FAST INTER-MODE DECISION AND SELECTIVE QUARTER-PEL REFINEMENT IN H.264 VIDEO CODING

Hasan F. Ates

Dept. of Electronics Engineering
Isik University
Sile, Istanbul, 34980 TURKEY
hfates@isikun.edu.tr

ABSTRACT

In H.264 video coding standard, there exist several inter - prediction modes that use macroblock partitions with variable block sizes. Choosing a rate-distortion optimal coding mode for each macroblock is essential for the best possible coding performance, but also prohibitive due to the heavy computational complexity associated with the required rate-distortion calculations. Likewise, sub-pel motion refinement improves the coding efficiency, but becomes a major computational bottleneck when integer-pel search is executed fast. In this paper, we present a simple strategy to reduce the complexity of quarter-pel refinement and inter-mode decision with minimum loss of coding efficiency. Based on the results of the half-pel motion estimation step, our method evaluates the likelihood of each inter-coding mode being optimal. Then, quarter-pel refinement and actual rate and distortion are computed for only those coding modes with sufficient chance of being optimal. We claim that this method minimizes optimal mode estimation error at a given level of refinement and mode decision complexity. Simulation results show that the algorithm speeds up quarter-pel search and inter-mode selection modules by a factor of about 6 with less than 0.12 dB PSNR loss.

Index Terms— mode decision, quarter-pel refinement, H.264 video coding.

1. INTRODUCTION

H.264 coding standard offers a rich set of coding modes to choose from when coding macroblocks (MBs) of a video frame [1]. These modes allow the encoder to try different MB partitions, multiple reference frames and inter/intra prediction methods in order to find a rate-distortion (R-D) optimal coding strategy for each MB. Unfortunately, making an optimal choice among all possible coding modes requires a substantial amount of computation to be performed for accurate rate and distortion calculations.

In H.264, the optimal motion vectors (MVs) for each partition can be refined down to quarter pixel accuracy. It has been shown that sub-pel refinement brings significant improvement in terms of coding efficiency. However, it is computationally expensive to find the optimal quarter-pel accurate MVs for each partition. When a fast method is used for integer-pel search, the complexity of sub-pel refinement step becomes a major concern for achieving real-time encoding.

In inter-mode decision (inter-MD), the encoder decides which of the possible MB partitions is optimal for motion compensated coding of the MB. The standard allows each MB (16×16 pixel) to be divided into two 16×8 or two 8×16 or four 8×8 subblocks. Each 8×8 block can be further partitioned into 8×4, 4×8 and 4×4 subblocks. Inter-MD is typically performed after integer-pel and sub-pel motion estimation (ME), when the optimal quarter-pixel accurate MVs of each mode become available.

A simple strategy to reduce the computational cost of sub-pel refinement is to perform mode decision (MD) immediately after integer-pel ME is over, and to execute sub-pel refinement only for the mode that is selected to be optimal. Simulations show that this method incurs significant coding loss, unless at least half-pel accurate MVs are available. In other words, half-pel refinement is essential for optimal mode selection, but performing quarter-pel refinement after MD has comparably less effect on coding performance.

Fast inter/intra MD has been extensively studied in literature. In this paper, we focus on inter-MD, but our method could easily be extended to intra-mode decision as well. For inter-MD, existing methods either use the results of partial ME and MD (such as sum of absolute (transformed) differences (SA(T)D), R-D costs of a subset of modes [2] or the variance/distribution of motion vectors [3], etc.), or propose computationally less expensive ways of estimating the rate and distortion of each coding mode [4]. Combined with fast integer-pel and sub-pel ME, these methods provide significant reduction in overall encoder complexity.

In this paper, we propose to estimate the optimal inter-mode based on the minimum SATD-based costs that are computed during half-pel ME. However, instead of making a hard decision about the optimal mode, we evaluate the likelihood of each mode being optimal based on the calculated SATDs. Modes with "sufficient" chance of being optimal are further compared after quarter-pel refinement and R-D cost calculation. Based on how this sufficiency is defined, the complexity of inter-MD module can be changed from almost no computation to the full complexity of R-D optimal decision.

