LAGRANGE MULTIPLIER SELECTION FOR 3-D WAVELET BASED SCALABLE VIDEO CODING

Fuzheng Yang[1,2]*, Shuai Wan*[1]*, Ebroul Izquierdo*[1]

[1] Multimedia and Vision Research Lab, Queen Mary, University of London
{fuzheng.yang, shuai.wan, ebroul.izquierdo}@elec.qmul.ac.uk
[2] State Key Lab of Integrated Service Networks, Xidian University

ABSTRACT

In this paper a thorough analysis on the theoretical rate distortion model and the rate distortion performance in an open-loop structure is conducted. A Lagrange multiplier selection for 3-D wavelet based scalable video coding is then derived. The proposed Lagrange multiplier is adaptive with respect to the characteristics of video content. Furthermore, it is especially suitable for 3-D wavelet based scalable video coding where quantisation steps are unavailable. Extensive experimental results have demonstrated the effectiveness of the proposed Lagrange multiplier selection.

Index Terms— *Lagrange multiplier, rate distortion model, scalable video coding, wavelet*

1. INTRODUCTION

Fine granularity and truly scalable video coding (SVC) can be achieved by 3-D wavelet models. Such models are based on spatial-temporal subbands decomposition and bit-plane coding, which provide scalability in a natural way. Over the last few years much research has been devoted to the corresponding decomposition and coding strategies. However, rate distortion (RD) optimisation for coder control still remains as an open issue. A common solution to the RD optimised coder control resides in Lagrangian optimisation, which chooses the coding parameters by finding the trade-off between distortion and bit rate. A critical problem resides in Lagrangian optimisation is how to select the Lagrange multiplier, which controls the RD trade-off in encoding. A simple and effective Lagrange multiplier selection has been proposed in [1]. This formula has been widely adopted and successfully applied to the latest H.264/AVC in an extended version [2][3]. Few techniques have been proposed with respect to improved Lagrange multiplier selection, including an adaptive one for H.264/AVC [4], and a parametric approximation based one for block-based motion estimation [5].

Although the Lagrange multiplier selection techniques in literature differ in formulation, they feature a crucial common aspect: the Lagrange multiplier is tied to the quantisation step. However, this feature arises as an overwhelming obstacle when conventional Lagrange multiplier selection is applied to 3-D wavelet based SVC. As some other SVC techniques, 3-D wavelet based SVC employs embedded quantisation and bit-plane coding, which results in an embedded bitstream with fine granularity scalability. Consequently, there is not specified value of the quantisation step during encoding. Instead, the bitsream can be truncated at various bit rates targeting at different applications. Therefore, in absence of the quantisation step the conventional Lagrange multiplier cannot be directly applied to fine granular SVC. Furthermore, the Lagrange multiplier is derived from RD models. However, traditional RD models cannot describe the RD behaviour of 3-D wavelet based SVC, due to its open-loop prediction structure. Accordingly, the consequential Lagrange multiplier selection for 3-D wavelet based SVC should be addressed differently.

In this paper an effective Lagrange multiplier selection is proposed for 3-D wavelet based SVC. Starting from an analysis on the theoretical RD model, the RD performance of wavelet based SVC is evaluated considering its open-loop structure. The Lagrange multiplier formulation is derived from the corresponding RD model for 3-D wavelet based SVC. The proposed Lagrange multiplier takes the form of a function of the targeted bit rate. Therefore it is suitable for 3-D wavelet based SVC where quantisation steps are unavailable. The rest of this paper is organised as follows. In section 2, the background of Lagrangian optimisation and Lagrange multiplier is outlined. A detailed description of theoretical RD analysis and the proposed Lagrange multiplier selection are given in section 3. In section 4 selected results from the experimental evaluation are reported to demonstrate the performance of the proposed method. The paper closes with concluding remarks in section 5.

2. BACKGROUND

RD optimisation for coder control aims at choosing the best coding parameter combination in order to minimise the distortion at a given coding bit rate. In view of Lagrangian optimisation, this issue can be formulated as minimising the cost function J using the Lagrange multiplier λ. The cost function is given by:

$$J = D + \lambda \cdot R \,, \tag{1}$$

where D is the distortion measure and R is the bit rate related to D. In theory, the Lagrange multiplier is given by

$$\lambda = -dD / dR \,. \tag{2}$$

The solution to (2) for traditional hybrid video coder control under mean squared distortion measure yields the following relationship between the Lagrange multiplier and the quantisation step Q:

