FAST MODE DECISION FOR H.264 BASED ON RATE-DISTORTION COST ESTIMATION

Yao-Chung Lin

Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA

Torsten Fink, Erwin Bellers

TriMedia Center of Excellence
NXP Semiconductors
San Jose, CA 95131, USA

ABSTRACT

A fast mode decision algorithm for an H.264 encoder is proposed which reduces computation by using a statistical dependency of macroblock rate-distortion (RD) costs. The algorithm skips the motion estimation and/or intra prediction mode decision with an adaptive threshold. The simulation results show that the proposed algorithm achieves an almost 2X speedup with negligible degradation in coding efficiency.

Index Terms— Video coding, H.264/AVC, mode decision

1. INTRODUCTION

The H.264 video coding standard, also known as MPEG-4/ AVC, offers a rich set of macroblock modes (7 partitions of Inter modes, 9 modes for Intra4x4, and 4 modes for Intra16x16) and has superior rate-distortion performance, which has up to 50% bit rate saving compared to previous standards, such as MPEG-1,-2, and H.263. The rate-distortion (RD) optimization technique adopted in the reference H.264 encoder [1] chooses the mode with minimum Lagrangian cost for a macroblock. To achieve the optimal rate-distortion performance, the encoder could check all modes and then select the best macroblock mode. This exhaustive method is not suitable for real-time implementation. Thus, a reduction of complexity is highly desirable.

Many researchers have studied fast motion estimation algorithms to reduce the computation cost, such as [2] and [3] which are adopted in the JM reference software [4]. As a large portion of computation cost is reduced by fast motion estimation algorithms, many studies have started to investigate fast fractional pel search, such as [5] and fast intra prediction mode decision, such as [6] [7], and [8]. For encoding a macroblock in an inter picture, it is well-known that the number of skipped macroblocks increases as the quantization parameter (QP) increases. If a mode decision algorithm can wisely choose macroblocks as the SKIP mode before checking all modes, a large portion of computation saving is obtained. The computation cost of the intra prediction mode decision with fast intra prediction algorithms is comparable to that of the Inter mode decision with fast motion estimation algorithms. Thus, additional saving would be obtained if the encoder skips the intra prediction mode decision. However, little work has been done in SKIP/Inter/Intra mode decision. In [9], the authors categorize a macroblock into three different error-risk regions based on features derived from the residuals of the best matched Inter and Intra predictions, and the motion vectors. The mode of macroblocks in risk-free region can be decided immediately by comparison between the features, while others are decided by minimum-risk criteria or minimum RD costs. In [10], the early SKIP mode decision is checked after applying motion search with 16x16 block size, and the boundary error is used to decide Inter or Intra mode for a macroblock. These methods extract the features by some high complexity procedures, such as motion search. Reliable features with light computation cost or even without touching the current macroblock would attract fast mode decision design.

We find that macroblock RD costs of a given mode have high correlation to those of the temporal and spatial adjacent macroblocks, especially in the previous co-located macroblock. This statistical dependency is adopted in our fast mode decision algorithm and yields an almost 2X speedup. In Section 2, we demonstrate the RD cost correlations and some related observations. Based on these observations, we propose a fast mode decision algorithm presented in Section 3. Simulation setup and results are presented in Section 4. Finally, we draw conclusions in Section 5.

2. RATE-DISTORTION COST CORRELATION

In the H.264 reference encoder, JM[4], the RD optimization is based on a Lagrangian optimization framework. For each mode, m, of a macroblock, the encoder calculates the Lagrangian RD cost by

$$J_m = D_m + \lambda_{mod} R_m, \forall m \in \{\text{available modes}\},$$

where D_m is the distortion measured by sum of squared difference (SSD) between reconstructed and original samples in the macroblock, R_m is the number of total bits for coding this macroblock in mode m, and λ_{mod} is the Lagrangian mul-

