FEATURE-BASED INTRA-PREDICTION MODE DECISION FOR H.264

Changsung Kim,[†] *Hsuan-Huei Shih*[‡] *and C.-C. Jay Kuo*[†]

† Integrated Media Systems Center and Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564
‡ ALi Microelectonics Corp., USA, 8105 Irvine Center Dr. #550, Irvine, CA 92618

ABSTRACT

A fast mode decision method for Intra prediction in H.264 is proposed in this work to reduce the encoder complexity. The proposed algorithm adopts a multi-stage sequential mode decision process that uses joint spatial and transform domain features to filter out unlikely candidate modes and, in the final stage, a simplified rate-distortion optimization method with a rate-distortion model is applied to a very limited number of candidates. It incorporates an early termination mechanism to save computations. Experimental results are given to demonstrate that the proposed scheme achieves a speed-up factor of 10-15 as compared with the current RD optimized mode decision with little quality degradation.

1. INTRODUCTION

The emerging H.264 coding standard incorporates many state-of-the-art techniques to achieve outstanding coding performance. The enhanced Intra prediction technique is one of the important factors that contribute to the success of H.264. This feature also makes H.264 Intra-frame coding better than the JPEG2000 still image compression standard [1]. To achieve high coding efficiency, the Intra prediction in H.264 employs the rate-distortion optimization (RDO) technique [2]. However, the computational complexity of the H.264 encoder is dramatically increased by this optimization technique so that it is difficult to implement the RDO in real-time applications such as video telephony and video conferencing. Only a few attempts have been made to address this complexity bottleneck of the Intra mode decision. Pan *et. al.* [3] and Jeon and Lee [4] proposed a fast mode decision scheme. However, the encoding speed is only approximately 30% faster than that of the RDO method. Thus, there is obviously a huge gap between the desired encoding speed and the actual one reached in [3] and [4].

In this work, we present a simple yet effective fast mode decision algorithm for H.264 Intra prediction. Based on the multi-stage mode decision concept, the proposed algorithm computes low cost features and checks whether the decision process should proceed to the next step or can be terminated earlier with the most probable mode at each stage. The proposed mode decision scheme is integrated with the H.264 JM7.3a codec for the performance test. It is compared with the RDO mode decision of H.264 in terms of the computational cost, the average PSNR and the coding bit-rate for sequences recommended in [5]. A speed-up factor of 10-14 is observed for the proposed Intra-prediction technique at the expense of little quality degradation.

2. INTRA PREDICTION FOR H.264

The H.264 standard exploits the spatial correlation between adjacent macroblocks/blocks for Intra prediction. That is, the current macroblock/blocks is predicted by adjacent pixels in the upper and the left macroblocks/blocks that are decoded earlier. H.264 offers a rich set of prediction patterns for Intra prediction, *i.e.* nine prediction modes for 4×4 luma blocks and four prediction modes for 16×16 luma blocks. Each mode has its own direction of prediction and the predicted samples are obtained from a weighted average of decoded values of neighborhood macroblocks/blocks [2]. To take the full advantage of these modes, the H.264 encoder can select the best mode using the rate distortion optimization (RDO).

The RDO mode decision exhaustively searches the best mode for each 4×4 block which produce the minimum rate-distortion cost given by

$$J(s, c, mode|QP, \lambda_{mode}) = SSD(s, c, mode|QP) + \lambda_{mode} \cdot R(s, c, mode|QP), \quad (1)$$

where QP is the macroblock quantization parameter, the Lagrangian multiplier λ_{mode} is selected to be $0.85 \cdot 2^{QP/3}$, $SSD(\cdot)$ means the sum of the squared differences between the original 4×4 luminance block denoted by s and its reconstruction c, and $R(\cdot)$ represents the number of bits associated with the chosen mode. As a result, the complexity of the Intra-mode decision is extremely high. To reduce the encoding complexity with little RD performance degradation, a fast Intra-mode decision method is proposed in the following section.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

3. PROPOSED FAST INTRA-MODE DECISION

In this section, we attempt to address the following two questions. First, what are the low-cost features that can be utilized to obtain the sub-optimal or the optimal mode in the R-D sense? Second, how can we use these features effectively in the decision process?

3.1. Feature-based Mode Filtering

To address the feature selection problem, both spatial and frequency domain features are selected. Intuitively speaking, a good prediction should produce a small value of the sum of absolute differences (SAD), which can be written as

$$SAD = \sum_{(x,y)\in b_k} |D(x,y) = I(x,y) - P(x,y)|, \qquad (2)$$

where I and P represent the current block and its prediction, respectively. With the Parseval theorem, the total energy (or the 2-norm) of the difference in the space domain and the transform domain should be the same. Thus, it is reasonable to say that a good prediction should also produce a small value of the sum of the absolute transform coefficent differences (SATD), which is defined by

$$SATD = \sum_{(x,y)\in b_k} |T\{D(x,y)\}|, \qquad (3)$$

where T is a certain 2D orthonormal transform. For computational simplicity, T is chosen to be the separable Hardamard transform with 4-point along each dimension in this work. Please note that SAD and SATD are generally different since they are the 1-norms (rather than the 2-norms) of the spatial- and the frequency-domain signals, respectively. In the first stage of the proposed mode decision process, we use these two features jointly for mode decision. This choice can be justified by the following observations.

