PARAMETER SELECTION IN LAGRANGIAN HYBRID VIDEO CODER CONTROL

Thomas Wiegand

Image Processing Department
Heinrich Hertz Institute
Berlin, Germany
`wiegand@HHI.de`

Bernd Girod

Information Systems Laboratory
Stanford University
Stanford, CA, USA
`bgirod@stanford.edu`

ABSTRACT

The Lagrangian coder control together with the parameter choice is presented that lead to the creation of the new hybrid video coder specifications TMN-10 for H.263 and TML for H.26L. An efficient approach for the determination of the encoding parameters is developed. It is shown by means of experimental results that the Lagrange parameter for the macroblock mode decision corresponds to the negative slope of the distortion-rate curve of the prediction error coding. This distortion-rate curve is parameterized by the quantization parameter of the DCT coefficients motivating the established dependency with the Lagrange parameter.

1. INTRODUCTION

The most successful class of today's video compression schemes are called hybrid codecs. One key problem in hybrid video compression is the operational control of the coder. The task of coder control is to determine a set of coding parameters, and thereby the bit-stream, such that distortion is minimized for a given rate and a given decoder. It has been demonstrated in the past that coder control algorithms that are based on Lagrangian bit-allocation techniques provide excellent performance. For an overview see [1].

But, the application of Lagrangian techniques to control a hybrid video coder is not straightforward in practise because of the choice of the other involved parameters. A critical issue that is common to most standardized hybrid video codecs is the choice of the quantization parameter Q for the DCT coefficients in combination with the Lagrange parameter. Thus, in this paper, we present a simple and efficient method to resolve this problem. As a result of the performance of the Lagrangian coder control and parameter choice a new test model has been created TMN-10 [2]. TMN-10 is the recommended encoding approach of the ITU-T video compression standard H.263+ [3]. Furthermore, the TML [4], which is the test model for the ITU-T H.26L project, is based on the presented approach.

This paper is organized as follows. First, the investigated Lagrangian control for an H.263-based video coder is explained. Second, the experiment that leads to the proposed choice of coder control parameters is presented. Then, the relationship obtained for the Lagrange parameters and DCT quantizer is interpreted and a justification is given for the proposed scheme. Finally, the efficiency of the proposed scheme is empirically verified.

2. BIT ALLOCATION IN HYBRID VIDEO CODING

Hybrid video coding consists of the motion compensation and the residual coding stage. The task for residual coding is to represent signal parts that are not sufficiently compensated by motion coding. From the view-point of bit-allocation strategies, the various modes relate to various bit-rate partitions. Rate-constrained mode decision minimizes

$$D_{\text{REC}}(\boldsymbol{S}_k, I_k|Q) + \lambda_{\text{MODE}} \cdot R_{\text{REC}}(\boldsymbol{S}_k, I_k|Q), \qquad (1)$$

where the macroblock mode I_k is varied over the set $\{\text{INTRA}, \text{SKIP}, \text{INTER}, \text{INTER+4V}\}$. Rate $R_{\text{REC}}(\boldsymbol{S}_k, I_k|Q)$ and distortion $D_{\text{REC}}(\boldsymbol{S}_k, I_k|Q)$ for the various modes are computed as follows.

For the INTRA mode, the 8×8 blocks of the macroblock $\boldsymbol{S}_k$ are processed by a DCT and subsequent quantization. The distortion $D_{\text{REC}}(\boldsymbol{S}_k, \text{INTRA}|Q)$ is measured as the SSD between the reconstructed and the original macroblock pixels. The rate $R_{\text{REC}}(\boldsymbol{S}_k, \text{INTRA}|Q)$ is the rate that results after run-level variable-length coding.

For the SKIP mode, distortion $D_{\text{REC}}(\boldsymbol{S}_k, \text{SKIP})$ and rate $R_{\text{REC}}(\boldsymbol{S}_k, \text{SKIP})$ do not depend on the DCT quantizer value Q of the current picture. The distortion is determined by the SSD between the current picture and the previous coded picture for the macroblock pixels, and the rate is given as one bit per macroblock, as specified by ITU-T Recommendation H.263 [5].

