Title
Fast Mode Decision For H.264 Video Coding in Packet Loss Environment

Authors
Zhang, Y
Cosman, P C

Publication Date
2011-09-01

FAST MODE DECISION FOR H.264 VIDEO CODING IN PACKET LOSS ENVIRONMENT

Yuan Zhang[1] *and Pamela C. Cosman*[2]

[1] Communication University of China, Beijing, 100024, China
[2] University of California, San Diego, CA, 92093-0407, USA

ABSTRACT

In this paper, we propose a fast mode decision scheme for H.264 video coding to address the requirements of both low complexity and error resilience in realtime video communications. Traditional fast mode decision schemes are usually designed based on feature analysis on the source videos. However, the rate-distortion behaviors of the coding modes change when channel errors are involved. Therefore, the existing fast algorithms may not be applicable in error-resilient video coding. We first study the end-to-end rate-distortion behaviors of various coding modes, and then derive a hierarchical mode decision scheme. Different from the traditional fast algorithms that separate skip and inter modes at the beginning, we propose a fast estimation of the coding costs of skip and intra modes in a packet-loss environment, and then narrow the mode decision into one of the two paths: non-intra and non-skip. Testing shows the significant time savings of the proposed algorithm.

Index Terms— Fast mode decision, error resilience, H.264, video coding

1. INTRODUCTION

The state-of-the-art H.264 video coding standard enables significant coding efficiency compared with previous standards. However, the computational complexity of the encoder is increased. Since it is widely used for power limited portable devices over wireless networks, it becomes a challenging task to reduce the computational complexity of the encoding process while achieving good quality of service.

In the H.264 video coding standard, a number of coding modes are defined to adapt the rate allocation to the video contents. Rate-Distortion Optimization (RDO) is a typical mode decision scheme for H.264 video encoding. However, traditional RDO-based mode decision is time-consuming. Many fast mode decision schemes have been proposed for realtime video applications [1]-[4]. Among them, early skip mode selection is almost always the first priority, since skip mode is frequently used [1]. The MB mode prediction approaches try to predict the best mode without checking all the coding mode types [3]-[4]. A time-efficient learning theoretic classification algorithm to discern between broad mode classes was proposed [3]. Intra/inter mode decision is also important for fast mode decision, and a feature-based algorithm is proposed, which decides the mode using the expected risk of choosing the wrong mode in a multidimensional simple feature space [4].

Realtime video applications are often deployed under wireless network conditions, making error-resilience necessary. The Rate-Distortion (R-D) behavior of every coding mode changes when packet loss is considered. Accordingly, the traditional RDO-based mode decision has been extended for a number of error-resilient coding methods, with the estimation of the end-to-end distortion in the encoder [5]-[7]. A recursive optimal per-pixel estimate (ROPE) algorithm estimates the pixel level end-to-end distortion by keeping track of the first and second moments of the reconstructed pixel value [5]. An error robust RDO method (ER-RDO), developed for packet-loss environments [6], was adopted in the H.264/AVC test model. ER-RDO estimates the expected overall end-to-end distortion by decoding K random realizations of the lossy channel at the encoder. It is complex if K is large for accurate estimation.

In many realtime video applications, the requirements of fast encoding and error-resilient coding should be addressed simultaneously. The previous fast mode decision algorithms, usually based on the pre-analysis and statistics of the R-D behaviors of the H.264 coding modes, may not perform well in a packet loss environment. In this paper, we propose a fast mode decision scheme for H.264 error resilient video coding based on the study of R-D behavior in a packet loss environment. The algorithm can discern between broad mode classes based on estimation of the end-to-end R-D cost of skip and intra modes. Computational complexity is reduced since only the modes belonging to the selected class need to be evaluated.

The rest of this paper is organized as follows. In Section 2, the R-D behaviors of various coding modes in terms of packet loss are studied. Section 3 describes the proposed method, including the estimation of coding costs of intra and skip modes. Experimental results and conclusions are presented in Section 4.