Aside from this flexibility in the approach, the algorithm provides a mechanism to distribute the overall complexity of quarter-pel refinement and inter-MD over MBs of a frame with as little loss of coding efficiency as possible. We claim that, when likelihood functions of inter-modes are accurately modeled, the proposed method minimizes total mode estimation error for a given level of encoder complexity. In other words, under the given modeling assumptions, the algorithm provides an optimal trade-off between encoder complexity and coding efficiency.

Section 2 describes the mode selection problem and the proposed approach. Section 3 explains the likelihood models used for mode selection and lays out the details of the algorithm. Section 4 compares the performance of the algorithm with the reference soft-

This research was supported by the TÜBİTAK Grant 104E125.

ware when rate-distortion optimization is on and off. The simulation results in Section 4 show that there is less than 0.12 dB PSNR loss, despite a major reduction in complexity. We conclude the paper in Section 5.

2. COMPLEXITY CONSTRAINED INTER-MODE DECISION

In order to achieve the superior coding performance promised by the H.264 standard, it is essential to choose an optimal coding mode for each MB of a given video sequence. This choice specifies whether to use intra or inter prediction for the MB. If inter-prediction is preferred, optimal partitioning, optimal reference frame and optimal MVs should also be determined. For R-D optimized decisions, all possible alternatives should be compared based on their actual rate-distortion costs:

$$\mathcal{J}_M = D + \lambda_M R, \tag{1}$$

where λ_M is an appropriate Lagrange multiplier used for MD. In practice, this is computationally not feasible. Instead, for each inter-mode, the optimal MVs and reference frame are determined by minimizing the following cost function:

$$\mathcal{J}_E = SA(T)D + \lambda_E R_{MV}, \tag{2}$$

where $SA(T)D$ stands for sum of absolute (Hadamard transformed) difference between the current MB and the reference frame, R_{MV} is the bitrate for coding the MVs and the reference frame index, and $\lambda_E = \sqrt{\lambda_M}$.

Inter-MD could also be performed by choosing the mode that has minimum $\mathcal{J}_E$ cost, which eliminates the computational burden of calculating $\mathcal{J}_M$. However, this could result in significant loss of coding efficiency that is not acceptable for high quality applications.

In this paper, we propose to use $\mathcal{J}_E$ costs to evaluate the likelihood of each mode being optimal, and compute the actual $\mathcal{J}_M$ for only those modes with sufficient chance of being optimal. $\mathcal{J}_E$ costs are collected from the half-pel ME step. If a mode is unlikely to be optimal, quarter-pel ME is skipped as well. In this manner, it is possible to reduce the overall complexity of quarter-pel refinement and inter-MD without compromising the coding performance. The rest of this section describes how the relative magnitudes of $\mathcal{J}_E$ could be used to perform accurate inter-MD with limited total complexity.

As mentioned above, the ultimate goal is to reduce computational complexity by eliminating unlikely modes from further processing. While doing this, we would like to minimize the total mode estimation error. This could be formulated as the following constrained optimization problem:

$$\text{minimize} \quad \sum_{n=1}^{N} P_n \quad \text{such that} \quad \sum_{n=1}^{N} C_n \leq CB, \tag{3}$$

where N is the number of MBs in a frame or in full video, and P_n, C_n are the mode misestimation probability and computational complexity of MB n, respectively. CB is the available budget for quarter-pel ME and inter-MD complexity. A similar formulation has been used before in [5, 6] to speed up integer-pel ME as well.

A mode error occurs when the optimal mode is eliminated due to its $\mathcal{J}_E$ cost. Therefore, the total mode estimation error depends on how $\mathcal{J}_E$ costs are used to evaluate the likelihood of each mode.