For a given 4×4 block, we can order the SAD and SATD values according to their magnitudes, which is called the rank order. It is worthwhile to point out that, for a given 4×4 block, there exist 9 Intra-prediction modes while there are $9\times9 = 81$ positions in the joint (SAD, SATD) feature rank-ordered domain. Thus, these 81 positions are actually mutually exclusive. Only 9 out of 81 will be filled in with count number "1" while the other 72 positions will have a count number of "0". Fig. 1 shows the histogram of the RDO mode with respect to the joint rank-ordered features for two QCIF test sequences of 300 frames long. As shown in the figure, most RDO modes fall in the window of 3×3 lowest ranks. Statistically, 93-95% of RDO modes are in such a window. Thus, we may search the RDO mode using the modes that fall in this window, which actually allows at most 3 distinctive Intra-prediction modes. Thus, we can narrow down the search range based on the joint feature of $(SAD, SATD)$. It is worthwhile to point out that there always exist some candidate modes in the 3×3 window in all experiments performed. However, if there is no candidate mode in this window, we can enlarge the search window from 3×3 to 4×4 to find the possible candidate mode.

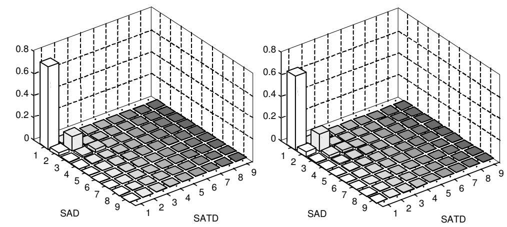


Fig. 1. The histogram of the RDO modes in the rank-ordered joint feature domain for test sequences "Akiyo" (left) and "Foreman" (right).

Another justification for the feature-based mode filtering is based on the rate-distortion analysis. Given a specific rank value (s_1, s_2) in the joint SAD-SATD feature domain. We can compute the average rate and distortion increase with respect to the RDO mode as

$$\Delta R(s_1,s_2) = N^{-1}_{s_1,s_2} \sum_{i\in S_{s_1,s_2}} [R_{s_1,s_2}(i) - R_{RDO}(i)],$$

$$\Delta D(s_1,s_2) = N^{-1}_{s_1,s_2} \sum_{i\in S_{s_1,s_2}} [D_{s_1,s_2}(i) - D_{RDO}(i)],$$

where i is the block index, S_{s_1,s_2} is the set of all events in which there exists a Intra-prediction mode in (s_1, s_2), N_{s_1,s_2} is the cardinality of S_{s_1,s_2}. $R_{s_1,s_2}(i)$ and $D_{s_1,s_2}(i)$ are the bit rate and the distortion for block i, respectively. Please note that $\Delta R(s_1, s_2)$ and $\Delta D(s_1, s_2)$ can be viewed as the posterior bit rate increase and distorion increase by choosing the mode associated with (s_1, s_2) as the RDO mode.

We plot $\Delta R(s_1, s_2)$ and $\Delta D(s_1, s_2)$ as the x- and y-axis in Fig. 2 for the Akiyo sequence, where the origin denotes the RDO mode and the modes within the lowest 3×3 rank window are labeled by the rank order. As shown in Fig. 2, we see that $(s_1, s_2) = (1, 1)$ is closest to the RDO mode. The next closest one is $(2, 1)$, and the third one is $(3, 1)$, and so on. This suggests an order to search for a good mode with an early termination rule. Similar R-D performance in the rank-ordered joint feature domain has been observed in other test sequences such as Foreman and Stefan.

Based on the above two observations, we conclude that the candidate modes to be considered in the next decision stage are those whose (s_1, s_2) feature vector are located in the lowest 3×3 window. The number of Intra-prediction modes existing in this region is typically 2 to 3, but no more than 3.

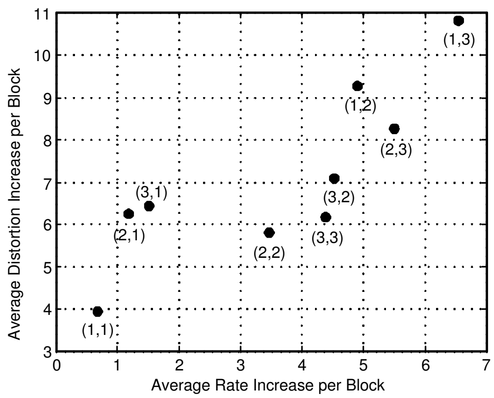


Fig. 2. The average rate (the x-axis) and the average distortion increase (the y-axis) for the Mobile squence of the QCIF format over 300 frames.

