Complexity Reduction Techniques Applied to The Compression of High Definition Sequences in Digital TV

Tiago A. da Fonseca and Ricardo L. de Queiroz

***Resumo*— O H.264/AVC é um codificador de modulação de código de pulso com estrutura híbrida em que são conectados módulos preditivos e de transformação. Neste trabalho propomos uma abordagem mais eficiente para a implementação do módulo de predição, notavelmente o estágio mais complexo computacionalmente desse codificador. A idéia é testar apenas subconjunto de modos de predição dominantes em vez de fazer o teste exaustivo de todos os modos recomendados pelo padrão H.264/AVC. Resultados mostram que, para seqüências de vídeo de alta resolução, a perda de qualidade é desprezível e abre a possibilidade de inserção de um controlador de complexidade de computacional para codificação de seqüências de vídeo de alta resolução, usadas em TV digital.**

***Palavras-Chave*— H.264/AVC, decisão de modos de predição, vídeo de alta resolução, redução de complexidade, TV digital.**

***Abstract*— H.264/AVC is a hybrid predictive-transform coder. In this paper we propose a more computational efficient approach to implement the prediction module, its most complex stage. The idea is to employ a subset of dominant prediction modes instead of testing all modes recommended by H.264/AVC standard. Results show that, for high definition sequences, the quality loss is negligible allowing us to control the compression complexity of high definition digital TV video sequences.**

***Keywords*— H.264/AVC, mode decision, high-definition, reduced complexity, Digital TV.**

I. INTRODUCTION

H.264/AVC is the latest international video coding standard [1]. Is was jointly developed by the Video Coding Experts Group (VCEG) of the ITU-T and the Moving Picture Experts Group (MPEG) of ISO/IEC. The many small improvements over previous encoding methods added up and promoted 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 well described in the literature [2]-[5], showing performance comparisons against other coders and also exploring less known features of the H.264/AVC.

II. MACROBLOCK PREDICTION IN H.264/AVC

H.264/AVC is a hybrid DPCM video codec, i.e. along with a transform module, it has a prediction module, a differential stage and a feedback loop [6]. In order to achieve substantial compression, a DPCM coder aims at performing the best possible prediction of the enconding signal in order to spend less bits to represent the residue.

Fig. 1 depicts the H.264/AVC encoder block diagram. Note that it can be divided into temporal (Inter) and spatial (Intra) predictions modules.

"Inter" prediction generates a prediction macroblock from one or more previously encoded video frames using block-based motion estimation and compensation. This model is responsible for almost 90% of the complexity of an H.264/AVC baseline encoder [7]. Important advances from earlier video standards include the support for a range of block sizes (16x16 and down, as in Fig. 2) and refined motion vectors (quarter-sample resolution for the luminance component).

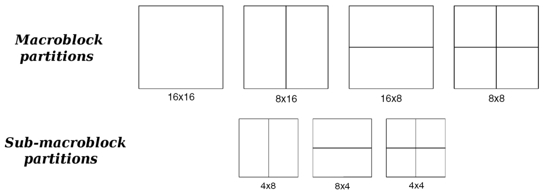


Fig. 2. Macroblock and submacroblock partitions for motion compensation in Inter Prediction.

In "Intra" prediction, a prediction block is formed based on planar extrapolation of previously encoded and reconstructed neighbouring pixels. The prediction is subtracted from the current block, prior to encoding. A macroblock can be partitioned in blocks of 4x4, 8x8 or 16x16 pixels. The former ones have a total of nine optional prediction modes for luminance while the latter has only four modes as illustrated in Fig. 3. The encoder typically selects the prediction mode for each block that minimizes the difference between the predicted block and the block to be encoded.

III. COMPLEXITY REDUCTION TECHNIQUES

Although very effective, the H.264/AVC prediction stage is rather complex due to the large set of models employed. Sub-optimal motion estimation techniques were proposed [9]-[10] and incorporated in the H.264/AVC reference software[1] [1]. The main idea is to apply heuristics to reduce the search for a block match. There is coding time reductions along with small rate-distortion performance losses compared to full-search motion estimation. In exploring the variety of

---

The authors are with Department of Electrical Engineering, Universidade de Brasilia, e-mail tiago@image.unb.br and queiroz@ieee.org. This work was supported by HP Brasil and by CNPq under grant 474912/2006-0.

