LETTER

A Novel Joint Rate Distortion Optimization Scheme for Intra Prediction Coding in H.264/AVC*

Qingbo WU[†a)], ***Student Member***, **Jian XIONG**[†], **Bing LUO**[†], **Chao HUANG**[†], ***Nonmembers***, ***and*** **Linfeng XU**[†], ***Student Member***

SUMMARY In this paper, we propose a novel joint rate distortion optimization (JRDO) model for intra prediction coding. The spatial prediction dependency is exploited by modeling the distortion propagation with a linear fitting function. A novel JRDO based Lagrange multiplier (LM) is derived from this model. To adapt to different blocks' distortion propagation characteristics, we also introduce a generalized multiple Lagrange multiplier (MLM) framework where some candidate LMs are used in the RDO process. Experiment results show that our proposed JRDO-MLM scheme is superior to the H.264/AVC encoder.

key words: *RDO, video coding, H.264/AVC*

1. Introduction

Rate distortion optimization (RDO) theory [1] plays an important role in video coding technology. By dealing with the mode decision process as a rate constrained distortion minimization problem [2], the RDO could select the optimal coding option at the cost of increased complexity. Then, the video coding becomes a trade-off task between the coding performance and the complexity. Some works put emphasis on improving rate-distortion (RD) performance. N'guessan et al. proposed a region of interest based method [3], which improved the video coding by introducing the human attention/saliency model [4]–[8]. Yang et al. proposed a temporal propagation model to improve the motion compensation module [9]. Other researchers have tried to reduce the complexity of the codec by refining the candidate intra modes [10]–[13]. In this paper, we focus on improving the RDO performance at low complexity cost.

In [2], Wiegand et al. proposed a Lagrangian multiplier (LM) determination method which derived the optimal LM as a dependent variable of the quantization parameter (QP). Because of its simplicity and efficiency, this method is widely employed in many hybrid video codecs. However, the encoding independence hypothesis in [2] limits its performance to achieve global optimal rate-distortion performance. Since the intra prediction depends on the neighboring reconstructed coding unit (CU), the distortion in the current block will impact the encoding performance in the subsequent CUs. In this paper, we improve the intra mode determination with a joint RDO (JRDO) model. A linear spatial distortion propagation model can be obtained by offline training. Based on this distortion propagation model, we propose a JRDO LM method. Since different contents in each block present different distortion propagation characteristics, a generalized multiple Lagrange multiplier (MLM) framework is designed to select the optimal coding option under multiple candidate LMs. Trough a series of theoretical derivations, the proposed JRDO-MLM scheme builds a robust framework to analyze the spatial JRDO problem, which makes an effort to approach the global optimal RDO solution.

The remainder of this paper is organized as follows. Section 2 describes the proposed JRDO LM derivation process. Section 3 introduces the MLM framework. The experimental results are presented in Sect. 4. Finally, we draw the conclusion in Sect. 5.

2. The Proposed Lagrangian Multiplier Determination Method

In this paper, we only discuss the H.264/AVC intra prediction coding, where the temporal dependency is not considered. Let's denote the number of CUs in a frame by N_c and denote the coding option for the ith CU by o_i. The option combination of all CUs can be denoted by o_c where $o_c = o_1 \bigcup o_2 \bigcup \cdots \bigcup o_{N_c}$. In LM method, the coding option determination process can be converted to the RDO problem by minimizing the cost of the Lagrangian formulation, i.e,

$$\min_{o_c} \sum_{i=1}^{N_c} J_i(o_c|\lambda) \quad \text{with } J_i(o_c|\lambda) = D_i(o_c) + \lambda \cdot R_i(o_c) \tag{1}$$

where $J_i(o_c|\lambda)$ is the Lagrangian cost function for the ith CU and λ is a pre-defined LM.