1-4244-0728-1/07/$20.00 ©2007 IEEE

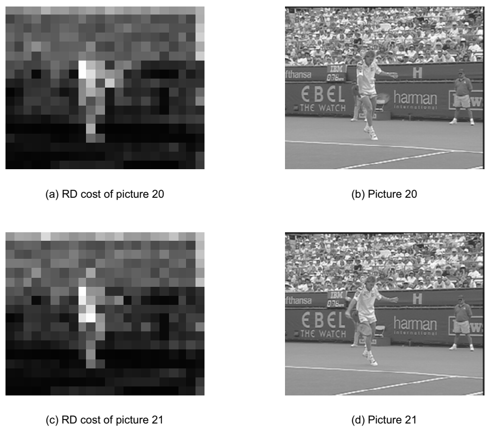

Fig. 1. The 20th and 21st pictures of Stefan sequence in CIF format and the corresponding Inter16X16 RD cost maps

tiplier, which is a function of QP,

$$\lambda_{mod}(QP) = 0.85 \times 2^{(QP-12)/3}$$

The multiplier, λ_{mod}, can be considered as a trade-off parameter between the rate and distortion. In this case, if encoder pays one bit to reduce more than λ_{mod} distortion in SSD, then less Lagrangian cost, J_m, is obtained. The optimal coding efficiency can be achieved by checking all available modes and selecting the minimum cost.

When computation resource is limited, checking all the the modes is rarely possible. Therefore, an additional computation resource constraint should be applied for joint optimization with rate and distortion. However, this leads to a contradiction that the RD costs can not be revealed without coding and coding consumes computations. Therefore, one way to deal with this problem is to estimate the RD costs before coding and allocate computation resource to the modes with high probability of having the minimum RD cost.

Fig. 1 demonstrates two adjacent pictures of the Stefan sequence in CIF format and the corresponding RD cost maps of Inter16x16 mode. In the RD cost maps, white color blocks indicate that the corresponding macroblocks have high RD costs, and darker ones for the lower RD cost macroblocks. As can be seen, the white blocks in the maps are the crowd and the body of Stefan in the original pictures. They are relatively hard to code and cost more bits for motion and residual. As for darker areas, they have flat texture or steady motion in the original pictures and are relatively easy to code. Like image and video signal, we expect that the RD costs have high correlation to those of the temporal and spatial adjacent macroblocks.

Fig. 2 presents the candidate macroblocks for estimating the RD cost of the current macroblock. The spatial candidates are the left, left-top, top, and right-top adjacent macroblocks.

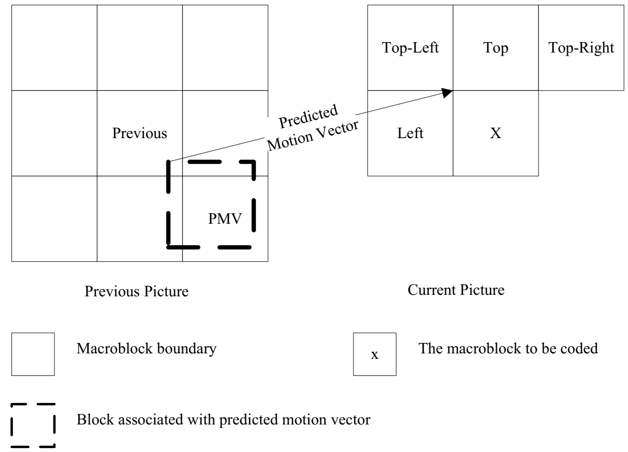

Fig. 2. The candidate macroblocks for estimating current RD cost

As for temporal candidates, we choose the co-located macroblock in the previous picture and the major dominating macroblock associated with the predicted motion vector (PMV). We obtain the correlation coefficients, ρ of two random variables, X and Y, by

$$\rho(X,Y) = \frac{Cov(XY)}{\sqrt{Var(X)Var(Y)}},$$

Table 1 shows the RD cost correlation coefficients of each mode, m, between the current and the candidate macroblocks, i.e. $\rho(Cost_m^{current}, Cost_m^{candidate})$. The data are collected from encoding the first 100 pictures of Bus, Carphone, Foreman, Mobile, and Stephan CIF sequences at 30 fps with QP equal to 28, 32, 36 and 40. The averaged correlation coefficients are presented. The encoder codes the macroblock in all available modes and records their RD costs. Here, PAxB means the Inter mode with partition AxB. The P8x8 is the Inter mode with submacroblock partitions, (8x8, 8x4, 4x8, and 4x4) and the minimum cost among the submacroblock partitions is recorded. The I4x4 and I16x16 are the intra prediction modes, and the minimum cost among the Intra modes is recorded. As for the picture prediction structure, only the first picture is coded as I-picture and remaining pictures are coded as P-picture. As can be seen in Table 1, the RD cost has high correlation, up to 0.97, to that of the previous co-located macroblock for all modes except the SKIP mode. Most test sequences demonstrate a similar relationship. The correlation would be low for sequences with scene change, flashes, and very fast motion. Note that RD cost of the candidate macroblock, PMV, does not have higher correlation than that of the previous co-located macroblock. A possible reason is that the predicted motion vector is not accurate and introduces more RD cost noise which lowers the correlation coefficient.