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

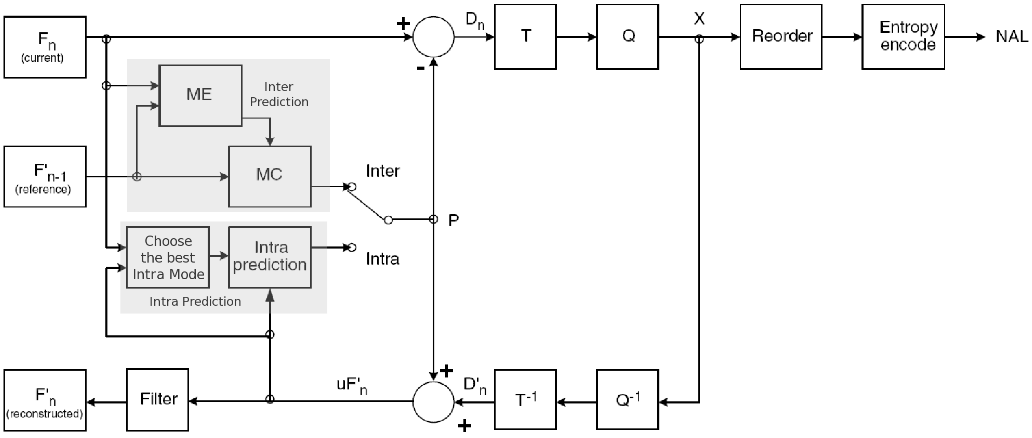
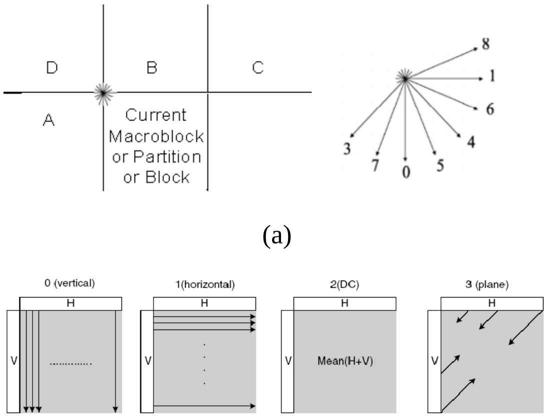


Fig. 1. H.264/AVC Encoder block diagram. Prediction stages are highlighted.



(a)

(b)

Fig. 3. Intra prediction modes and their respective planar extrapolation directions for (a) 4x4/8x8 blocks and (b) 16x16 blocks. In (a), mode 2 is DC prediction.

macroblock partitions available in H.264/AVC, there are works [11],[12] that apply motion estimation only on the most probable partition, determined through techniques that explore informations on neighbouring blocks.

Intra-prediction tests can also be reduced by means of selection of the most probable best mode according to heuristics [15],[16].

Another approach is to treat the encoding computational complexity as a scarce resource, generalizing the rate-distortion analisys and adding a third optimization variable. This concept is well suited to the emerging field of wireless digital video communications, where energy and delay constraints are stringent. [13],[14]

IV. PREDICTION MODE BIAS

As described, H.264/AVC is structured as a hybrid DPCM video coder. Its prediction stage is rather complex due to the many tests of various prediction modes available to each macroblock. For instance, to encode P-frames in H.264/AVC *High Profile*, we can use the following set of Inter- and Intra-frame prediction modes:

- P16×16: motion compensated prediction for 16×16 pixels macroblocks;
- P16×8: motion compensated prediction for 16×8 pixels macroblocks;
- P8×16: motion compensated prediction for 8×16 pixels macroblocks;
- P<=8×8: motion compensated prediction for macroblocks whose size is less then or equal to 8×8 pixels;
- I16MB: intra prediction for 16×16 pixels macroblocks;
- I8MB: intra prediction for 8×8 pixels macroblocks;
- I4MB: intra prediction for 4×4 pixels macroblocks;
- SKIP: zero residue motion compesanted prediction for 16×16 pixels macroblocks.

However, when compressing high definition 1080p video sequences (1920×1080 pixels per frame), we verify that the prediction modes applied to encode the signals become concentrated in small classes. The frequency profile of selected prediction modes for different sequences and resolutions, ranging from QCIF (176×144 pixels) to 1080p (1920×1080 pixels), is presented in Figs. 4 through 6. The test sequences were the following standard ones and its downsampled versions:

- Pedestrian Area (1920x1080, 25 fps, progressive) is a shot of a pedestrian area. The camera is static at a low position while pedestrians pass by.
- Sunflower (1920x1080, 25 fps, progressive) is a very detailed shot. There is a bee at the sunflower, with small color differences. The camera is fixed and the scene has slow global motion.
- Rush-hour (1920x1080, 25 fps, progressive) is a shot of rush-hour in Munich. There are many cars moving slowly, a high depth of focus and the camera is fixed.
- Riverbed (1920x1080, 25 fps, progressive) is a shot of a riverbed seen through the water. Challenging compression.