To further simplify the global RDO problem in (1), an independent assumption among CUs is made in [2], which produces a suboptimal solution. Ideally, we can get better LM by introducing the distortion propagation into the Lagrangian cost function. As discussed in [14], the quantization errors meet Markov property in predictive coders. So, we only consider the spatial prediction dependency in the neighboring CUs. The problem in (1) can be reduced to a JRDO model

Manuscript received August 20, 2013.
Manuscript revised October 29, 2013.

[†]The authors are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China 611731.

[*]This work was supported by the Fundamental Research Funds for the Central Universities (ZYGX2012YB007).

a) E-mail: wqb.uestc@gmail.com

DOI: 10.1587/transinf.E97.D.989

$$\min_{o^l_{c,i}} \sum_{j=i}^{i+1} J_j(o^l_{c,i}|\lambda) \quad \text{with } J_j(o^l_{c,i}|\lambda) = D_j(o^l_{c,i}) + \lambda \cdot R_j(o^l_{c,i}) \tag{2}$$

where $o^l_{c,i}$ is the local neighboring CUs' coding option combination and $o^l_{c,i} = o_i \bigcup o_{i+1}$.

To sequentially solve o_i for the ith CU in (2), we can assume that the subsequent CU's coding option is known which can be denoted by o^*_{i+1}. The problem in (2) can be rewritten as

$$o^l_{c^*,i} = \min_{o^l_{c^*,i}} \sum_{j=i}^{i+1} J_j(o^l_{c^*,i}|\lambda) \quad \text{with } J_j(o^l_{c^*,i}|\lambda) = D_j(o^l_{c^*,i}) + \lambda \cdot R_j(o^l_{c^*,i}) \tag{3}$$

where $o^l_{c^*,i} = o_i \bigcup o^*_{i+1}$ and o_i is the only undetermined variable to be solved.

Since the joint rate distortion cost can be estimated with the spatial distortion propagation, let's formulate the distortion propagation process as

$$\bar{D}_{i+1} = f(\bar{D}_i) \tag{4}$$

where $\bar{D}_i$ is the mean square error (MSE) of the ith CU and $f(\cdot)$ is the spatial distortion propagation function. Here, we replace the rate of each CU with a R-D function $r(\cdot)$ which can be represented as

$$\bar{R}_i = r(\bar{D}_i) \tag{5}$$

where $\bar{R}_i$ is the mean rate of the ith CU.

Then, it is easy to be shown that the optimal LM for the JRDO framework is

$$\lambda = -\frac{\mathrm{d}\left[\bar{D}(Q) + f(\bar{D}(Q))\right]/\mathrm{d}Q}{\mathrm{d}\left[r\left(\bar{D}(Q)\right) + r\left(f(\bar{D}(Q))\right)\right]/\mathrm{d}Q} \tag{6}$$

where $\bar{D}(Q)$ is the distortion-to-quantizer relation function and Q is the quantization step.

Thanks to the works of the predecessors in [15] and [16], we can obtain the expressions of $r(\cdot)$ and $\bar{D}(Q)$

$$r\left(\bar{D}(Q)\right) = c \log_2\left(\frac{d}{\bar{D}(Q)}\right) \tag{7}$$

$$\bar{D}(Q) = \frac{Q^2}{12} \tag{8}$$

where c and d are the parameters to describe the functional relationship between rate and distortion.

Based on a large number of statistical analysis, we find that the adjacent blocks' MSE is monotonic and near-linear dependency with the current block's MSE. For clarity, an intuitive statistical result for the *news* sequence is shown in Fig. 1. Accordingly, we employ the linear fitting function to model the spatial distortion propagation process, i.e.,

$$f(\bar{D}_i) = a \cdot \bar{D}_i + b \tag{9}$$

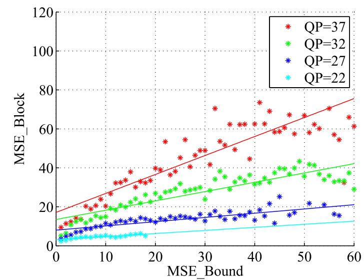


Fig. 1 The relationship of adjacent blocks' distortions. The *x-axis* denotes the boundary blocks' MSE and the *y-axis* denotes current block's MSE.

