FAST MACROBLOCK INTRA AND INTER MODES SELECTION FOR H.264/AVC

E. Arsura, L. Del Vecchio, R. Lancini, L. Nisti

CEFRIEL – Politecnico di Milano
Via Fucini 2, 20133 Milano, Italy
arsura,delvecch, rosa,nisti@cefriel.it

ABSTRACT

The major complexity source in H.264 is the rate distortion (R-D) optimization. This work proposes a method to reduce the complexity of R-D analysis, paying a little video degradation, by estimating the encoding modes that are more likely used by the encoder. Since the encoding modes depend mainly on the typology of the video and on the quantization parameter used to encode each block, the work proposes to group the modes into subsets that depend on the quantizer level and the characteristics of the video. In particular, the motion compensability of each frame modulates the intra modes, while texture difficulty addresses the inter modes. The R-D algorithm evaluates only the modes included into the resulting subset.

The usage of subset modes reduces the complexity of the encoding process (40-70%) with respect to the original R-D optimized encoder, with a slight PSNR reduction (0-0,30dB).

1. INTRODUCTION

The latest video coding standard H.264/AVC [1] achieves an improved coding efficiency over all existing standards through added features and functionalities, at the expense of an increased implementation cost. In particular, motion estimation at different block sizes and an improved number of mode decisions contribute to a better video quality increasing a lot the complexity. The available MB modes [2] in H.264/AVC include two sizes for intra coding mode that are 4x4 and 16x16, and various inter coding modes such as skip, inter-16x16, inter-16x8, etc.. up to 4x4. The classical encoder has to examine all these modes to achieve the best prediction. This process is very computational demanding especially when R-D optimization is used, because the coder performs a complete encoding process including motion estimation (only for inter modes), transformation, quantization and entropy coding. Anyway if it is possible to predict a subset which includes the most probable modes that the encoder is going to use to encode a macroblock, the computational complexity could be strongly reduced. Recently fast inter mode selection algorithms were proposed in [3], [4] and [5] to alleviate the encoder complexity due to the motion estimation and inter mode decisions. These algorithms reduce the set of inter modes that the encoder needs to check depending on parameters like the complexity of the content, temporal similarity and AZCB (All-zero Coefficient Block) detection. In [6] a fast inter prediction mode decision based on pre-encoding process is presented. Other works [7] propose a fast mode decision algorithm to speed up intra coding prediction modes for both 4x4 and 16x16 sizes, based on edge direction histogram and directional fields.

This paper describes a new approach for both the inter and intra mode size selections, that takes into account the image content type (complexity and predictability of each frame) and also the quantization parameter used to encode each macroblock. In fact the mode selection also depends on the quantization parameter, since the video quality due to different quantization levels influences the choices on the different encoding tools.

In Section 2 a description of the proposed algorithms for inter and intra size mode selection is presented; Section 3 describes the obtained results. Conclusions are given in Section 4.

2. PROPOSED ALGORITHMS

A series of simulations [8] proved that the inter mode decision depends on the typology of the video which includes the complexity of the texture to be encoded and the motion content. Smaller block modes (i.e. 4x4, 4x8, etc..) are used only if the image is rich in details and the movement of macroblocks is very complex so that small parts have different moving directions.

The number of macroblocks intra coded depends mainly on the motion uncompensability of one frame from another; the uncompensability measures the amount of new contents in a video segment. Another factor that influences the mode decision is the quantization parameter: for example, if the video presents

0-7803-9332-5/05/$20.00 ©2005 IEEE

homogeneous regions and the quantizer is high, the skip mode is the most frequent choice of the encoder.
Based on these observations, this work proposes two methods which define subsets of the encoding block modes for both intra and inter cases. For each macroblock, the encoder will search the best encoding mode with R-D optimization technique, evaluating only the modes included into the subset addressed by the quantization parameter and the video typology of that macroblock.