The computation of the Lagrangian costs for the INTER and INTER+4V coding modes is much more demanding than for INTRA and SKIP. This is because of the block motion estimation step. The size of the blocks can be either 16×16 pixels for the INTER mode or 8×8 pixels for the INTER+4V mode. Let the Lagrange parameter λ_{MOTION} and the decoded reference picture s' be given. Rate-constrained motion estimation for a block $\boldsymbol{S}_i$ is conducted by minimizing the Lagrangian cost function

$$\boldsymbol{m}_i = \underset{\boldsymbol{m} \in \boldsymbol{\mathcal{M}}}{\operatorname{argmin}} \left\{ D_{\text{DFD}}(\boldsymbol{S}_i, \boldsymbol{m}) + \lambda_{\text{MOTION}} R_{\text{MOTION}}(\boldsymbol{S}_i, \boldsymbol{m}) \right\}, \qquad (2)$$

with the distortion term being given as

$$D_{\text{DFD}}(\boldsymbol{S}_i, \boldsymbol{m}) = \sum_{(x,y) \in \boldsymbol{\mathcal{A}}_i} |s[x,y,t] - s'[x-m_x, y-m_y, t-m_t]|^p \qquad (3)$$

wit $p = 1$ for the sum of absolute differences (SAD) and $p = 2$ for the sum of squared differences (SSD). $R_{\text{MOTION}}(\boldsymbol{S}_i, \boldsymbol{m})$ is the bit-rate required for the motion vector. The search range $\boldsymbol{\mathcal{M}}$ is ± 16 integer pixel positions horizontally and vertically and the prior decoded picture is referenced ($m_t = 1$). Depending on the use of SSD or SAD, the Lagrange parameter λ_{MOTION} has to be adjusted as discussed in the next section. The motion search that minimizes (2) proceeds first over integer-pixel locations. Then, the best of those integer-pixel motion vectors is tested whether one of the surrounding half-pixel positions provides a cost reduction in (2). This step is regarded as half-pixel refinement and yields the resulting motion vector $\boldsymbol{m}_i$. The resulting prediction error signal $u[x, y, t, \boldsymbol{m}_i]$ is similar to the INTRA mode processed by a DCT and subsequent quantization. The distortion D_{REC} is also measured as the SSD between the reconstructed and the original macroblock pixels. The rate R_{REC} is given as the sum of the bits for the motion vector and the bits for the quantized and run-level variable-length encoded DCT coefficients.

3. EXPERIMENTAL DETERMINATION OF THE CODER CONTROL PARAMETERS

The Lagrange parameter λ_{MODE} controls the macroblock mode decision when evaluating (1). The Lagrangian cost function in (1) depends for the INTER modes on the MCP signal and the DFD coding. The MCP signal is obtained by minimizing (2), which depends on the choice of λ_{MOTION}, while the DFD coding is controlled by the DCT quantizer value Q. Hence, for a fixed value of λ_{MODE}, a particular setting of λ_{MOTION} and Q yields a minimum Lagrangian cost function in (1). One approach to find those values for λ_{MOTION} and Q is to evaluate the product space of these two parameters. However, this approach requires a prohibitive amount of computation. Therefore, the relationship between λ_{MODE} and Q is considered first while fixing λ_{MOTION}. The parameter λ_{MOTION} is adjusted according to $\lambda_{\text{MOTION}} = \lambda_{\text{MODE}}$ when considering the SSD distortion measure in (2). This choice is motivated by theoretical [6] and experimental results that are presented later.

To obtain a relationship between Q and λ_{MODE}, the minimization of the Lagrangian cost function in (1) is extended by the macroblock mode type INTER+Q, which permits changing Q by a small amount when sending an INTER macroblock. More precisely, the macroblock mode decision is conducted by minimizing (1) over the set of macroblock modes

$$\{\text{INTRA}, \text{SKIP}, \text{INTER}, \text{INTER+4V}, \ldots$$
$$\text{INTER+Q}(-2), \text{INTER+Q}(-1), \text{INTER+Q}(1), \text{INTER+Q}(2)\},$$

where, for example, INTER+Q(-2) stands for the INTER macroblock mode being coded with DCT quantizer value reduced by two relative to the previous macroblock. Hence, the Q value selected by the minimization routine becomes dependent on λ_{MODE}. Otherwise the algorithm for running the rate-distortion optimized video coder remains unchanged.