In the following formulation, imagine that seven different partitions are represented by the modes, $m_1 = 16{\times}16$, $m_2 = 16{\times}8$, $m_3 = 8{\times}16$, $m_4 = 8{\times}8$, $m_5 = 8{\times}4$, $m_6 = 4{\times}8$, $m_7 = 4{\times}4$. We define an additional mode, m_0, to represent that the MB is divided into 8×8 subblocks and each subblock can be of mode m_j where $4 \leq j \leq 7$. That is;

$$\hat{m} = m_0 \Rightarrow \hat{m}^k = m_j,\ \exists j \geq 4,\ 1 \leq k \leq 4, \tag{4}$$

where $\hat{m}$ represents the optimal coding mode of the MB, and $\hat{m}^k$ represents the optimal coding mode of the 8×8 subblock k.

To evaluate the likelihood of optimality for each mode, we compute the difference between the mode's $\mathcal{J}_E$ cost and the minimum cost:

$$\delta\mathcal{J}_i = \mathcal{J}_E^i - \min_{0 \leq j \leq 3}(\mathcal{J}_E^j), \quad 0 \leq i \leq 3, \tag{5}$$

Likewise, for each 8×8 subblock, a similar measure is computed using subblock $\mathcal{J}_E$ costs:

$$\delta\mathcal{J}_i = \mathcal{J}_E^i - \min_{4 \leq j \leq 7}(\mathcal{J}_E^j), \quad 4 \leq i \leq 7. \tag{6}$$

In this formulation, the cost of mode m_0 is equal to the sum of the minimum costs of four subblocks:

$$\mathcal{J}_E^0 = \sum_{k=1}^{4} \min_{4 \leq j \leq 7}(\mathcal{J}_E^{j,k}). \tag{7}$$

There is a strong correlation between the magnitude of $\delta\mathcal{J}_i$ and the likelihood of mode m_i being optimal. For the mode with minimum $\mathcal{J}_E$, this value is equal to zero; this mode is automatically qualified as a candidate for being the optimal choice. For other modes, $\delta\mathcal{J}_i \geq 0$ and the likelihood of being optimal drops as $\delta\mathcal{J}_i$ increases (except for m_0; see next section).

The next section describes how $\delta\mathcal{J}_i$ could be used to evaluate the likelihood of optimality for m_i, and how this formulation provides a simple solution to the optimization problem of Equation 3.

3. LIKELIHOOD FUNCTIONS AND ALGORITHM FLOW

Given the value $\delta\mathcal{J}_i$ and the mode with minimum cost, we define the conditional probability of each mode m_i being optimal as follows. For $0 \leq i \leq 3$:

$$L_j^i(\delta\mathcal{J}_i) = P(\hat{m} = m_i \mid \delta\mathcal{J}_i, \tilde{m} = m_j), \tag{8}$$

where $\tilde{m} = m_j$ indicates $j = \arg\min_{0 \leq l \leq 3}(\mathcal{J}_E^l)$. A similar definition applies to 8×8 subblock modes m_i, $4 \leq i \leq 7$.

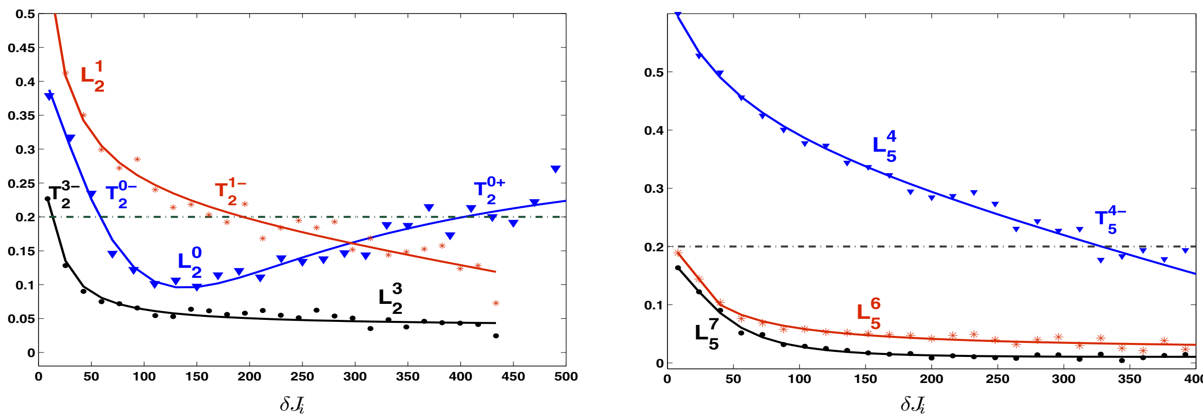
$L_j^i(\delta)$ gives the likelihood of optimality for m_i as a function of δ when m_j has the lowest $\mathcal{J}_E$ cost. Now, for a given MB, the mode misestimation probability P is equal to the sum of $L_j^i(\delta\mathcal{J}_i)$ over all modes m_i which are eliminated from inter-MD process:

