCLUSIONS

We proposed a fully compliant complexity control framework for H.264/AVC that should work for most applications ranging from mobile devices to HD. Rather than testing all prediction modes available, we search for dominant mode subsets, i.e. the encoder only tests the estimated "most frequently selected modes" until a complexity budget is met. The estimation is done by carrying full mode search in a reduced population of macroblocks and using the resulting dominant set information while processing the next frame. Our tests have shown that the RD performance is only moderately affected by skipping prediction modes, which means the skipped modes are not always very relevant, even though there are many mismatches. Nevertheless, we achieved significant complexity reduction. We plan to further work on making complexity targets to change along the frame such that we would still meet a complexity budget for the whole frame of group of frames, but to adapt depending on each macroblock characteristics.

6. REFERENCES

[1] JVT of ISO/IEC MPEG and ITU-T VCEG, "Advanced video coding for generic audiovisual services," Tech. Rep. 14496-10:2005, March 2005.

[2] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, John Wiley & Sons Ltd, 2003.

[3] T. Wiegand, G. J. Sullivan, G. Bjontergaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[4] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions," *Proc. SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC*, Aug. 2004.

[5] R. L. de Queiroz, R. S. Ortis, A. Zaghetto, and T. A. Fonseca, "Fringe benefits of the H.264/AVC," in *Proc. International Telecommunication Symposium*, 2006, pp. 208–212.

[6] K. Sayood, *Introduction to Data Compression*, Morgan Kuffmann Publishers, 2000.

[7] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *IEEE Circuits and Systems Magazine*, pp. 7–28, Jan-Mar 2004.

[8] Z. Chen, P. Zhou, and Y. He, "Hybrid unsymmetrical-cross multi-hexagon-grid search strategy for integer pel motion estimation in H.264," *Proc. Picture Coding Symposium*, Apr. 2003.

[9] Y.-Y. Huang, B.-Y. Hsieh, S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 4, pp. 507–522, Apr. 2006.

[10] H.-Y. C. Tourapis, A. M. Tourapis, and P. Topiwala, "Fast motion estimation within the H.264 Codec," *Proc. of Intl. Conf. on Multimedia and Expo. ICME*, vol. 3, pp. 517–520, July 2003.

[11] T.-Y. Kuo and C.-H. Chan, "Fast variable block size motion estimation for H.264 using likelihood and correlation of motion field," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 10, pp. 1185–1195, Oct. 2006.

[12] B.G. Kim, S.-K. Song, and C.-S. Cho, "Efficient inter-mode decision based on contextual prediction for the P-slice in H.264/AVC video coding," *Proc. IEEE International Conference on Image Processing*, pp. 1333–1336, September 2006.

[13] C. S. Kannangara, I.E.G. Richardson, M. Bystrom, J. Solera, Y. Zhao, A. MacLennan, and R. Cooney, "Complexity reduction of H.264 using lagrange multiplier methods," *Proc. IEEE Int. Conf. on Visual Information Engineering*, April 2005.

[14] C. S. Kannangara, Y. Zhao, I. E. Richardson, and M. Bystrom, "Complexity control of H.264 based on a bayesian framework," *Proc. Picture Coding Symposium*, November 2007.

[15] B. La, M. Eom, and Y. Choe, "Fast mode decision for intra prediction in H.264/AVC encoder," *Proc. IEEE International Conference on Image Processing*, vol. V, pp. 321–324, September 2007.

[16] C. Hwang, S.S. Zhuang, and S.-H. Lai, "Efficient intra mode selection using image structure tensor for H.264/AVC," *Proc. IEEE International Conference on Image Processing*, vol. V, pp. 289–292, September 2007.

[17] R. Almeida and F. Pereira, "H.264/AVC low complexity encoding: a novel fast coding mode selection solution," in *Proc. Conf. on Telecommunications - ConfTele*, Peniche, Portugal, May 2007.

[18] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 645–658, May 2005.

[19] E. Akyol, D. Mukherjee, and Y. Liu, "Complexity Control for Real-Time Video Coding," *Proc. IEEE International Conference on Image Processing*, vol. I, pp. 77–80, September 2007.

[20] R. Vanam, E. A. Riskin, S. S. Hemami, and R. E. Ladner, "Distortion-complexity optimization of the h.264/mpeg-4 avc encoder using the GBFOS algorithm," in *DCC '07: Proc. of the 2007 Data Compression Conference*, Washington, DC, USA, 2007, pp. 303–312, IEEE Computer Society.

[21] T. A. da Fonseca and R. L. de Queiroz, "Complexity reduction techniques applied to the compression of high definition sequences in digital TV," *Proc. XXVI Simposio Brasileiro de Telecomunicacoes*, pp. 43–47, September 2008.

[22] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, April 2001.