Table 1 Analytic expression of parameters.

Parameters Analytic Expression			
m_1	$3p^2k_1^2$	m_6	$144k_2l_2$
m_2	$5p^2k_1l_1 + 2pk_1$	m_7	$60pk_1$
m_3	$2p^2l_1^2 + 2pl_1 + 48pk_1k_2$	m_8	$48pl_1$
m_4	$24k_2 + 36pk_1l_2 + 36pk_2l_1$	m_9	$432k_2$
m_5	$144k_2^2 + 24l_2 + 24pl_1l_2$	m_{10}	$288l_2$

where a and b are the fitting parameters.

Since the impact of spatial distortion propagation increases as QP becoming larger, we further represent a and b as the dependent values of Q, i.e.,

$$a = k_1 \cdot Q + l_1 \tag{10}$$

$$b = k_2 \cdot Q + l_2 \tag{11}$$

where k_1, l_1 and k_2, l_2 are the fitting parameters for a and b respectively, and we can obtain these four parameters by an off-line training method.

Finally, we can obtain the JRDO LM by plugging (10)~(14) into (9), i.e.,

$$\lambda = w \cdot \frac{\begin{pmatrix} m_1 \cdot Q^6 + m_2 \cdot Q^5 + m_3 \cdot Q^4 + \\ m_4 \cdot Q^3 + m_5 \cdot Q^2 + m_6 \cdot Q \end{pmatrix}}{m_7 \cdot Q^3 + m_8 \cdot Q^2 + m_9 \cdot Q + m_{10}} \tag{12}$$

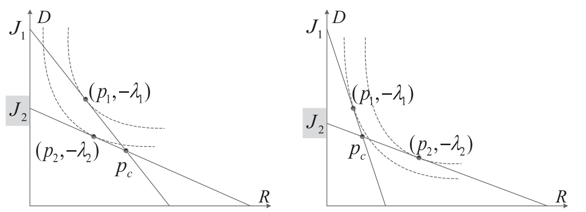
where $w = (\ln 2)/c$ and the other parameters are shown in Table 1.

3. Generalized Multiple-LM Framework

The single LM use the same LM for every CU in a frame, which can't capture the image contents variation's impact on the spatial distortion propagation. Accordingly, we propose a generalized multiple-LM framework (MLM), where multiple candidate LMs are used in the RDO process and a new coding option determination criterion is designed for the MLM framework.

Let's denote the candidate coding options and LMs of o_i by p_u and λ_u, respectively. The R-D curve that goes through p_u is denoted by $RD_C(p_u)$ and the tangent line of the R-D curve that goes through p_u is denoted by $RD_L(p_u, \lambda_u)$. Let's represent $RD_L(p_u, \lambda_u)$ by

$$D - D(p_u) = -\lambda_u (R - R(p_u)) \tag{13}$$



(a) p_2 is superior to p_1 (b) p_2 is inferior to p_1

Fig. 2 RDO process of the MLM framework. The curves are the R-D curves for each coding option and the solid lines are the tangent lines of each R-D curve.

Table 2 Criterion of optimal coding option selection.

LM Relationship	Rate Interval	Optimal Coding Option $(o_i, -\lambda_i)$
$\lambda_1 < \lambda_2$	$R(p_2) < R(p_c)$	$(p_1, -\lambda_1)$
	$R(p_2) \geq R(p_c)$	$(p_2, -\lambda_2)$
$\lambda_1 > \lambda_2$	$R(p_2) < R(p_c)$	$(p_2, -\lambda_2)$
	$R(p_2) \geq R(p_c)$	$(p_1, -\lambda_1)$

where $D(p_u)$ and $R(p_u)$ represent the distortion and rate under coding option p_u respectively.