978-1-4577-1303-3/11/$26.00

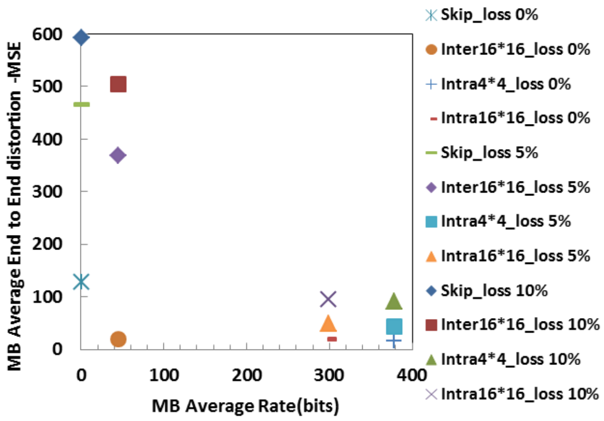

Figure 1: Average R-D operating points of various modes in different packet loss rates.

2. PROBLEM ANALYSIS

The objective of the fast mode decision is to accelerate the encoding process while achieving acceptable video quality at the decoder side. Therefore, the R-D behaviors and the distributions of selected coding modes in a packet loss environment are examined.

2.1 Rate-Distortion Behaviors with Packet Loss

Figure 1 shows the average R-D operating points of various coding modes in different packet loss rates. The end-to-end distortion is computed as the average of the distortion in the decoder by simulating the decoding process 100 times. Only the first frame is intra-coded; all the remaining are inter-coded. Each row of macroblocks composes a slice and is transmitted in a single packet. The statistics on Foreman (100 frames, 30 fps, QCIF, QP=28) are presented in this paper. The statistics on other sequences show similar properties.

In an error-free environment, usually skip mode has high distortion but low rate. The inter and intra modes have similar distortions, but the intra mode usually has higher rate than inter mode. In most cases, the intra mode is not selected as the optimal coding mode. Therefore, the pre-decision between skip and inter modes is the key to accelerate the overall mode decision. However, Figure 1 also indicates that the R-D behaviors change when the packet loss rate increases. The end-to-end distortions of both inter and skip modes increase significantly, while that of the intra mode is almost unchanged. Obviously, there is a significant gap between the R-D operating points of skip mode and intra mode.

2.2 Distributions of Coding Modes with Packet Loss

The percentage of time that a coding mode is optimal is determined by its R-D behavior. In Tables 1 and 2, we present these distributions. Because, as shown in Figure 1, the end-to-end distortion of inter mode increases dramatically under lossy conditions, whereas intra mode distortion increases only slightly, some percentage of inter decisions will convert to intra. Macroblocks in background and motionless areas, which are usually skip-coded, will tend not to transfer to intra mode under lossy conditions, because of both the high error resiliency of such areas, and the near zero rate cost of skip mode. This explains the trends seen in Tables 1 and 2: use of intra mode increases under lossy conditions, use of inter mode goes down, especially for P_8×8, and use of skip mode stays constant or increases slightly. The intra update rate greatly depends on the video content. In general, video with large motion usually needs a large number of intra blocks. This is because the intra coding cost (considering the end-to-end distortion) may be smaller than inter mode, which makes the skip/non-skip decision not the first priority task. Table 1 and 2 were derived using the error-resilient coding algorithm in the H.264 reference software [6].