Figure 1 shows the relative frequency of chosen macroblock quantizer values Q for several values of λ_{MODE}. The Lagrange parameter λ_{MODE} is varied over seven values: 4, 25, 100, 250, 400, 730, and 1000, producing seven normalized histograms for the chosen DCT quantizer value Q that are depicted in the plots in Fig. 1. In Fig. 1, the macroblock Q values are gathered while coding 100 frames of the video sequences *Foreman*, *Mobile & Calendar*, and *News*. The quantizer value Q does not vary much given a fixed value of λ_{MODE}. Moreover, as experimental results show, the gain when permitting the variation is rather small, indicating that fixing Q as in TMN-10 might be justified.

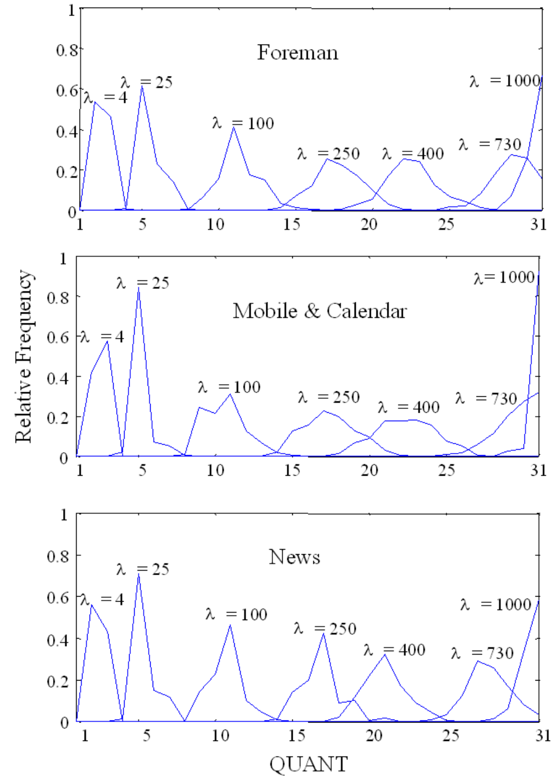

Fig. 1. Relative frequency vs. macroblock Q for various values of the Lagrange parameter λ_{MODE}. The relative frequencies of macroblock Q values are gathered while coding 100 frames of the video sequences *Foreman* (top), *Mobile & Calendar* (middle), and *News* (bottom).

As can already be seen from the histograms in Fig. 1, the peaks of the histograms are very similar among the four sequences and they are only dependent on the choice of λ_{MODE}. This observation can be confirmed by looking at the left-hand side of Fig. 2, where the average macroblock quantizer values Q from the histograms in Fig. 1 are shown. The bold curve in Fig. 2 depicts the function

$$\lambda_{\text{MODE}}(Q) \approx 0.85 \cdot Q^2\,, \qquad (4)$$

which is an approximation of the relationship between the macroblock quantizer value Q and the Lagrange parameter λ_{MODE} up to Q values of 25. H.263 allows only a choice of $Q \in \{1, 2, \ldots, 31\}$. In the next section, a motivation is given for the relationship between Q and λ_{MODE} in (4). Particularly remarkable is the strong dependency between λ_{MODE} and Q, even for sequences with widely varying content. Note, however, that for a given value of λ_{MODE}, the chosen Q tends to be higher for sequences which require higher amounts of bits (*Mobile & Calendar*) in comparison to sequences requiring smaller amounts of bits for coding at that particular λ_{MODE} (*News*) – but these differences are rather small.