3.2. Early Termination Thresholds

The proposed fast mode decision algorithm is shown in Fig. 3. Due to the relative location of modes as shown in Fig. 2, we may weigh the $SATD$ feature more than the SAD feature. Thus, we first compute $SATD$ and then SAD. If the difference between the smallest SATD value (*i.e.* $SATD_1$) and the second smallest SATD value (*i.e.* $SATD_2$) is relatively large, the probability of the mode associated with $SATD_1$ to be the RDO mode is higher. Thus, we can choose that particular mode as the final mode directly without further processing. The same arguement applies to the SAD feature.

Thus, let us define

$$T = \frac{SAD_2 - SAD_1}{SAD_1}, \quad T' = \frac{SATD_2 - SATD_1}{SATD_1}.$$

To determine thresholds for T and T' for early termination, we examine the posterior error probabilities as functions of T and T' for several test sequences. The results are plotted in Fig. 4. When T or T' is larger, the error probability is smaller. We choose the threshold value such that the error probability is less than a constant, say, 10%. A different constant will give a different tradeoff between the complexity and performance degradation.

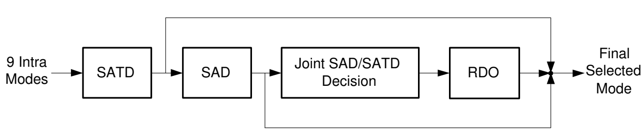
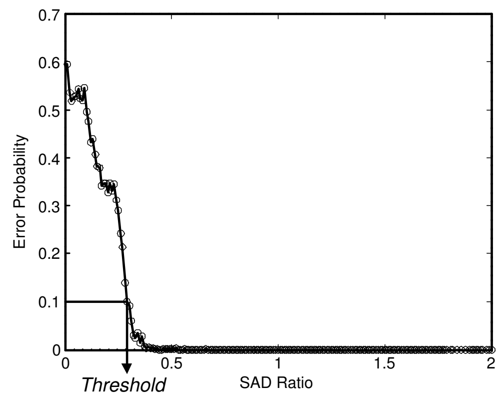


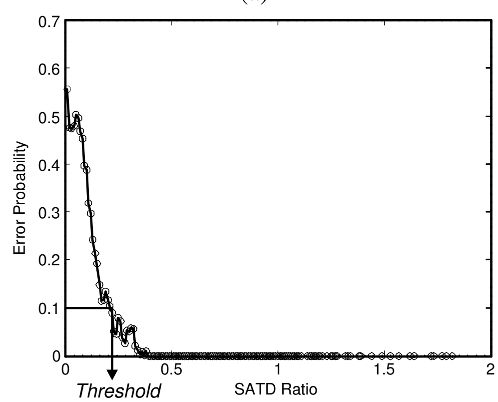
Fig. 3. The proposed multi-stage Intra-prediction mode decision scheme.

3.3. RDO Mode Search

As stated in Sec. 3.1, after the mode filtering process using joint SAD/SATD features, there are 2-3 candidate modes remaining for further consideration.



(a)



(b)

Fig. 4. Selection of early termination thresholds: (a)$P_{e|T}$, (b)$P_{e|T'}$.

In the next stage, one straightforward way is to apply the RDO method to these candidate modes directly. However, since the RDO complexity is very high, the rate-distiortion (R-D) model can be employed in the RDO mode search for further complexity reduction. In this work, the R-D model proposed in [6] is used when the quantization parameter is greater than 16 (*i.e.* excluding very high rate video) due to the limitation of the accuracy of the R-D model. Otherwise, the conventional RD optimization process as described in Sec. 2 is adopted.

4. EXPERIMENTAL RESULTS

The proposed mode decision scheme has been integrated with the H.264 JM7.3a codec for the performance evaluation. It is compared with the RDO mode decision of H.264 in terms of the computational cost (the average CPU time per call for the mode decision routine) and the average PSNR as a function of the coding bit-rate for test sequences recommended in [5]. Fig. 5 shows the RSNR performance and the computational complexity for the Akiyo sequence of 300 frames at various coding rates. The results obtained from the exhaustive RDO search and the proposed fast mode decision algorithm are represented by the circled and squared curves, respectively. We see that the proposed scheme gives almost identical RD performance while providing a speed-up factor ranging from 10.8 to 13.6. Note that the speedup factor is defined to be the ratio of the encoding time of the mode decision routine using the RDO technique and that of the proposed scheme.