Table 1: Distribution of various coding modes in packet loss environments (%) - Foreman QCIF, 128 kbps, IPPP

Loss Rate	Skip	P_16×16	P_16×8	P_8×16	P_8×8	Intra
0%	13.99	42.98	12.93	15.72	15.37	0.01
3%	13.49	43.19	12.27	15.16	13.12	2.77
5%	14.10	43.28	11.78	14.15	12.48	4.21
10%	15.08	42.98	11.17	13.82	10.14	6.81
20%	16.30	43.79	9.63	11.67	7.27	11.3

Table 2: Distribution of various coding modes in packet loss environments (%) - News QCIF, 64 kbps, IPPP

Loss Rate	Skip	P_16×16	P_16×8	P_8×16	P_8×8	Intra
0%	77.25	8.51	3.67	4.56	5.79	0.22
3%	77.00	8.43	3.57	4.57	5.99	0.44
5%	77.15	8.39	3.33	4.72	5.80	0.62
10%	77.42	7.92	3.41	4.66	5.47	1.13
20%	77.58	8.60	3.42	3.99	4.58	1.83

2.3 Discussion

When channel-induced distortion becomes dominant in the end-to-end distortion, the attempt to use MB-related features (e.g., motion activity) to predict the coding mode may not be helpful any more. Moreover, the previous analysis has indicated that the priority of intra mode increases with channel errors in the optimal coding mode selection, as shown in Figure 1. Therefore, in the first step of mode decision it is reasonable to consider intra, in addition to skip that is still a dominant mode. The comparison between the coding costs of intra and skip can separate the decision into two paths: non-intra and non-skip.

3. PROPOSED SCHEME

Figure 2 gives the flowchart of the proposed mode selection scheme. First, the non-intra or the non-skip mode decision path is selected according to the comparison between the estimated end-to-end R-D costs of skip and intra modes. The end-to-end R-D cost of skip mode J_{skip} can be directly calculated without pre-encoding the macroblock, which will be described in detail in Section 3.1. Therefore, $J_{skip,est}$ equals J_{skip}. However, the R-D cost of intra mode J_{intra} cannot be directly derived. The method to get its estimate $J_{intra,est}$ will be described in Section 3.2.

After the skip or intra mode is excluded based on the first decision, the problem is converted to two sub-problems: intra/inter mode decision and skip/inter mode decision. In the intra/inter path, the macroblock is pre-encoded with intra to calculate its end-to-end R-D cost J_{intra}. The true R-D cost J_{intra} is compared with J_{skip} to correct a possible classification error in the previous step. If J_{intra} is larger, then the decision switches to the inter/skip path; otherwise, it continues along the intra/inter path, starting from the calculation of the R-D cost of inter 16×16 (P_16×16) denoted $J_{inter16\times16}$. If J_{intra} is still larger, an inter mode will be selected, following the traditional RDO-based mode decision scheme; otherwise, intra is selected as optimal. The inter/skip mode decision path has a similar process, as shown in Figure 2.

3.1 Estimation of Skip Mode Coding Cost

The end-to-end distortion of an inter or skip mode consists of source, error-propagated, and error concealment distortions. Suppose p is the packet loss rate, REF lists the reference frames and m_J lists the motion vectors of all sub-blocks in macroblock m in frame n in terms of coding option o. The end-to-end rate distortion cost is:

$$J(n,m,o) = (1-p)\left(D_s(n,m,o) + D_{ep}(REF, m_J)\right) + pD_{ec}(n,m) + (1-p)\lambda R \tag{1}$$

where $D_s(n, m, o)$, $D_{ep}(REF, m_J)$, $D_{ec}(n, m)$ and R denote the macroblock-level source distortion, error-propagated distortion, error concealment distortion and number of coding bits, respectively. The error propagated distortion D_{ep} can be recursively calculated after the current frame has been encoded, and stored as a distortion map for further reference. The new Lagrange multiplier in a packet loss environment was derived as $(1-p)\lambda$ [7], where λ is the Lagrange multiplier used in the error-free environment.