For different LMs, there is always an intersection point between two R-D tangent lines labeled by p_c. Then, we can solve the R-D point of p_c as

$$R(p_c) = \frac{\begin{pmatrix} D(p_2) - D(p_1) - \\ \lambda_1 R(p_1) + \lambda_2 R(p_2) \end{pmatrix}}{\lambda_2 - \lambda_1} \tag{14}$$

We show two positional relations for R-D points $(D(p_1), R(p_1))$ and $(D(p_2), R(p_2))$ where $\lambda_1 > \lambda_2$ in Fig. 2. In Fig. 2 (a), p_2 is under $RD_L(p_1, \lambda_1)$ when the slope of this line is less than zero and p_2 is on the left of p_c. The the distortion value in $RD_L(p_1, \lambda_1)$ is greater than $D(p_2)$ when the rate is $R(p_2)$. Since $RD_L(p_1, \lambda_1)$ is the tangent line of $RD_C(p_1)$, the distortion value in $RD_C(p_1)$ is equal or greater than the one in $RD_L(p_1, \lambda_1)$ under the same rate. Then, we know that the distortion value in $RD_C(p_1)$ is also greater than $D(p_2)$ when the rate is $R(p_2)$. Since p_2 is in $RD_C(p_2)$, we can conclude that p_2 will achieve better R-D performance than p_1. In Fig. 2 (b), an opposite result can be found as p_2 is on the right of p_c. Based on this observation, we design a mode determination criterion for MLM framework as shown in Table 2.

Here, we employ a MB-level MLM scheme by exploring multiple candidate LMs. First, we compute the rate and distortion under all available coding options. Second, we find the optimal coding options under each candidate LM. Third, the final coding option is selected from the refined coding options in the second step based on the criterion in Table 2.

4. Experimental Results

To verify the performance of our proposed JRDO scheme, we implement the proposed method on the VCEG KTA2.4r1†. Here, both the common simulation conditions (QP={22,27,32,37}) and low bitrate simulation conditions (QP={36,40,44,48}) are involved in our experiment. In this intra only simulation, the H.264/AVC High Profile is used as the benchmark. All distortion propagation parameters (k_1, l_1, k_2, l_2) are obtained by off-line training. The training set is collected from the open access database††. To be fair, the test sequences are selected from the recommendation [17], which are not involved in the training set. The parameter w is set to 3.7. We denote the single LM scheme that employs our proposed LM by JRDO-SLM. The MLM scheme which combines our JRDO LM and the conventional LM is denoted by JRDO-MLM-x, where x indicates the numbers of the candidate LMs. For JRDO-MLM-2, the JRDO LM parameters are $(k_1, l_1) = (0.0411, -0.0502)$ and $(k_2, l_2) = (1.3270, 0.9419)$. Since the distortion propagation is more significant at high QPs, we add an extra JRDO-MLM-4 test under low bit rate conditions, where two additional LMs are derived with the parameters $(k_1, l_1) = (0.0562, -0.1098)$, $(k_2, l_2) = (1.7365, -0.5345)$ and $(k_1, l_1) = (0.0214, 0.0647)$, $(k_2, l_2) = (0.6983, 1.4767)$.

For evaluating the coding efficiency, BDPSNR (Bjonteggard Delta PSNR) and BDBR (Bjonteggard Delta Bit-Rate) [18] are used in our experiment. To further evaluate the complexity of different schemes, the percentage of difference of coding time ($\Delta T\%$) is employed, i.e.,

$$\Delta T = \frac{T_{pro} - T_{anc}}{T_{anc}} \times 100 \tag{15}$$

where T_{pro} and T_{anc} denote the coding time of the proposed scheme and the anchor respectively.