2.1 Intra Mode Selection

For the intra mode subset selection, the proposed algorithm evaluates the uncompensability of each frame: the uncompensability is a parameter that determines how well a certain $frame_{i+1}$ is predictable by a previous $frame_i$. For example, if one frame corresponds to a scene change, the level of its uncompensability will be very high.
The main idea to determine the uncompensability is to identify particular scene details that are called *feature points*.
The feature points are used in conjunction with $frame_i$ motion vector to make a forward motion compensation with the aim to verify that the identified feature point is also present in $frame_{i+1}$. The number of not matched points is the level of uncompensability.
A feature point is obtained considering the variance of a NxN block, computed on the original frame as:

$$\sigma^2 = \frac{1}{N^2}\sum_{i=0}^{N}\sum_{j=0}^{N}(x_{ij}-\mu)^2$$

Where: μ is the mean of the NxN block and x_{ij} is the original block pixel.
If $\sigma^2 \geq TH_1$ then, the corresponding block is a candidate feature point. TH_1 is an optimal threshold computed over different of simulations.

The forward motion compensation matching criteria M, is based on the difference between the mean luminance of the feature point and the motion compensated one, compared with another threshold TH_2:

$$M = \begin{cases} 1 & |mean_y(FP) - mean_y(MCFP)| \geq TH_2 \\ 0 & otherwise \end{cases}$$

where: FP is the feature point in $frame_i$, $MCFP$ is the moto-compensated FP in $frame_{i+1}$ and $mean_y(...)$ is the mean luminance of the considered block. TH_2 is an appropriate threshold that has been chosen after a series of simulations over different video sequences.
The uncompensability (Muc) is computed as:

$$Muc = \frac{1}{frameBs}\sum_{i=1}^{frameBs} M_i$$

where: $frameBs$ is the number of blocks that constitutes the frame.

QP	Muc < T_{M1}		T_{M1}<Muc<T_{M2}		Muc>T_{M2}	
	4x4	16x16	4x4	16x16	4x4	16x16
1-7	No	No	Yes	Yes	Yes	Yes
8-12	No	No	Yes	Yes	Yes	Yes
13-17	No	No	Yes	Yes	Yes	Yes
18-22	No	No	Yes	No	Yes	Yes
23-27	No	No	Yes	No	No	Yes
28-32	No	No	No	No	No	Yes
33-37	No	No	No	No	No	Yes
38-42	No	No	No	No	No	Yes
43-51	No	No	No	No	No	Yes

Table 1: Subsets for intra modes.

The proposed algorithm uses three subsets to modulate the intra modes: they depend on quantization step size and the values of Muc. The statistic collection, in fact, reveals that at high quantizer values (i.e. QP>28) is not convenient to evaluate the smaller mode Intra-4x4 because it is not very often used. Also, the evaluation of the intra modes depends on the values of the Muc: if Muc is very small ($Muc<T_{M1}$), no one of the intra modes will be evaluated by the encoder; otherwise if the Muc is high ($Muc>T_{M2}$) at least the Intra 16x16 mode will be evaluated because it is very probable that a scene change or a sudden change in the video has been revealed. The complete subset selection containing the evaluation of the Intra modes is represented in Table 1.
The thresholds T_{M1} and T_{M2} have been chosen after a series of simulations on different video sequences.

2.2 Inter Mode Selection

The inter modes evaluated by motion estimation depend on the difficulty of the texture of each frame and the quantization parameter. The level of details for each frame is measured by estimating the bits that might be needed to encode that frame in intra mode with quantizer set to 1, in a similar way of the computation for the Difficulty Hint in MPEG-7 [9]. This work proposes a logarithmic model as a good estimator of the generated rate $R(Q)$ of each macroblock when the quantizer Q is 1:

$$D = R(\sigma^2) = \sum_{i=1}^{frameBs} a \cdot \log(b\sigma_i^2 + c)$$

where: R is the estimated rate for a certain frame, σ_i^2 is the variance of the original block i and a,b,c are model parameters. The constant parameters a,b,c were obtained empirically by fitting data derived by encoding some test sequences.

The inter modes are modulated by D and by the quantization step size. In general, for higher values of the quantizer (low resolution) the evaluation of the smallest modes (i.e smaller than 8x8) is discharged. Together with the QP, the subset is addressed comparing D with a threshold T_D: at low values of D ($D<T_D$) some of the smallest modes are also discharged from evaluation. Otherwise, if D is high and the QP is small (highest resolution), all the modes are considered: this case corresponds to images that are rich in details and need to be encoded with finest quality. The complete subset selection for the inter modes is reported in Table 2.
The threshold T_D has been chosen after a series of simulations on different video sequences.