$$P = \sum_{i=0}^{3} (1 - I_i) L_j^i(\delta\mathcal{J}_i) + I_0 \frac{1}{4} \sum_{k=1}^{4} P_0^k, \tag{9}$$

where $I_i = 1$ indicates that mode m_i is qualified for further evaluation, and $I_i = 0$ means m_i is not considered to be the optimal inter-mode. P_0^k is the subblock mode error probability:

$$P_0 = \sum_{i=4}^{7} (1 - I_i) L_j^i(\delta\mathcal{J}_i). \tag{10}$$

Note that, there are two sources of error for mode m_0: either m_0 is totally skipped, or it is further considered for optimality but the optimal partition of one or more of the subblocks is skipped. $1/4$ factor controls the contribution of subblock errors to the overall MB



(a) (b)

Fig. 1. Likelihood functions: (a) modes m_0, m_1, m_3 (when $\tilde{m} = m_2$); (b) modes m_4, m_6, m_7 (when $\tilde{m}^k = m_5$)

error; when the modes of all four subblocks are misestimated, that counts as one MB error.

The complexity of the MB can be expressed in a similar form:

$$C = \sum_{i=1}^{3} I_i C_m + I_0 \sum_{k=1}^{4} \sum_{i=4}^{7} I_i^k \frac{C_m}{4}, \qquad (11)$$

where all partitions are assumed to have equal complexity that is C_m. Simulations reveal that this assumption is fairly accurate.

As in [5], it is not hard to show that the total mode estimation error is minimized when modes are selected based on the magnitude of the likelihood function ($0 \le i \le 7$):

$$I_i = \begin{cases} 1, & \text{if} \quad L_j^i(\delta\mathcal{J}_i) \ge \alpha \\ 0, & \text{else} \end{cases} \qquad (12)$$

The optimal solution for I_0 is actually more involved than others. But, the performance of the algorithm is not much affected by the use of this general expression for I_0 as well. This solution is actually analogous to the well-known water-filling solution in optimization theory. The threshold α is chosen such that the overall complexity constraint, $\sum_{n=1}^{N} C_n \le CB$, is satisfied.

To sum up, the inter-mode selection algorithm works as follows:

1. For each mode, perform integer-pel and half-pel ME, and calculate $\mathcal{J}_E^j$, $0 \le i \le 7$.
2. Find the mode with minimum cost, i.e. $\tilde{m} = m_j$. Then, for $0 \le i \le 3$, compute $L_j^i(\delta\mathcal{J}_i)$ and I_i.
3. If $I_0 = 1$, then for each subblock k, $1 \le k \le 4$:
   (a) Find the mode with minimum cost, i.e. $\tilde{m}^k = m_j$. Then, for $4 \le i \le 7$, compute $L_j^i(\delta\mathcal{J}_i)$ and I_i.
   (b) For $4 \le i \le 7$, if $I_i = 1$, then proceed as usual: perform quarter-pel refinement and compute $\mathcal{J}_M^i$.
   (c) Based on $\mathcal{J}_M^i$, choose optimal subblock mode $\hat{m}^k$.
4. If $I_0 = 1$, $\mathcal{J}_M^0 = \sum_{k=1}^{4} \min_{4 \le j \le 7}(\mathcal{J}_M^{j,k})$.
5. For $1 \le i \le 3$, if $I_i = 1$, then proceed as usual: perform quarter-pel refinement and compute $\mathcal{J}_M^i$.
6. Based on $\mathcal{J}_M^i$, choose optimal MB mode $\hat{m}$.