$$\lambda = c \cdot Q^2 \,. \tag{3}$$

This widely adopted formulation (3) takes a different form when applied to H.264/AVC, as $\lambda = 0.85 \cdot 2^{(QP-12)/3}$, where the quantisation parameter (QP) in H.264/AVC is introduced instead of Q, whereas in essence it is in accordance with (3). It is noticeable that these equations are based on the distortion measure of sum of squared difference (SSD). In the most common case in motion estimation, where sum of absolute difference (SAD) is utilised instead of SSD as the distortion measure, the Lagrange multiplier for motion estimation λ_{motion} should be adjusted by

$$\lambda_{motion} = \sqrt{\lambda} \,. \tag{4}$$

As indicated in section 1, the conventional Lagrange multiplier functions are not applicable to embedded coding, in absence of the quantisation step. Furthermore, wavelet based SVC is usually based on an open-loop prediction structure, which results in a quite different RD performance compared with the traditional closed-loop based hybrid video coders [6]. Therefore, a different Lagrange multiplier should be derived for wavelet based SVC.

3. LAGRANGE MULTIPLIER SELECTION FOR WAVELET BASED SVC

According to the theoretical RD model given by [7], the average distortion $\overline{D}$ of the transformed coefficients can be expressed as follows under high resolution quantisation hypothesis:

$$\overline{D} = \frac{1}{12} 2^{2\overline{H}_d} 2^{-2\overline{R}} \,, \tag{5}$$

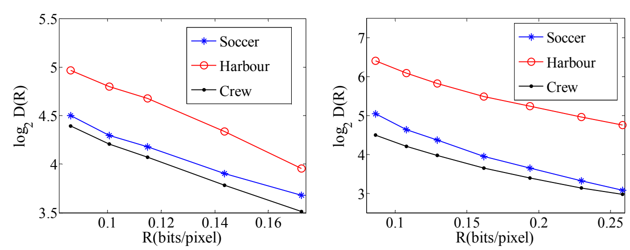
where $\overline{H}_d$ is the averaged differential entropy, and $\overline{R}$ is the average number of bits per sample (R_m), given by $\overline{R} = \left(\sum_{m=0}^{N-1} R_m \right) / N$. It is noticeable that here the bits for coding every coefficient are respectively summed and the coefficients are considered as being coded independently. However, the assumed independent coding for each coefficient is usually not employed in video coding. Instead, techniques such as the run-length coding in traditional block-DCT based video coding and the zero-tree coding in the wavelet based one, have already broken the independence among the coefficients. In consequence, the realised average bit rate in video codec is far less than $\overline{R}$ in (5). To evaluate the RD performance of video coding, (5) is written as

$$\log_2 \overline{D} = \log_2 \left(\frac{1}{12} \right) + 2\overline{H}_d - 2\overline{R} \,. \tag{6}$$

The above formulation implies that $\log_2 \overline{D}$ should have a linear relationship with $\overline{R}$ with a slope of -2. However, extensive experiments have revealed that the linear relationship does exist for the commonly used bit rates in SVC. Therefore the relationship between $\log_2 \overline{D}$ and $\overline{R}$ can be considered as piecewise linear, where the corresponding slope is different in different range. Experimental results using state-of-the-art SVC codecs are given in Fig.1. Here the advanced JSVM [8] and the wavelet based SVC codec aceSVC [9] were used. As depicted in Fig.1, the log distortion rate curves using both codecs are all with a slope of approximately –12. Therefore the RD model given in (6) is modified as

$$\overline{D} = \frac{1}{12} 2^{2\overline{H}_d} \cdot 2^{-\gamma \overline{R}} \,, \tag{7}$$

with γ is the corresponding slope. Empirically $\gamma = -12$ for $\overline{R} = 0.1 - 0.3$.



(a) (b)

Fig. 1 Log distortion versus bit rate. (a) JSVM; (b) aceSVC.

Specifically for 3-D wavelet based video coding, generalised RD models, such as the one in (7), are still inadequate to describe the distortion propagation effect inherent in the open-loop structure. As depicted in Fig.2, 3-D wavelet based SVC usually employs the motion-compensated temporal filtering (MCTF) with an open-loop. In an open-loop structure, the quantisation distortion in higher temporal levels will inevitably propagate to lower ones through MCTF. The distortion propagation problem has been thoroughly analysed in [6].

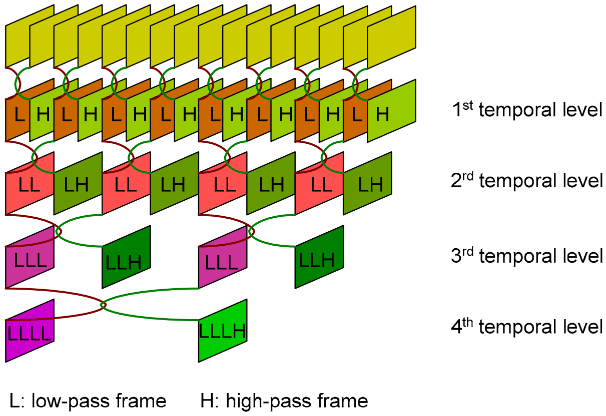


Fig. 2. Temporal decomposition using MCTF (Haar filtering).