4. INTERPRETATION OF THE LAGRANGE PARAMETER

The Lagrange parameter is regarded as the negative slope of the rate-distortion curve [7, 8, 9]. It is simple to show that if

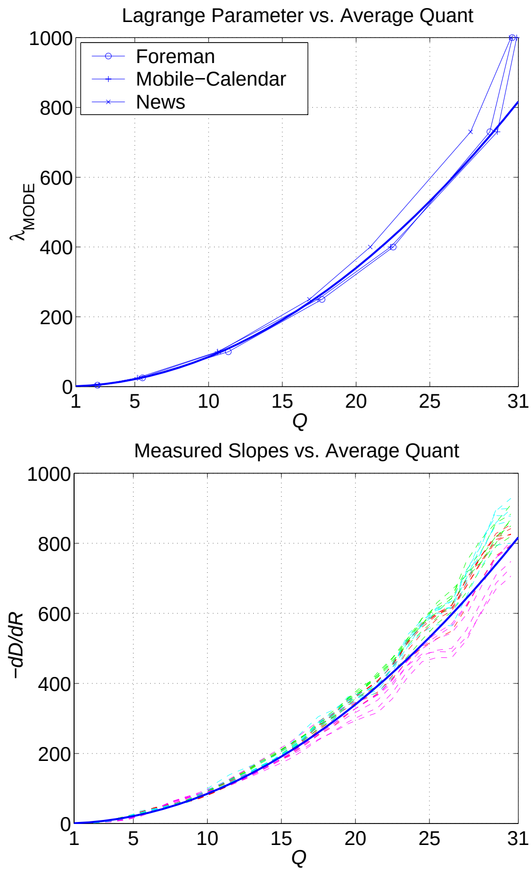

Fig. 2. Lagrange parameter λ_{MODE} vs. average macroblock Q (top) and measured slopes (bottom).

the distortion-rate function $D_{\text{REC}}(R_{\text{REC}})$ is strictly convex then $J_{\text{MODE}}(R_{\text{REC}}) = D_{\text{REC}}(R_{\text{REC}}) + \lambda_{\text{MODE}} R_{\text{REC}}$ is strictly convex as well. Assuming $D_{\text{REC}}(R_{\text{REC}})$ to be differentiable everywhere, the minimum of the Lagrangian cost function is given by setting its derivative to zero, i.e.

$$\frac{\mathrm{d}J_{\text{MODE}}}{\mathrm{d}R_{\text{REC}}} = \frac{\mathrm{d}D_{\text{REC}}}{\mathrm{d}R_{\text{REC}}} + \lambda_{\text{MODE}} \stackrel{!}{=} 0, \tag{5}$$

which yields

$$\lambda_{\text{MODE}} = -\frac{\mathrm{d}D_{\text{REC}}}{\mathrm{d}R_{\text{REC}}}. \tag{6}$$

A typical high-rate approximation curve for entropy-constrained scalar quantization can be written as [10]

$$R_{\text{REC}}(D_{\text{REC}}) = a \log_2 \left(\frac{b}{D_{\text{REC}}} \right), \tag{7}$$

with a and b parameterizing the functional relationship between rate and distortion. For the distortion-to-quantizer relation, it is assumed that at sufficiently high rates, the source probability distribution can be approximated as uniform within each quantization interval [11] yielding

$$D_{\text{REC}} = \frac{(2Q)^2}{12} = \frac{Q^2}{3}. \tag{8}$$

Note that in H.263 the macroblock quantizer value Q is approximately double the distance of the quantizer reproduction levels.

The total differentials of rate and distortion are given as

$$\begin{aligned} \mathrm{d}R_{\text{REC}} &= \frac{\partial R_{\text{REC}}}{\partial Q}\mathrm{d}Q = \frac{-2a}{Q \ln 2}\mathrm{d}Q \qquad \text{and} \\ \mathrm{d}D_{\text{REC}} &= \frac{\partial D_{\text{REC}}}{\partial Q}\mathrm{d}Q = \frac{2Q}{3}\mathrm{d}Q \end{aligned} \tag{9}$$

Plugging these into (6), provides the result

$$\lambda_{\text{MODE}}(Q) = -\frac{\mathrm{d}D_{\text{REC}}(Q)}{\mathrm{d}R_{\text{REC}}(Q)} = c \cdot Q^2 \tag{10}$$

where $c = \ln 2/(3a)$. Although the assumptions here may not be completely realistic, the derivation reveals at least the qualitative insight that it may be reasonable for the value of the Lagrange parameter λ_{MODE} to be proportional to the square of the quantizer value. As shown above by means of experimental results, 0.85 appears to be a reasonable value for use as the constant c.