QP	**D < T_D**				
	Skip	**16x16**	**16x8**	**8x16**	**Other**
1-7	Yes	Yes	Yes	Yes	Yes
8-12	Yes	Yes	Yes	Yes	Yes
13-17	Yes	Yes	Yes	Yes	Yes
18-22	Yes	Yes	Yes	Yes	Yes
23-27	Yes	Yes	Yes	No	Yes
28-32	Yes	Yes	No	No	No
33-37	Yes	Yes	No	No	No
38-42	Yes	Yes	No	No	No
43-51	Yes	Yes	No	No	No
QP	**D > T_D**				
	Skip	**16x16**	**16x8**	**8x16**	**Other**
1-7	Yes	Yes	Yes	Yes	Yes
8-12	Yes	Yes	Yes	Yes	Yes
13-17	Yes	Yes	Yes	Yes	Yes
18-22	Yes	Yes	Yes	Yes	Yes
23-27	Yes	Yes	Yes	Yes	Yes
28-32	Yes	Yes	Yes	Yes	No
33-37	Yes	Yes	Yes	No	No
38-42	Yes	Yes	No	No	No
43-51	Yes	Yes	No	No	No

Table 2: Subsets for inter modes.

3. RESULTS

The proposed algorithm for intra size decision described in Sec.2.1 has been included in JM8.5 reference encoder [10] and simulated on CIF sequences. The tests were performed using different sequences; here we give the results obtained with two sequences: SEQ1 and SEQ2. SEQ1 is the concatenation of Foreman (300 frames), Miss America (150 frames), Akiyo (300 frames), Table tennis (150 frames), Mobile&calendar (300 frames), Claire (168 frames) and Paris (282 frames); SEQ2 is the concatenation of All monitor (300 frames), Container (300 frames), Mother&daughter (300 frames) and News (300 frames). The simulations were performed using concatenated sequences because the intra modes are more spurred in presence of scene changes. The tests were performed using a baseline H.264 configuration (with R-D lagrangian optimization). The *PSNR_Y* average values and the obtained bitrate are reported in Table 3 for SEQ1 and in Table 4 for SEQ2. The PSNR reductions are very slight: the maximum lost is of about 0.61 dB using both intra and inter methods (SEQ 2 at QP=50). Generally the loss is between 0 and 0.30 dB and it is not perceptible. The bitrate increases between 0 and 10%.

QP	PSNR (dB)			
	Original	Fast Intra	Fast Inter	Fast Intra+Inter
10	50.26	50.20	50.26	50.20
20	42.61	42.58	42.61	42.58
30	35.49	35.54	35.31	35.34
40	29.08	29.10	28.86	28.84
50	23.24	23.16	23.14	23.01
QP	Rate (kbit/s)			
	Original	Fast Intra	Fast Inter	Fast Intra+Inter
10	6051	6082	6051	6082
20	1836	1837	1836	1837
30	437	442	452	456
40	80	81	87	89
50	24	24	24	24

Table 3: PSNR and bitrate results for SEQ1 sequence.

QP	PSNR (dB)			
	Original	Fast Intra	Fast Inter	Fast Intra+Inter
10	50.31	50.11	50.31	50.11
20	42.63	42.60	42.63	42.60
30	36.07	36.18	35.98	36.05
40	30.35	30.34	30.18	30.08
50	25.14	24.74	24.98	24.53
QP	Rate (kbit/s)			
	Original	Fast Intra	Fast Inter	Fast Intra+Inter
10	5210	5279	5210	5279
20	945	944	945	944
30	136	139	140	144
40	32	33	34	36
50	10	10	10	10

Table 4: PSNR and bitrate results for SEQ2 sequence.

The total encoding time is pictured in Figure 1 for SEQ1 and in Figure 2 for SEQ2: using both the methods the average time saving ranges from 40% to 70%. It is noticeable that most of the gain is obtained using the Fast Intra mode decision.

An analysis on other fast mode decision algorithms showed that the computational time can be reduced of about 50% for [4], [5] and [6], and up to 84% for [3].