We can observe that when we increase the resolution, prediction modes tend to polarize themselves around bigger macroblock partitions. The only sequence that does not follow this tendency is Riverbed due the high occurrency of Intra-

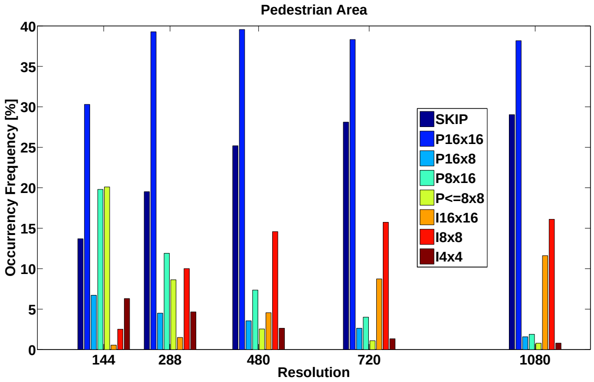


Fig. 4. Prediction modes occurrency frequency × resolution for Pedestrian Area sequence.

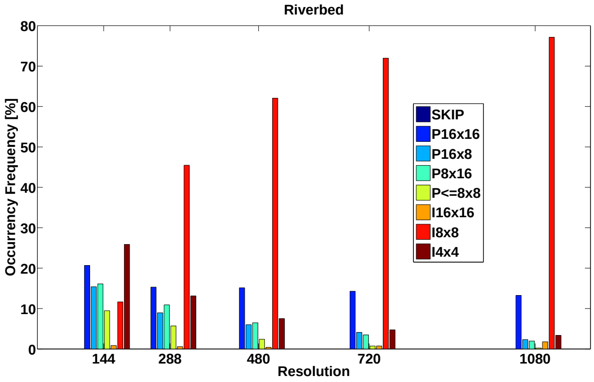


Fig. 5. Prediction modes occurrency frequency × resolution for Riverbed sequence.

predicted 8×8 pixels macroblocks for higher resolutions. However, the general behavior suggests that some computational effort could be saved when encoding high definition video sequences by avoiding small-sized partitions in motion compensated predictions.

V. REDUCED MODE SET PREDICTION

The analisys of Figs. 4 through 6 suggests that the encoder can save computational time if it was kwnown, *a priori*, the frequency distribution of prediction modes. Thus, one can prune less frequent prediction modes.

In order to achieve complexity reduction based on frequency distribution of prediction modes, we propose the metodology illustrated in Fig. 7. First, we randomly select macroblocks which will compose the sampling population, employed to preview the frequency distribution of the next frame; then, we proceed descend sorting the prediction modes and, finally, selecting the dominant modes, here arbitrated as the set of modes which corresponds at least to 80% of the choices.

Applying the current frame prediction mode frequency distribution as prediction of next frame distribution is an error prone approach, since we are considering the frequency distribution as stationary. This is not allways verified, however, in practice, this is a good approximation. Errors in determining the dominant modes will be reflected by a small degradation on the encoder rate-distortion performance.

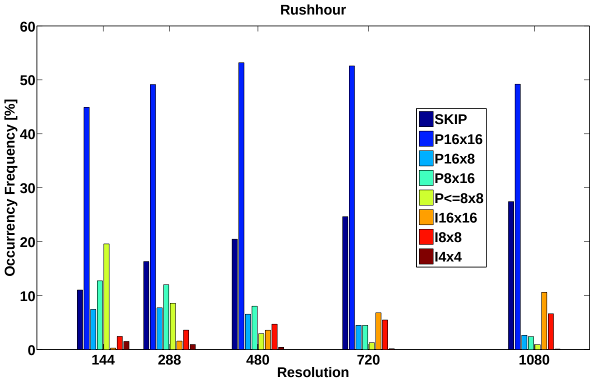


Fig. 6. Prediction modes occurrency frequency × resolution for Rushhour sequence.

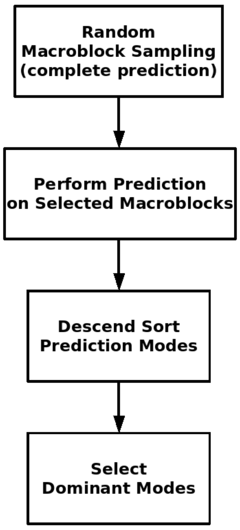


Fig. 7. Dominant prediction mode selection metodology.