For confirmation of the relationship in (10), an experiment has been conducted to measure the rate-distortion slopes $\mathrm{d}D_{\text{REC}}(Q)/\mathrm{d}R_{\text{REC}}(Q)$ for a given value of Q. The experiment consists of the following steps:

1. The hybrid video coder is run employing quantizer values $Q_{\text{REF}} \in \{4, 5, 7, 10, 15, 25\}$. The resulting bit-streams are decoded and the reconstructed frames are employed as reference frames in the next step.
2. Given the coded reference frames, the MCP signal is computed for a fixed value of

$$\lambda_{\text{MOTION}} = 0.85 \cdot Q_{\text{REF}}^2 \tag{11}$$

   when employing the SSD distortion measure in the minimization of (2). Here, only 16×16 blocks are utilized for half-pixel accurate motion compensation. The MCP signal is subtracted from the original signal providing the DFD signal that is further processed in the next step.
3. The DFD signal is encoded for each frame when varying the value of the DCT quantizer in the range $Q = \{1, \ldots, 31\}$ for the INTER macroblock mode. The other macroblock modes have been excluded here to avoid the macroblock mode decision that involves Lagrangian optimization using λ_{MODE}.
4. For each sequence and Q_{REF}, the distortion and rate values per frame including the motion vector bit-rate are averaged, and the slopes are computed numerically.

Via this procedure, the relationship between the DCT quantizer value Q and the negative slope of the distortion-rate curve has been obtained as shown on the right-hand side of Fig. 2. This experiment shows that the relationship in (10) can be measured using the rate-distortion curve for the DFD coding part of the hybrid video coder. This is in agreement with the experiment that is employed to establish (4).

5. EFFICIENCY EVALUATION FOR THE PARAMETER CHOICE

The choice of the encoding parameters has to be evaluated based on its effect on rate-distortion performance. Hence, in

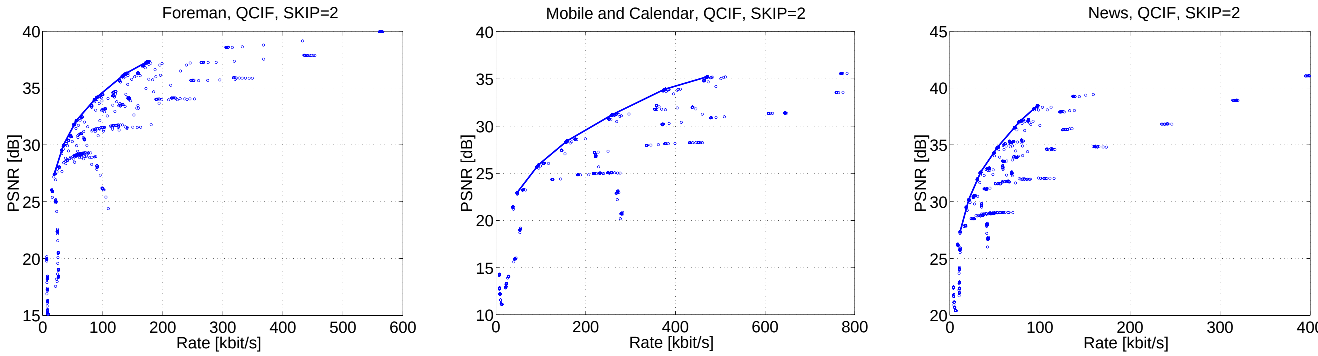

Fig. 3. PSNR in dB vs. bit-rate in kbit/s when running TMN-10 with various λ_{MODE}, λ_{MOTION}, and Q combinations for the video sequences *Foreman* (left), *Mobile & Calendar* (middle), and *News* (right).