Since $D_{ec}(n, m)$ is independent of o, it is unnecessary to be calculated for the mode selection. Therefore, the final formula is:

$$J'(n,m,o) = D_s(n,m,o) + D_{ep}(REF, m_J) + \lambda R \tag{2}$$

Since the rate of skip-coded macroblocks is nearly zero, the calculation of its R-D cost can be further simplified:

$$J_{skip} = D_{s,skip} + D_{ep}(REF, pred\_MV) \tag{3}$$

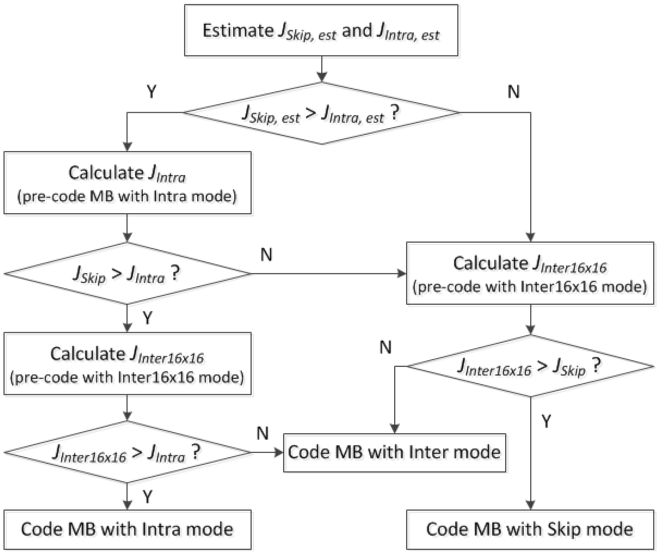

Figure 2: Flowchart of the proposed mode selection.

The source distortion $D_{s,skip}$ is the distortion between the current macroblock and the motion-compensated macroblock in the reference picture REF with zero displacement from the predicted motion vector $pred\_MV$, which can be calculated directly. The error-propagated distortion D_{ep} can be derived from the previous error propagation map [7]. Since J_{skip} can be directly calculated almost without computational cost, its estimate $J_{skip,est}$ in Figure 2 equals J_{skip}.

3.2 Estimation of Intra Mode Coding Cost

Using (2), the end-to-end R-D cost of intra mode is:

$$J_{intra} = D_{s,intra} + \lambda R \tag{4}$$

because the error-propagated distortion D_{ep} does not exist. J_{intra} can be calculated by pre-encoding the macroblock with various intra prediction directions, a time-consuming process. In the first step of the proposed scheme, an estimate of J_{intra} is used instead. Based on our statistics, J_{intra} of a macroblock has strong correlation with the coding cost of its co-located macroblock in the previous frame, indicated by J_{prev_mb}. Therefore, we use a linear function to estimate J_{intra} as:

$$J_{intra,est} = k \times J_{prev\_mb} \tag{5}$$

where k is based on a number of training sequences. In this paper, we set $k = 1.0$, $k = 2.2$, and $k = 3.0$, if the modes of the co-located macroblock in the previous frame are intra, inter, and skip, respectively.

4. EXPERIMENTAL RESULTS

The proposed algorithm was implemented in H.264 reference software JM12. The error-resilient video coding algorithm proposed in [7] is taken as a reference in the comparison; it does an exhaustive search of all coding modes. 5 sequences in QCIF format at 30 frames per second (fps) are used. Only the first frame is encoded as an I frame, and the remaining frames are encoded as P frames. There are four packets per frame. We assume that the packet containing the parameter set and packets of the first frame are conveyed reliably. The packet loss situation is simulated according to the error resilience testing conditions specified in [8]. Each sequence is decoded 100 times.