A question that arises is what is the size of the fast-predicted macroblock population, i.e. how many macroblocks are to be submitted to suboptimal prediction tests, and what is the suitable sampling population size, which will be used in the prediction of next frame dominant modes. The proposed methodology considers all "full-prediction" macroblocks of the current frame to estimate the mode occurrency frequency, in order to predict the mode for the next frame. The remaining macroblocks are submitted only to dominant mode predictions. The curves in Fig. 8 relate complexity savings against the population size of fully tested macroblocks, using the proposed method and the high-definition sequences Pedestrian Area, Riverbed, Rushhour and Sunflower.

The analisys of Fig. 8 suggests that the size of the fast predicted macroblock population for the set of employed test sequences has a direct relation to the achieved complexity savings. This opens an opportunity to provide a complexity controlled compression through the supression of less frequent prediction modes.

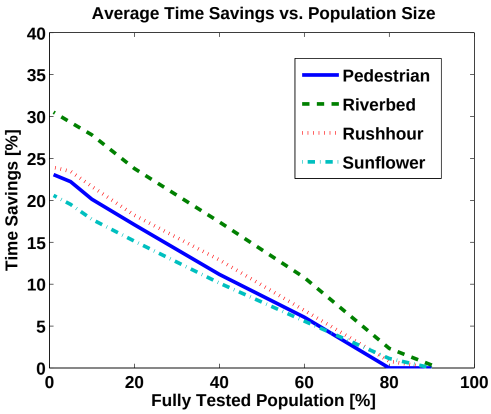


Fig. 8. Time savings *vs.* Population size for different HD video sequences.

VI. EXPERIMENTAL RESULTS

The proposed modification was implemented in the H.264/AVC reference software JM12.3. Each sequence was made up of 20 frames. Fast full-search motion estimation was employed and the results were obtained by varying the QP (quantization parameter) over the range $12 \leq QP \leq 36$. Within this range, we respect HD video constraints of quality and rate for broadcasting. Fig. 9 presents rate-distortion perfomance curves for Pedestrian Area sequence, for different sample sizes: Original (100% of macroblocks are fully tested), 90% of fully-tested macroblocks, 80% of fully-tested macroblocks 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 diferences between them.

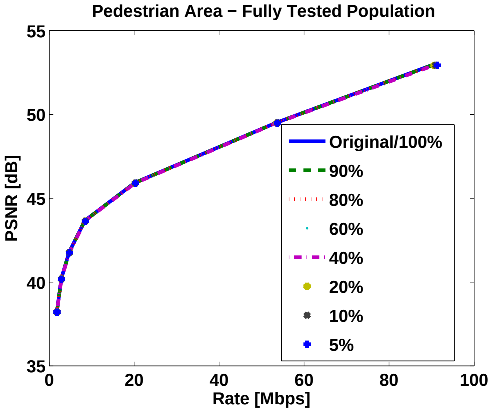


Fig. 9. Different tested population size rate-distortion curves for Pedestrian Area. The curves are essentially co-located.

A more detailed approach is to plot the average difference between the performance curves for different sampling population sizes, as show in Figs. 10 and 11. The average PSNR and bitrate differences between RD-curves were evaluated as described in [18].

We can observe a very small quality loss when predicting only through dominant modes. This is due to eventual mismatches between the best prediction mode evaluated by the two methods. The rate loss is somewhat tolerated (below 5%) if the population size remains above 10%. Even though computational savings are relatively small, fast full-search motion estimation was enabled, i.e. motion estimation of one macroblock prediction mode is fully reused to other modes. Actually, in fast full-search motion estimation, the encoder carries 4×4-pixel-block motion estimation and makes greater partitions motion estimations by grouping the results (SAD/SSD) of previously stored blocks. Thus, for sequences where the intra-predicted macroblocks are more frequent, like Riverbed, the computational savings are greater due to the fact that motion estimated prediction modes are not included in dominant set for some frames. Once a motion estimated prediction mode is available, motion estimation is performed and can be reused for other partitions.