Table 2 shows the mean penalty in RD costs for the selected mode, m, given the optimal mode, m^*, of a macroblock,

Table 1. RD cost correlation coefficients between the current and candidate macroblocks

	Left-Top	Top	Right-Top	Left	Previous	PMV
SKIP	0.11	0.28	0.11	0.39	0.66	0.60
P16X16	0.39	0.50	0.39	0.62	**0.91**	0.90
P16X8	0.40	0.52	0.40	0.61	**0.91**	0.90
P8X16	0.40	0.50	0.40	0.64	**0.91**	0.90
P8X8	0.42	0.52	0.42	0.64	**0.92**	0.91
I16X16	0.42	0.51	0.46	0.69	**0.97**	0.97
I4X4	0.44	0.51	0.44	0.66	**0.97**	0.97

Table 2. Averaged RD cost penalty

		Selected Mode (m)						
		SKIP	P16X16	P16X8	P8X16	P8X8	I16X16	I4X4
Optimal Mode (m*)	SKIP	0	533	941	959	1585	11298	13406
	P16X16	39661	0	903	927	1494	15581	20409
	P16X8	56326	1825	0	2219	1304	15795	21135
	P8X16	48257	1533	1933	0	1224	15607	21101
	P8X8	112420	3432	2777	2962	0	18424	25537
	I16X16	240070	6041	5674	5391	4395	0	5423
	I4X4	43543	1462	1792	1819	2409	2250	0

i.e. $Penalty(m^*, m) = Cost(m) - Cost(m^*)$. Here, the QP in this setting is 32 which yields the Lagrangian multiplier for mode decision, λ_{mod}, equal to 86.4 for P-picture. The relationship between RD performance and the penalty is described as follows. The penalty, p, will increase p/λ_{mod} bits given the same distortion, SSD, or increase the SSD by p given the same number of bits. In Table 2, if the best mode of one macroblock is SKIP mode, this macroblock would receive extra 6 bits or its PSNR would drop to 34.5 dB from 35.0 dB at the same bit rate when an encoder selects P16x16 to code. The observation in Table 2 shows several intuitions. First, wrongly selecting the coded mode as SKIP mode would receive huge penalty. Second, if the optimal mode is in Intra group but the coded mode is in Inter group, the RD cost would increase greatly, and vice versa. Finally, less penalty is received while selecting wrong modes within the same group (Inter or Intra) as the optimal mode.

In summary, we observe that macroblock RD costs of a mode is highly correlated to that of the previous co-located macroblock, except the SKIP mode. This statistical dependency is used in our proposed mode decision algorithm. The observation in RD cost penalty suggests that the mode decision algorithm should carefully make two decisions, SKIP/Inter, and Inter/Intra. The mode decision within Inter or Intra group has been studied in several literatures aforementioned and is not within the scope of this paper.