Table 3: Performance evaluation for the proposed scheme

Test Seq	Loss Rate	ΔPSNR (dB) Encoder	ΔPSNR (dB) Decoder	ΔBitrate (%)	Time Saving (%)
News QP:32	0%	-0.06	-	-1.12	31.7
	3%	-0.02	+0.56	-0.82	35.0
	5%	-0.05	+0.33	-3.22	34.6
	10%	-0.04	+0.42	-5.47	33.3
	20%	-0.06	+0.01	-11.7	33.4
Foreman QP:28	0%	-0.02	-	-0.19	13.1
	3%	-0.05	-0.10	-0.89	16.9
	5%	-0.05	+0.06	-3.09	16.0
	10%	-0.06	-0.13	-4.52	17.6
	20%	-0.07	-0.23	-9.09	17.1
Akiyo QP:30	0%	-0.04	-	-2.23	37.2
	3%	-0.03	-0.04	-3.13	37.0
	5%	-0.03	-0.04	-4.23	35.5
	10%	-0.05	-0.02	-6.40	36.1
	20%	-0.04	-0.02	-11.2	36.5
Carphone QP:32	0%	+0.02	-	-1.34	19.5
	3%	-0.08	-0.25	-1.45	24.5
	5%	-0.07	-0.17	-2.96	23.2
	10%	-0.12	-0.18	-3.75	22.6
	20%	-0.14	+0.17	-8.87	22.3
Container QP:28	0%	-0.03	-	-1.49	38.2
	3%	-0.02	-0.03	-1.96	36.3
	5%	-0.02	-0.07	-0.82	36.9
	10%	-0.03	-0.13	-2.69	37.3
	20%	-0.03	-0.11	-4.17	36.7

Comparing the proposed algorithm to the reference, we use ΔPSNR Encoder and ΔPSNR Decoder to indicate the gains in video quality measured at the encoder and at the decoder, respectively, and use ΔBitrate to indicate the percentage increase of bit rate. As shown in Table 3, the proposed algorithm achieves similar R-D performance to the exhaustive mode decision approach while reducing the coding time significantly. The time savings from the proposed scheme is mainly achieved from the pre-determination of the inter/intra and skip/inter paths. It depends on the percentage of skip- and intra-coded blocks. For example, the time savings on Foreman is less than others, because it has fewer skip- and intra-coded blocks. Another fast algorithm can be applied in the following stages, to improve the time savings.

Conclusion: In this paper, we examined the problem of interference between fast mode decision and error resilience in video coding. We proposed a fast mode decision algorithm based on the estimation of end-to-end R-D cost. By wisely skipping a number of motion estimations and intra prediction mode decisions, the algorithm achieves up to 38% time savings with virtually no degradation in R-D performance. The time savings is achieved by a simple first step non-intra and non-skip modes classification. For future work, fast inter mode selection should be considered to achieve better R-D cost prediction and yield greater time savings.

Acknowledgment

This work was supported in part by National Natural Science Foundation of China (61001177).

REFERENCES

[1] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," IEEE Trans. Circuits and Systems for Video Tech., vol.16, no.12, pp.1557-1561, Dec. 2006.

[2] C. S. Kannangara, I. E. G. Richardson, M. Bystrom, J. R. Solera, Y. Zhao, A. MacLennan and R. Cooney, "Low complexity skip prediction for H.264 through Lagrangian cost estimation," IEEE Trans. Circuits and Systems for Video Tech., vol.16, no.2, pp.202-208, Feb. 2006.

[3] A. Jagmohan, K. Ratakonda, "Time-efficient learning theoretic algorithms for H.264 mode selection," Intl. Conf. on Image Processing 2004, pp.749-752.

[4] C. Kim and C-C. J. Kuo, "Feature-based Intra/Inter Coding Mode Selection for H.264/AVC," IEEE Trans. Circuits and Systems for Video Tech., vol.17, no.4, pp.441-453, April 2007.

[5] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," IEEE J. on Selected Areas in Comm., vol. 18, pp. 966-976, June 2000.

[6] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," in Proc. Packet Video Workshop, Apr. 2002.

[7] Y. Zhang, W. Gao, Y. Lu, Q.M. Huang, and D.B. Zhao. "Joint Source-Channel Rate-Distortion Optimization for H.264 Video Coding over Error-Prone Networks," IEEE Trans. Multimedia. Vol. 9, No. 3, pp. 445-454, April 2007.

[8] S. Wenger, "Common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing," ITU-T VCEG document VCEG-N79r1, Sep. 2001.