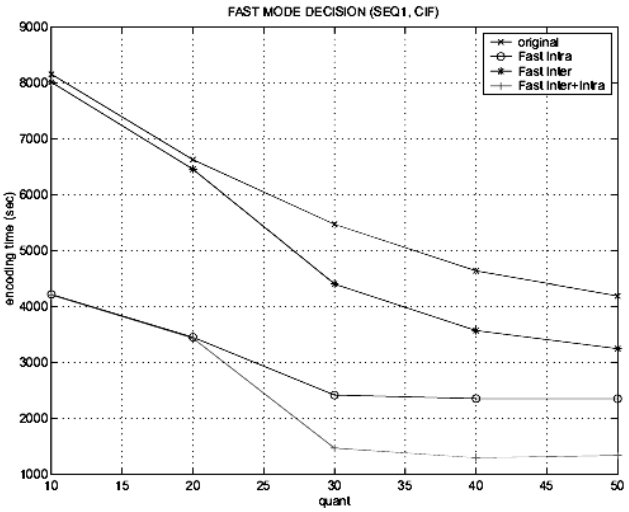

Figure 1: Encoding time saving for the SEQ1 sequence using the proposed methods.

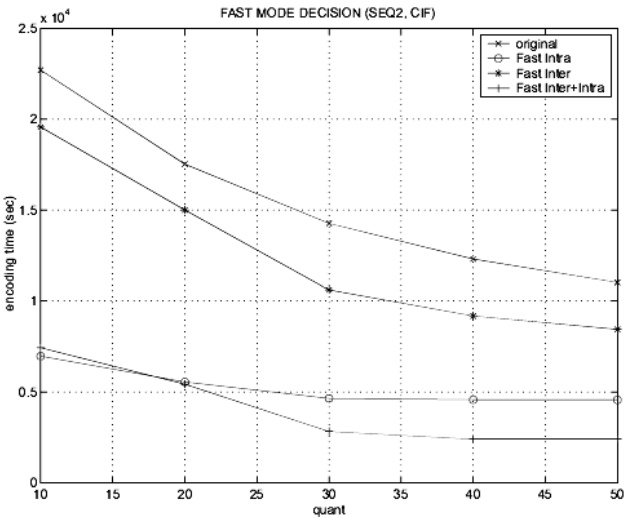

Figure 2: Encoding time saving for the SEQ2 sequence using the proposed methods**.**

4. CONCLUSIONS AND FURTHER RESEARCH

This paper presented two algorithms for reducing the encoding complexity related to intra and inter modes evaluation, in particular when R-D optimization is used. These methods select for each frame a subset of modes that will be evaluated by the coder. The simulation results showed that the proposed methods permit to reduce the complexity of the encoding process with a small drop in PSNR.

These methods can also be used to reduce the complexity of H.264 encoder that does not use the R-D optimization.

Further works are in progress for a fine-tuning of the different proposed tools, focusing on the usage of the inter selection method at macroblock level instead of frame resolution. More simulations are also in progress to test the proposed methods with other resolutions including QCIF and SD.

5. REFERENCES

[1] *ITU-T Rec.H.264, ISO/IEC 14496-10 AVC, Advanced Video Coding*, Final Draft International Standard, JVT-G050r1, Geneva, Switzerland, May 2003.

[2] M. Yang, W. Wang, "Fast Macroblock Mode Selection Based on Motion Content Classification in H.264/AVC," ICIP 2004, Singapore, 2004.

[3] H. Kim, Y. Altunbasak, "Fast Mode Decision for Inter Prediction in H.264," ICIP 2004, Singapore, 2004.

[4] Y.H. Kim, J.W. Yoo, S.W. Lee, J. Shin, "Adaptive mode decision for H.264 encoder," in Electronic Letters, Vol.40 No.19, September, 2004.

[5] A. C. Yu, G. Martin, "Advanced block size selection algorithm for inter frame coding in H.264/MPEG-4 AVC," in ICIP 2004, Singapore, 2004.

[6] Q. Dai, D. Zhu, R. Ding, "Fast mode decision for inter prediction in H.264," in ICIP 2004, Singapore, 2004.

[7] F. Pang, X. Lin, R. Susanto, K.P. Lim, Z. Li, G. Feng, D. Wu, S. Wu, "Fast mode decision for intra prediction," document JVT-G013, 7th meeting, Pattaya, Thailand, 7-14 March 2003.

[8] D. Alfonso, D. Bagni, L. Celetto, L. Pezzoni, "Detailed rate-distortion analysis of H.264 video coding standard and comparison to MPEG-2/4," in Proceedings of SPIE 2003, vol. 5150, p.891-902, 2003.

[9] P.M. Kuhn, T. Suzuki, A. Vetro, "MPEG-7 Transcoding Hints for Reduced Complexity and Improved Quality," Proc. in International Packet Video Workshop, April, 2001.

[10] "http://iphome.hhi.de/suehring/tml/download/"