4.1 Coding Performance

The detailed coding results under both the common and low bitrate conditions are shown in Table 3. It can be found that our proposed JRDO-MLM scheme achieves superior R-D performance under both test conditions. The robustness of JRDO-MLM scheme is better than the JRDO-SLM scheme. Relative to the anchor scheme, the JRDO-SLM works not well for some sequences like *ParkScence* and *RaceHorses*. This is consistent with our discussion in Sect. 3, i.e., the single LM can't capture different block contents' impact on distortion propagation. In the JRDO-MLM scheme, we effectively improve the robustness with the LM switching strategy. As shown in Table 3, the JRDO-MLM-4 scheme achieves better performance than JRDO-MLM-2. That is, the coding gain of JRDO-MLM scheme is positively associated with the number of candidate LMs.

In addition, it should be noted that the JRDO-MLM scheme can achieve more significant coding gains under low bitrate conditions. This is consistent with the fact that the impact of the distortion propagation increases as QP becoming larger.

†http://iphome.hhi.de/suehring/tml/download/KTA/
††http://media.xiph.org/video/derf/

Table 3 Coding gains in terms of BDBR (%) and BDPSNR (dB).

Sequence		Common Test				Low-bitrate Test					
		JRDO-SLM		JRDO-MLM-2		JRDO-SLM		JRDO-MLM-2		JRDO-MLM-4	
		BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR	BDBR	BDPSNR
1080p	BasketballDrive	-0.3042	0.0101	-0.5350	0.0157	-0.5677	0.0211	-0.7734	0.0309	-0.8200	0.0323
	BQTerrace	0.0829	-0.0099	-0.3530	0.0209	0.2849	-0.0118	-0.9154	0.0437	-1.3288	0.0637
	Cactus	-0.1687	0.0072	-0.5656	0.0229	-0.3001	0.0119	-1.0379	0.0412	-1.2592	0.0511
	ParkScence	0.2636	-0.0143	-0.5315	0.0244	0.2559	-0.0076	-1.2044	0.0414	-1.1885	0.0411
	Kimono1	-0.5176	0.0181	-0.9720	0.0348	-0.9162	0.0458	-1.2914	0.0640	-1.2325	0.0620
Average 1080p		-0.1288	0.0022	-0.5914	0.0237	-0.2487	0.0119	-1.0445	0.0442	-1.1657	0.0501
720p	vidyo1	-0.1894	0.0104	-0.5935	0.0309	-0.8677	0.0520	-1.1359	0.0661	-1.2841	0.0750
	vidyo3	-0.3310	0.0195	-0.4851	0.0281	-0.6756	0.0369	-0.8677	0.0495	-1.1009	0.0630
	vidyo4	-0.3971	0.0186	-0.6805	0.0324	-0.4981	0.0250	-0.8264	0.0399	-1.0061	0.0491
Average 720p		-0.3058	0.0162	-0.5864	0.0305	-0.6805	0.0380	-0.9434	0.0518	-1.1304	0.0623
Average Total		-0.2173	0.0092	-0.5889	0.0271	-0.4646	0.0250	-0.9940	0.0480	-1.1481	0.0562

Table 4 Complexity increase in term of $\triangle T$ (%).

Sequence		Common Test		Low-bitrate Test		
		JRDO-SLM	JRDO-MLM-2	JRDO-SLM	JRDO-MLM-2	JRDO-MLM-4
1080p	BasketballDrive	0.04	11.58	2.11	16.85	48.77
	BQTerrace	-0.08	12.26	-1.18	15.22	49.46
	Cactus	0.06	11.88	1.18	14.23	48.17
	ParkScence	-0.69	11.30	-0.76	16.92	48.19
	Kimono1	0.04	11.15	4.58	14.52	47.70
Average 1080p		-0.13	11.63	1.19	15.55	48.46
720p	vidyo1	-0.57	12.05	5.20	16.83	53.80
	vidyo3	-0.50	12.10	3.69	16.40	55.09
	vidyo4	-0.39	12.20	2.50	15.91	53.19
Average 720p		-0.49	12.12	3.80	16.38	54.03
Average Total		-0.31	11.88	2.90	15.97	51.25

4.2 Complexity Analysis

To compare the computational complexity of different schemes, we show the $\triangle T$ results under both the common and low bitrate conditions in Table 4. It can be seen that for JRDO-SLM scheme the complexity is close to conventional RDO scheme since we only replace the LM with our proposed JRDO model. For JRDO-MLM-2 scheme, the average $\triangle T$s are 11.88% and 15.97% under the common and low bitrate conditions, respectively. For JRDO-MLM-4 scheme, the average $\triangle T$ increases to 51.25%. Since the MLM framework needs to explore all available LMs, the complexity linearly increases with the number of the LMs.