3. PROPOSED FAST MODE DECISION ALGORITHM

Based on the statistical dependency described in the previous section, we propose a fast algorithm to deal with the SKIP/Inter and Inter/Intra mode decisions. The flow chart of the proposed mode decision algorithm is presented in Fig. 3. For each macroblock, i, the encoder first evaluates the RD cost of the SKIP mode, $J_{SKIP}(i)$. If $J_{SKIP}(i)$ is greater than the threshold, $\hat{J}_{Inter}(i) - \alpha\sigma_{Inter}$, the encoder uses fast motion estimation algorithm to check all Inter modes and get the minimum RD cost of Inter modes, $J_{Inter}(i)$, otherwise, the macroblock is coded as skip mode. Here, $\hat{J}_{Inter}(i)$ and σ_{Inter} denote the predicted Inter mode RD cost of macroblock i and standard deviation of the prediction error, respectively. The parameter, α, can be decreased for speedup or increased for better coding efficiency. The predictor, $\hat{J}_{Inter}(i)$, is updated as the checked RD cost, $J_{Inter}(i)$, and the prediction error standard deviation, σ_{Inter}, will be updated after collecting a certain number of prediction errors, $e_{Inter}(i) = \hat{J}_{Inter}(i) - J_{Inter}(i)$. A similar procedure is applied to make the Inter/Intra decision. After obtaining the minimum RD cost of the Inter modes, $J_{Inter}(i)$, the encoder compares it with the threshold, $\hat{J}_{Intra}(i) + \beta\sigma_{Intra}$. When $J_{Inter}(i)$ is greater than the threshold, the encoder invokes intra prediction mode decision to get the minimum RD cost of Intra modes, $J_{Intra}(i)$ and updates the RD cost predictor, $J_{Intra}(i)$, and the prediction error standard deviation, σ_{Intra}, are updated. Finally, the macroblock is coded as the one of the checked modes with the minimum RD cost.

As for the initial value of prediction error variance, we find that the variances of the RD cost prediction error are highly related to the QP. The variance is exponentially increasing as QP increases, and the variance-QP curve in log-scale has the similar slope as the Lagrangian multiplier, $\lambda_{mod}(QP)$. Based on this observation, we empirically set the variances, σ^2_{Inter} and σ^2_{Intra}, as $10^5 \times \lambda_{mod}(QP)$. The standard deviations of the RD prediction error are derived from the variances which will be updated after collecting a certain number of prediction errors.

4. SIMULATION RESULTS

The proposed mode decision algorithm is implemented in JM 11.0 encoder. A default decision algorithm in the JM serves as the reference algorithm for comparison. The test common settings are described as follows. The fast motion estimation algorithm, EZPS, is used in all experiments. The motion estimation is conducted over a search range of ± 32 pels with quarter pel precision. One reference picture is used for motion search. The GOP structure is IPPP, i.e. the first picture is coded as I-picture and remaining pictures are coded as P-pictures. All Inter and Intra modes are turned on. Our proposed algorithm is compared to the default algorithm in JM11.0 including early skip detection and selective intra prediction mode decision as specified in [10]. The two parameters in our algorithm, α and β, are set to one. We encode the first 150 frames of CIF sequences with four quantization parameters, 28, 32, 36, and 40. The calculation of the difference between two RD curves is specified in [11]. The difference is

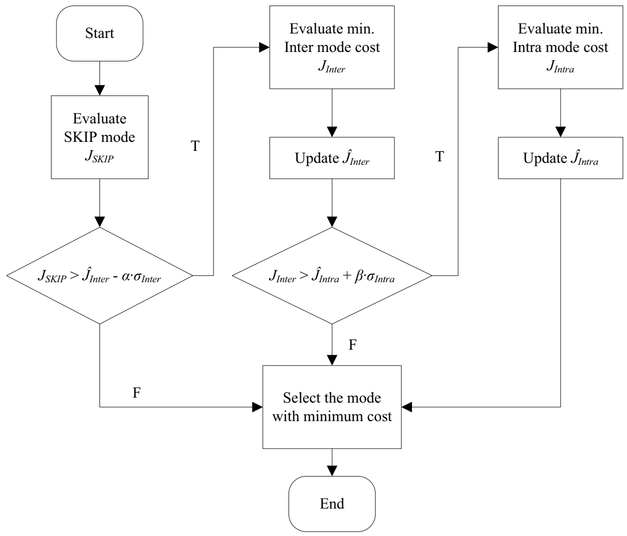

Fig. 3. Flow chart of the proposed mode decision algorithm

Table 3. Simulation Results

Sequence	BD PSNR (dB)	BD Bit Rate (%)	Time Saving (%)
Akiyo	0.02	-0.43	33.38
Bus	-0.06	1.36	35.76
CarPhone	-0.10	2.45	44.89
FootBall	-0.21	4.61	32.99
Foreman	-0.09	2.11	42.88
Mother&Daughter	-0.03	0.65	45.30
Mobile	-0.03	0.59	49.47
News	-0.02	0.43	57.57
Stefan	-0.07	1.56	51.93
Average	**-0.07**	**1.48**	**43.80**

represented as Bjontegaard Delta (BD) PSNR or BD bit rate.