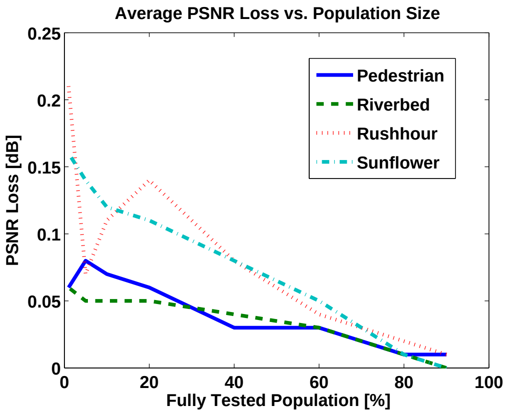


Fig. 10. Average PSNR Loss *vs.* Population size for different HD video sequences.

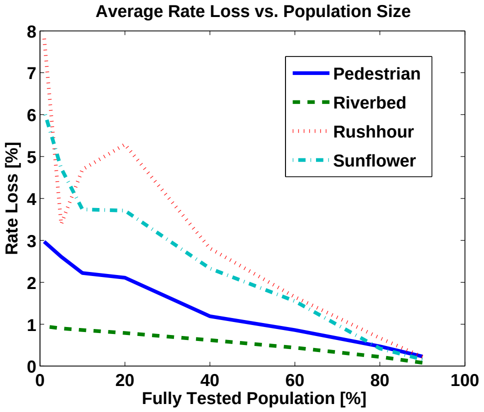


Fig. 11. Average Rate Loss *vs.* Population size for different HD video sequences.

VII. CONCLUSIONS

We propose a reduced-complexity way of carrying the prediction mode tests in H.264/AVC for high-definition sequences. Rather than testing all prediction modes available, we search for a "dominant" mode subset. In tests with broadcast quality coding of HD sequences, results show that the rate-distortion performance is only weakly affected by the prediction mode prunning, although the complexity reduction is significant. The method does not require a new decoder implementation because only non-normative codec aspects are modified. Future work will concentrate on heuristics implementation for sub-optimal motion estimation techniques [8]-[10], and on the design of a complexity controlled implementation of the H.264/AVC codec in order to reduce the computational complexity and cost of a digital TV encoder.

REFERENCES

[1] JVT of ISO/IEC MPEG and ITU-T VCEG, "Advanced Video Coding for Generic Audiovisual Services", March 2005.

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

[3] T. Wiegand, G. J. Sullivan, G. Bjøntergaard, 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 et al., "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*, August, 2004.

[5] R. L. de Queiroz, R. S. Ortis, A. Zaghetto, and T. A. da Fonseca, "Fringe benefits of the H.264/AVC", *Proc. of Intl. Telecom. Symp.*, Fortaleza, Brazil, pp. 208-212, Sep. 2006.

[6] K. Sayood, *Introduction to Data Compression*, 2nd edition, Morgan Kaufmann Publishers, 2000.

[7] 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 Trans. on Circuits and Systems for Video Technology*, Vol 16, No. 4, pp. 507-522, April 2006.

[8] Z. Chen, P. Zhou, Y. He, "Fast Integer Pel and Fractional Pel Motion estimation in for JVT", *JVTF017r1.doc*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 6th meeting, Awaji, Island, JP, December, 2002.

[9] Z. Chen, P. Zhou, Y. He, "Hybrid Unsymmetrical-cross Multi-Hexagon-grid Search Strategy for Integer Pel Motion Estimation in H.264", *Picture Coding Symposium*, April, 2003.

[10] H.-Y. C. Tourapis, A. M. Tourapis, P. Topiwala, "Fast Motion Estimation within the H.264 Codec", *Proc. of International Conference on Multimedia and Expo. ICME*, Vol. 3, pp. 517-520, 2003.

[11] T.-Y. Kuo, 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, October, 2006.

[12] B.G. Kim, S.-K. Song, C.-S. Cho, "Efficient Inter-Mode Decision Based on Contextual Prediction for the P-Slice in H.264/AVC Video Coding", *IEEE International Conference on Image Processing*, pp. 1333-1336, September, 2006.

[13] Z. He, Y. Liang, L. Chen, I. Ahmad, 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.

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

[15] B. La, M. Eom, Y. Choe, "Fast Mode Decision for Intra Prediction in H.264/AVC Encoder", *IEEE International Conference on Image Processing*, Vol. V, pp. 321-324, September, 2007.

[16] C. Hwang, S.S. Zhuang, S.-H. Lai, "Efficient Intra Mode Selection Using Image Structure Tensor for H.264/AVC", *IEEE International Conference on Image Processing*, Vol. V, pp. 289-292, September, 2007.

[17] Y.-Y. Huang, B.-Y. Hsieh, S.-Y. Chien, S.-Y. Ma, 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, April, 2006.

[18] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves.", Coc. VCEG-M33, Apr. 2001.