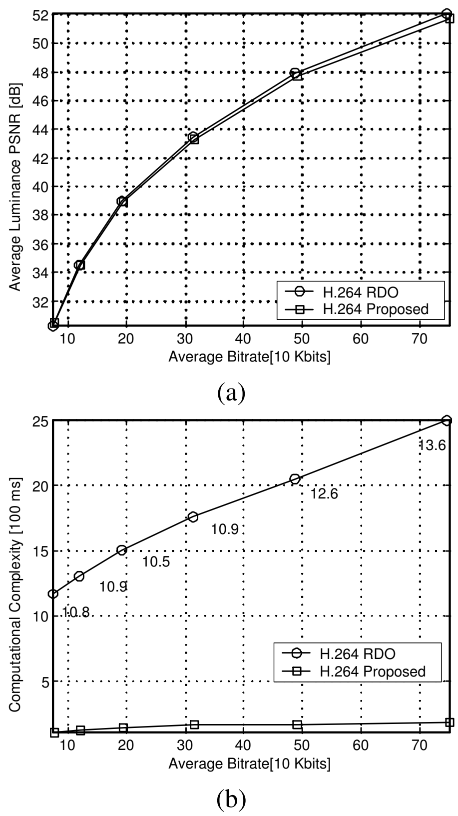


Fig. 5. Comparison of (a) the R-D performance and (b) the computational complexity for the QCIF Akio sequence.

More RD performance and complexity results for Akio, Foreman, Table Tennis (T. T.) and Mobile sequences are given in Tables I, II and III by varying quantization parameters (QPs) from 10 to 40. As compared to the full RDO search, the average rate increase (R) and the average distortion loss (D) for each sequence are: R:0.18% and D:1.3% for Akio, R:0.32% and D:1.3% for Foreman, R:0.5% and D:1.4% for Table Tennis and R:1.05% and D:1.3% for Mobile. The speed-up factor becomes larger as the QP goes smaller (or, equivalently, the bit rate goes higher). Most sequences follow the same trend as shown in Table III. To conclude, a feature-based multi-stage intra-prediction mode decision scheme was proposed for H.264, and a speed up factor of around 10-15 times was achieved without noticeable R-D performance degradation when it was implemented in JVT reference software JM7.3a in this work.

Table I. Rate comparison

R:RDO		Rate (Kbit/frame) [QP=10∼40]					
M:Proposed		10	16	22	28	34	40
Akio	R	74.6	48.8	31.2	19.1	11.8	7.4
	M	75.1	49.1	31.5	19.4	12.1	7.6
Foreman	R	109	72.5	43.8	24.8	13.8	8.0
	M	109	73.0	44.2	25.0	14.1	8.2
T. T.	R	110	73.3	44.8	26.1	15.7	9.2
	M	111	74.0	45.3	26.6	15.9	9.4
Mobile	R	208	159	118	81.5	51.6	26.8
	M	210	161	119	82.8	52.4	27.4

Table II. Distortion comparison

R:RDO		PSNR (dB) [QP=10∼40]					
M:Proposed		10	16	22	28	34	40
Akio	R	52.0	47.9	43.4	38.9	34.5	30.3
	M	51.7	47.7	43.3	38.9	34.5	30.4
Foreman	R	51.8	46.4	41.3	36.9	32.7	28.7
	M	51.3	46.2	41.2	36.8	32.7	28.8
T. T.	R	51.9	46.5	41.2	36.9	32.9	28.8
	M	51.3	46.2	41.1	36.8	32.8	28.8
Mobile	R	52.0	46.4	40.6	34.8	29.2	24.1
	M	51.3	45.9	40.1	34.4	28.9	24.0

Table III. Complextiy comparison

	Speedup Factor (scale) [QP=10∼40]					
	10	16	22	28	34	40
Akio	13.58	12.60	10.85	10.49	10.86	10.82
Foreman	13.53	13.42	10.37	11.25	10.13	9.20
T. T.	14.45	14.29	10.56	11.32	9.29	8.84
Mobile	14.27	14.84	12.14	15.39	14.31	11.88

5. REFERENCES

[1] Till Halbach, "Performance comparison: H.26L intra coding vs. JPEG2000," in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, JVT 4^{th} Meeting Klagenfurt, Austria, July 2002.

[2] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits, System and Video Technology*, vol. 7, no. 1-19, July 2003.

[3] Feng Pan and Xiao Lin *et. al.*, "Fast mode decision for intra prediction," in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, JVT 7^{th} Meeting Pattaya II, Thailand, March 2003.

[4] B. Jeon and J. Lee, "Fast mode decision for H.264," in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, JVT 10^{th} Meeting, Waikoloa, Hawaii, December 2003.

[5] JVT Test Model Ad Hoc Group, "Evaluation sheet for motion estimation," in *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, Draft version 4, Febrary 2003.