Table 1. Thresholds for modes m_i, $0 \le i \le 3$ (QP=32).

$\tilde{m}$	T^{0-}	T^{0+}	T^{1-}	T^{2-}	T^{3-}
m_0	∞	∞	180	47	47
m_1	50	∞	∞	0	0
m_2	58	403	197	∞	11
m_3	54	336	215	8	∞

Table 2. Thresholds for modes m_i, $4 \le i \le 7$ (QP=32).

$\tilde{m}^k$	T^{4-}	T^{5-}	T^{6-}	T^{7-}
m_4	∞	14	9	0
m_5	320	∞	0	0
m_6	366	10	∞	0
m_7	500	65	55	∞

For the algorithm, the conditional likelihood functions, $L_j^i(\delta\mathcal{J}_i)$, should be derived. We use a training set of video sequences coded at various compression levels to compute the likelihood functions. Simulations show that quantization parameter QP affects the probability of different partitions being optimal, and hence the likelihood functions. Therefore, for each QP, we derive a different set of likelihood functions.

Figure 1(a) and (b) show the derived distributions for QP=32, when $\tilde{m} = m_2$ and $\tilde{m}^k = m_5$ respectively. In these figures, rational functions (having second order numerators and denominators) are fitted through the empirical functions. For $\alpha = 0.2$, the points in each curve that fall below the constant line correspond to $\delta\mathcal{J}_i$ values for which mode m_i will be eliminated from mode selection. In other words, for $0 \le i \le 7$:

$$I_i = \begin{cases} 0, & \text{if} \quad T_j^{i-} < \delta\mathcal{J}_i < T_j^{i+} \\ 1, & \text{else} \end{cases} \qquad (13)$$

where,

$$L_j^i(T_j^{i-}) = L_j^i(T_j^{i+}) = \alpha. \qquad (14)$$

Table 1 gives the defined threshold values for modes m_i, $0 \le i \le 3$, and Table 2 gives the threshold values for subblock modes, $4 \le i \le 7$, when QP=32. Note that $T_j^{i-} = 0$ means the likelihood function is always below $\alpha = 0.2$, in which case the corresponding mode is

Table 3. Performance comparison with R-D optimized MD.

carphone	δPSNR (dB)	δbitrate(%)	Speed-up
Likely-MD	-0.09	+1.83	5.7
$\mathcal{J}_E^i$-based	-0.25	+5.05	9.0
foreman	δPSNR (dB)	δbitrate(%)	Speed-up
Likely-MD	-0.12	+3.08	6.1
$\mathcal{J}_E^i$-based	-0.24	+5.81	9.5
tennis	δPSNR (dB)	δbitrate(%)	Speed-up
Likely-MD	-0.11	+3.20	6.1
$\mathcal{J}_E^i$-based	-0.20	+5.20	8.6

always skipped. $T_j^{i-} = \infty$ means the likelihood function is always above $\alpha = 0.2$, in which case the corresponding mode is always further processed.

Figure 1 implies that the likelihood functions are typically monotonically decreasing, except for mode m_0. When $\delta\mathcal{J}_0$ gets very high, it usually means all $\mathcal{J}_E$ costs are high and motion compensated prediction is poor for all modes. In that case, it is likely for m_0 mode prediction to improve substantially with quarter-pel refinement and m_0 to become the optimal mode. That's why the likelihood of m_0 tends to slightly increase as $\delta\mathcal{J}_0$ gets higher.

4. SIMULATION RESULTS

The simulations are performed for video sequences *carphone* (QCIF), *foreman* (CIF), *tennis* (SIF) at 30 fps, (search range $[-16, 16]$). All frames except the first one are coded as P-frames. Two reference frames are allowed. The CAVLC entropy coder is used, with quantization parameter values of $QP = 24, 28, 32, 36$.