order to verify that the particular choice of the relationship between λ_{MODE}, λ_{MOTION}, and Q provides good results in rate-distortion performance, the H.263+ coder is run using the TMN-10 algorithm for the product space of the parameter sets $\lambda_{\text{MODE}}, \lambda_{\text{MOTION}} \in \{0, 4, 14, 21, 42, 85, 191, 531, 1360, 8500\}$ and $Q \in \{4, 5, 7, 10, 15, 25\}$. For each of the 600 combinations of the three parameters, the sequences *Foreman*, *Mobile & Calendar*, and *News* are encoded, and the resulting average rate-distortion points are depicted in Fig. 3. The rate-distortion points obtained when setting $\lambda_{\text{MODE}} = \lambda_{\text{MOTION}} = 0.85Q^2$ are connected by the line in Fig. 3 and indicate that this setting indeed provides good results for all tested sequences. Although not shown here, it has been found that also for other sequences as well as other temporal and spatial resolutions, similar results can be obtained.

So far, SSD has been used as distortion measure for motion estimation. In case SAD is used for motion estimation, λ_{MOTION} is adjusted as

$$\lambda_{\text{MOTION}} = \sqrt{\lambda_{\text{MODE}}}. \tag{12}$$

Using this adjustment, experiments show that both distortion measures SSD and SAD provide very similar results.

6. CONCLUSIONS

The presented Lagrangian coder control together with the parameter choice has emerged as a practical and widely accepted optimization approach to hybrid video coding within the ITU-T. Thus, the encoder test models TMN-10 for H.263 and TML for H.26L have been created by the ITU-T Video Coding Experts Group based on this approach. In comparison to TMN-9 [12], the threshold-based predecessor of TMN-10, the overall performance gain of TMN-10 is typically between 5 and 10% in bit-rate savings when comparing at a fixed reconstruction quality of 34 dB PSNR.

The main contribution of this paper is an efficient approach for choosing the encoding parameters. The lack of such a method has been for a long time an obstacle for the consideration of Lagrangian coder control in practical systems. It is been shown by means of experimental results that the Lagrange parameter λ_{MODE} corresponds to the negative slope of the distortion-rate curve of the prediction error coding. This rate-distortion curve is parameterized by the quantization parameter of the DCT coefficients motivating the established dependency to the Lagrange parameter.

The strong dependency between Q, λ_{MODE}, and λ_{MOTION} offers a simple treatment of each of these quantities as a dependent variable of another. For example, the rate control method may adjust the macroblock quantizer value Q occasionally so as to control the average bit-rate of a video sequence, while treating λ_{MODE} and λ_{MOTION} as dependent variables using (4) for both or (12) for λ_{MOTION} in case the SAD is employed for motion estimation.

7. REFERENCES

[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] ITU-T/SG16/Q15-D-65, "Video Codec Test Model, Near Term, Version 10 (TMN-10), Draft 1", Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d65.doc, Apr. 1998.

[3] ITU-T Recommendation H.263 Version 2 (H.263+), "Video Coding for Low Bitrate Communication", Jan. 1998.

[4] ITU-T/SG16/Q15-K-59, "H.26L Test Model Long-Term Number 5 (TML-5)", Download via anonymous ftp to: standard.pictel.com/video-site/0008_Por/q15k59.doc, Oct. 2000.

[5] ITU-T Recommendation H.263, "Video Coding for Low Bitrate Communication", June 1996.

[6] B. Girod, "Rate-Constrained Motion Estimation", in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, Chicago, IL, USA, Sept. 1994, vol. 2308, pp. 1026–1034.

[7] H. Everett III, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources", *Operations Research*, vol. 11, pp. 399–417, 1963.

[8] Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.

[9] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[10] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1994.

[11] H. Gish and J. N. Pierce, "Asymptotically Efficient Quantizing", *IEEE Transactions on Information Theory*, vol. 14, pp. 676–683, Sept. 1968.

[12] ITU-T/SG16/Q15-C-15, "Video Codec Test Model Number 9 (TMN-9)", Download via anonymous ftp to: standard.pictel.com/video-site/9712_Eib/q15c15.doc, Dec. 1997.