5. Conclusion and Future Work

In this paper, we proposed a spatial distortion propagation based JRDO model and MLM framework. Since the JRDO model could minimize both the distortions in the current block and the neighboring block, the proposed scheme achieves superior R-D performance over H.264/AVC High Profile. In our future work, a more flexible spatial distortion propagation model will be studied to be compatible with the latest HEVC codec.

References

[1] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.74–90, Nov. 1998.

[2] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," ICIP'01, pp.542–545, 2001.

[3] S. N'guessan and N. Ling, "Human attention region-of interest in i-frame for video coding," VCIP'12, pp.1–5, 2012.

[4] W.H. Cheng, C. Wei-Ta, and W. Ja-Ling, "A visual attention based region-of-interest determination framework for video sequences," IEICE Trans. Inf. & Syst., vol.E88-D, no.7, pp.1578–1586, July 2005.

[5] H. Li and K.N. Ngan, "A co-saliency model of image pairs," IEEE Trans. Image Process., vol.20, no.12, pp.3365–3375, 2011.

[6] H. Li and K.N. Ngan, "Saliency model based face segmentation in head-and-shoulder video sequences," J. Vis. Commun. Image R., vol.19, no.5, pp.320–333, 2008.

[7] F. Meng, H. Li, G. Liu, and K.N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," IEEE Trans. Multimedia, vol.14, no.5, pp.1429–1441, 2012.

[8] W. Luo, H. Li, G. Liu, and K.N. Ngan, "Global salient information maximization for saliency detection," Signal Process.-Image Commun., vol.27, no.3, pp.238–248, 2012.

[9] T. Yang, C. Zhu, X. Fan, and Q. Peng, "Source distortion temporal propagation model for motion compensated video coding optimization," ICME'12, pp.85–90, 2012.

[10] J. Xiaocong, S. Jun, Y. Huang, S. Jia, and T. Ikenaga, "Fast h.264/avc direct mode decision based on mode selection and predicted rate-distortion cost," IEICE Trans. Inf. & Syst., vol.E94-D, no.8, pp.1653–1662, Aug. 2011.

[11] K. Bharanitharan, D. Jiun-Ren, C. Bo-Wei, and W. Jhing-Fa, "Selective intra block size decision and fast intra mode decision algorithms for h.264/avc encoder," IEICE Trans. Inf. & Syst., vol.E95-D, no.11, pp.2720–2723, Nov. 2012.

[12] H. Li, K.N. Ngan, and Z. Wei, "Fast and efficient method for block edge classification and its application in H.264/AVC video coding," IEEE Trans. Circuits Syst. Video Technol., vol.18, no.6, pp.756–768, June 2008.

[13] Z. Wei, K.N. Ngan, and H. Li, "An efficient intra-mode selection algorithm for h.264 based on edge classification and rate-distortion estimation," Signal Process.-Image Commun., vol.23, no.9, pp.699–710, 2008.

[14] D. Arnstein, "Quantization error in predictive coders," IEEE Trans. Commun., vol.23, no.4, pp.423–429, April 1975.

[15] N.S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[16] H. Gish and J.N. Pierce, "Asymptotically efficient quantizing," IEEE Trans. Inf. Theory, vol.14, no.5, pp.676–683, Sept. 1968.

[17] F. Bossen, Common test conditions and software reference configurations. ITU-T and ISO/IEC, 2010. JCTVC-B300.

[18] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T VCEG, VCEG-M33, 2001.