Simulations are carried out for $\alpha = 0.2$. The algorithm is incorporated into JM software version 8.2, and used together with fast integer-pel search method UMHexagonS [7]. The results in Table 3 compare the performance of our approach (Likely-MD) with $\mathcal{J}_E^i$ based MD (in which the optimal mode is chosen as the one that minimizes $\mathcal{J}_E^i$ cost). The table gives average PSNR loss in dB (at equal bitrates) and percentage change in bitrate (at equal PSNR) with respect to R-D optimal MD (based on $\mathcal{J}_M^i$). The last column shows the total speed-up factor of quarter-pel and inter-MD modules with respect to the full R-D optimal decision.

Compared to the R-D optimized MD, Likely-MD achieves a speed-up factor of about 6 with less than 0.12 dB PSNR loss. Compared to $\mathcal{J}_E^i$-based decision, the PSNR loss is halved with around 30-35% reduction in speed-up factor. Figure 2 confirms, for *carphone*, that PSNR loss is indeed reduced by almost 50% at different bitrates and QP values. Hence, the algorithm provides an efficient compromise between the two extremes while not sacrificing too much from both coding efficiency and execution speed. By changing α, the nature of this compromise could also be changed.

It is a limiting factor on the performance to adopt a fixed set of likelihood functions for all video sequences at equal QP. The probability of different partitions change especially when the level of detail and motion content in a frame change. Since the training set and the testing sequences consist of video sequences with moderate-to-high motion content, this limitation doesn't have a significant effect on the performance. Nevertheless, in the future, we plan to design an adaptive scheme that updates likelihood functions based on the content of a given frame.

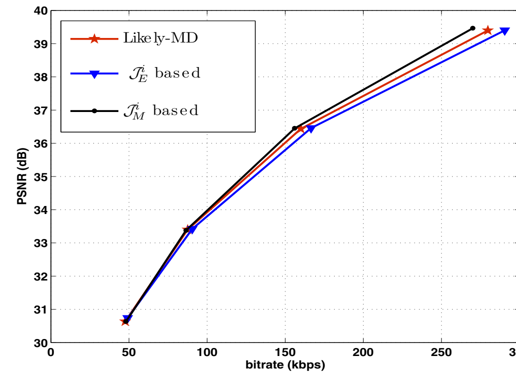


Fig. 2. PSNR vs. bitrate for *carphone*.

5. CONCLUSION AND FUTURE WORK

In this paper, we develop a probabilistic framework to evaluate the likelihood of optimality for each inter-mode, based on the results of half-pel ME step. The proposed method provides an optimal trade-off between total mode estimation error and inter-MD complexity.

The selective processing approach developed in this paper is also applicable to various other parts of the mode decision process, such as half-pel ME, reference frame selection or intra-mode decision. In this manner, overall encoder complexity could be significantly reduced with an adjustable amount of coding loss.

6. REFERENCES

[1] I.G. Richardson, *H.264 and MPEG-4 Video Compression*, Wiley, West Sussex, England, 2003.

[2] P. Yin, H. Cheong, A.M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sept. 2003, vol. 3, pp. 14–17.

[3] Y. Shen, D. Zhang, C. Huang, and J. Li, "Fast mode selection based on texture analysis and local motion activity in H.264/JVT," in *Proc. IEEE Int. Conf. Comm. Circuits Systems*, June 2004, vol. 1, pp. 539–42.

[4] M.G. Sarwer and L.-M. Po, "Fast bit rate estimation for mode decision of H.264/AVC," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17, no. 10, pp. 1402–7, Oct 2007.

[5] H.F. Ates and Y. Altunbasak, "Rate-distortion and complexity optimized motion estimation for H.264 video coding," *accepted for publication in IEEE Trans. Circuit Syst. Video Technol.*, 2007.

[6] H.F. Ates, B. Kanberoglu, and Y. Altunbasak, "Rate-distortion and complexity joint optimization for fast motion estimation in H.264 video coding," in *Proc. IEEE Int. Conf. Image Processing*, Atlanta, Oct. 2006.

[7] Z. Chen, P. Zhou, and Y. He, "Fast motion estimation for JVT," in *JVT-G016.doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG*, 7th Meeting: Pattaya II, Thailand, 7-14 March 2003.