Table 3 shows the rate distortion difference and the time saving between our proposed algorithm and the reference algorithm. The time saving is calculated by

$$\triangle Time(\%) = \frac{Time_{Reference} - Time_{Proposed}}{Time_{Reference}} \times 100\%, \quad (1)$$

where $Time_{Proposed}$ and $Time_{Reference}$ are encoding times for the proposed and reference algorithms respectively. As can be seen in Table 3, our algorithm yields 43% time saving on average compared to the default mode decision algorithm with negligible quality (-0.07dB) or bit rate (+1.41%) loss. For fast motion video sequences, such as Football, the algorithm still has 26% to 41% time saving but with 0.2 dB quality loss or 4.6% bit rate loss. For steady motion video sequence, such as News, the algorithm achieves up to 57% time saving with negligible degradation of rate and distortion loss due to more macroblocks selected as SKIP mode. Note that the proposed algorithm achieves only 33% time saving in Akiyo sequence. We believe that the reference algorithm has already obtained significant time saving.

5. CONCLUSION

We find that the RD costs for the macroblock mode decision in the H.264 encoder have high correlation to that of the previous co-located macroblock. A fast mode decision algorithm based on this statistical dependency is proposed. By wisely skipping a number of motion estimations and intra prediction mode decisions, the algorithm achieves on average 43% time saving with negligible quality and rate loss in the simulation. For future work, more sophisticatedly using the dependency, such as using spatial correlations, can achieve better RD cost prediction and yield greater time saving and less coding efficiency degradation. This statistical dependency can be used for joint rate-distortion-computation optimization.

ACKNOWLEDGEMENT

The authors would like to thank Professor Bernd Girod, Mr. Manoj Koul, and Arno Glim for helpful discussions.

6. REFERENCES

[1] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 11, November 1998.

[2] Zhibo Chen, Peng Zhou, and Yun He, "Fast integer pel and fractional pel motion estimation for JVT," JVT-F017, 6th JVT meeting: Awaji, Island, JP, December 2002.

[3] Alexis Michael Tourapis, Hye-Yeon Cheong, and Pankaj Topiwala, "Fast ME in the JM reference software," JVT-P026, 16th JVT meeting: Poznan, PL, July 2005.

[4] Joint Video Team, "Reference software JM11.0," http://iphome.hhi.de/seuhring/tml/download.

[5] Zhibo Chen, Cheng Du, Jinghua Wang, and Yung He, "PPFPS-A paraboloid prediction based fractional pixel search strategy for H.26L," in *International Symposium on Circuits and Systems (ISCAS)*. IEEE, May 2002.

[6] Feng Pan, Xiao Lin, Rahardja Susanto, Keng Pang Lim, Zheng Guo Li, Ge Nan Feng, Da Jun Wu, and Si Wu, "Fast mode decision for intra prediction," JVT-G013, 7th JVT meeting: Pattaya II, Tailand, March 2003.

[7] Yu-Wen Huang, Bing-Yu Hsieh, Tung-Chien Chen, and Lian-Gee Chen, "Analysis, fast algorithm, and VLSI architecture design for H.264/AVC intra frame coder," *IEEE Transcation on Circuit and systems for video technology*, vol. 15, no. 3, march 2005.

[8] Takeshi Tsukuba, Isao Nagayoshi, Tsuyoshi Hanamura, and Hideyoshi Tominaga, "H.264 fast intra-prediction mode decision based on frequency characteristic," in 13^{th} *European Signal Processing Conference*. EURASIP, September 2005.

[9] Changsung Kim and C.-C. Jay Kuo, "A feature-based approach to fast h.264 intra/inter mode decision," in *ISCAS*. IEEE, 2005.

[10] Byeungwoo Jeon and Jeyun Lee, "Fast mode decision for h.264," JVT-J033, 10th JVT meeting: Waikoloa, Hawaii, USA, December 2003.

[11] Gisle Bjontegaard, "Calculation of average psnr differences between rd-curves," VCEG-M033, 13th VCEG meeting: Austin Texas, USA, April 2001.