They revealed a linear relationship between the average frame distortion within adjacent temporal levels. Accordingly the average distortion for the original video sequences after T levels temporal decomposition is as

$$\bar{d}_L^{(0)} = \sum_{t=1}^{T} B^{(t)} \prod_{j=1}^{t-1} A^{(j)} \bar{d}_H^{(t)} + \prod_{t=1}^{T} A^{(j)} \bar{d}_L^{(T)}, \tag{8}$$

where $\bar{d}_L^{(0)}$, $\bar{d}_H^{(t)}$ and $\bar{d}_L^{(T)}$ denote the average quantisation error of the zero temporal level, high pass frames in the t^{th} temporal level, and low pass frames in the highest temporal level, respectively. $A^{(t+1)}$, $B^{(t+1)}$ in (8) are parameters dependent on the motion estimation algorithm, the wavelet filter pair used for MCTF, and the video sequence characteristics. Generally, under the high resolution hypothesis one can always make the following assumption: $\bar{d}_H^{(t)} = \bar{d}_L^{(T)} = d_q$, where d_q is the average quantisation distortion. Therefore (8) can be approximated as

$$\bar{d}_L^{(0)} = \beta d_q. \tag{9}$$

Using the proposed model (7) to compute d_q and substituting (7) to (9) we obtain:

$$\bar{D}(R) = \frac{\beta}{12} 2^{2\bar{H}_d} 2^{-\gamma \bar{R}}. \tag{10}$$

Following [7], (10) can be further computed as

$$\bar{D}(R) = \beta \frac{\pi e}{6} \sigma^2 2^{-\gamma \bar{R}}, \tag{11}$$

where $\sigma^2 = \left(\prod_{m=0}^{N-1} \sigma_m^2 \right)^{1/N}$ with σ_m^2 the variance of each wavelet coefficient. Therefore it is easy to compute the Lagrange multiplier as:

$$\lambda = -\frac{dD}{dR} = -\beta \frac{\pi e}{6} \sigma^2 (-\gamma \ln \gamma) 2^{-\gamma R}. \tag{12}$$

However, in practice, it is very difficult to determine the standard deviation or the variance for a given coefficient, as required by (12). Previous studies suggest that the distribution of the transformed coefficients can be modelled by the Laplacian distribution [10]. Based on this observation the variance of the each coefficient can be estimated by $\sigma_m^2 = k_m \sigma_f^2$, where k_m is a parameter related to the wavelet transform, and σ_f^2 is the variance of the residual pixel values before wavelet transform. Moreover, using mean absolute difference (MAD) σ_f can be further approximated by $\sigma_f \approx \sqrt{2}MAD$ [10], resulting a more practical Lagrange multiplier expression as:

$$\lambda = -\frac{dD}{dR} = \alpha MAD 2^{-\gamma R}, \tag{13}$$

with $\alpha = 7.5$ an empirical value suitable for different sequences. The MAD can be obtained by pre-analysis of the sequence using motion compensation before encoding, or obtained and updated during encoding. As an effective measure to the video complexity, the MAD captures the characteristics of the sequence and maps them into the Lagrange multiplier selection. Therefore the proposed Lagrange multiplier selection is self-adaptive to the sequence content.

Correspondingly, if SAD is used as the distortion measure in motion estimation as in (4), the Lagrange multiplier for motion estimation is adjusted as

$$\lambda_{motion} = \sqrt{\alpha MAD 2^{-\gamma R}}. \tag{14}$$

Clearly, following (13) and (14), the optimal Lagrange multiplier can be determined once the target bit rate is available.

4. EXPERIMENTAL RESULTS

The performance of the proposed Lagrange multiplier selection has been extensively evaluated, using the 3-D wavelet based aceSVC [9] to generate anchors for comparison. It is noticeable that in our experiments the Lagrange multiplier in motion estimation (λ_{motion}) is considered. Therefore Lagrange multiplier is predicted using (14). For the sake of conciseness the results reported in this paper include four test sequences only: Foreman, Soccer, Crew and Mobile in CIF format. Similar results can be observed for other sequences.

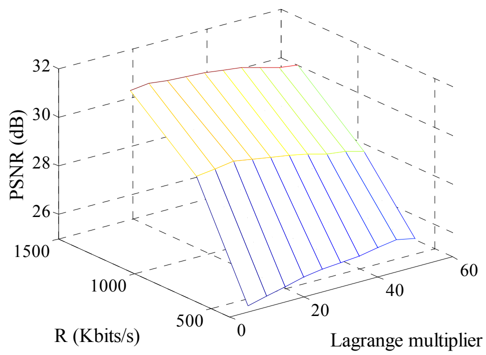
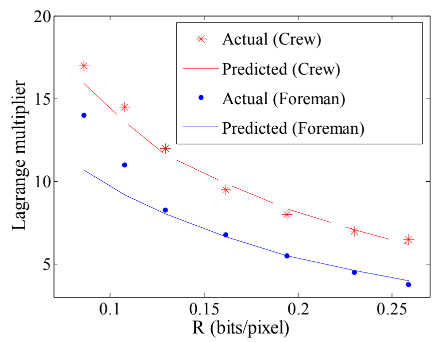
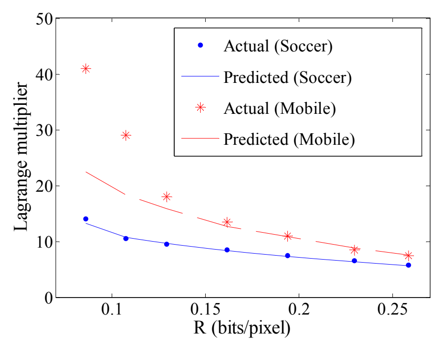


Fig. 3 RD performance with various Lagrange multiplier



(a)



(b)

Fig.4 Lagrange multiplier versus coding bit rate

Fig.3 shows the RD performance of "Mobile" with various Lagrange multiplier values. The observation indicates that the optimal value of the Lagrange multiplier varies with the targeted bit rates. The actual data of the optimal Lagrange multiplier values were experimentally obtained through exhaustive searching. The predicted values of the Lagrange multiplier were computed using (14). Fig. 4 illustrates the fitness of the proposed Lagrange multiplier with the actual data. It can be observed that the optimal Lagrange multiplier can be well approximated using model (14) for all sequences at the commonly used bit rates, which confirms the adaptability and accuracy of the proposed Lagrange multiplier selection. However, larger prediction errors can be observed for very low bit rate points. The observed mismatch is due to the RD model at low bit rates takes a different slope, which should result in a different γ. In addition, specified RD model for very low bit rates takes a different formulation from (5), which cannot be well captured by the piecewise linear model. This issue is currently under investigation.

5. CONCLUSION

In this paper a Lagrange multiplier selection for 3-D wavelet based SVC is proposed. The proposed method considers the RD performance of video coding based on the open-loop structure. Given the target bit rate, the optimal Lagrange multiplier can be accurately approximated at middle to high range of the commonly used bit rates. Considering the characteristics of the sequence, the proposed Lagrange multiplier is content adaptive. It is especially suitable for 3-D wavelet based SVC where quantisation steps are unavailable. Future work includes deriving a Lagrange multiplier selection targeting very low bit rates. Important issues such as Lagrange multiplier selection for a range of bit rates and the corresponding RD optimisation for 3-D wavelet based SVC are also the subject of further developments.

REFERENCE

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

[2] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control", in *Proc. IEEE Int. Conf. Image Processing 2001*, Thessaloniki, Greece, 2001, pp. 542-545.

[3] *"Advanced video coding for generic audiovisual services", ITU-T Rec. H.264- ISO/IEC 14496-10 (AVC),* Mar., 2005.

[4] M. Jiang, and N. Ling, "On Lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control", *IEEE Trans. on CSVT*, vol. 16, no. 5, pp. 663-669, May 2006.

[5] P. Sangi, J. Heikkila, and O. Silven, "Selection of the Lagrange multplier for block-based motion estimation criteria", in *Proc. IEEE ASSP*, Canada, 2004, pp. 325-328.

[6] M. Wang and M. van der Schaar, "Operational Rate-Distortion Modeling for Wavelet Video Coders". IEEE Trans. Signal Process., vol. 54, no. 9, pp. 3505–3516, Sept. 2006.

[7] S. Mallat and F. Falzon. "Analysis of low bit rate image transform coding". IEEE Trans. Signal Process., vol. 46, no. 4, pp. 1027–1042, Apr. 1998.

[8] ISO/IEC-JTC1 and ITU-T, "JSVM 3.0 Reference Encoding Algorithm Description", ISO/IEC JTC1/SC29/WG11, N7311, 73th MPEG Meeting, Poland, July 2005.

[9] M. Mrak, N. Sprljan,et.al. "Performance evidence of software proposal for wavelet video coding exploration group". ISO/IEC JTC1/SC29/WG11 M13146 2006, Switzerland.

[10] I. Pao and M. Sun, "Modeling DCT Coefficients for Fast Video Encoding", IEEE Trans. on CSVT, vol. 9, No.4, pp.608-616, June.1999.