

ON JOINT OPTIMIZATION OF MOTION COMPENSATION, QUANTIZATION AND BASELINE ENTROPY CODING IN H.264 WITH COMPLETE DECODER COMPATIBILITY

En-hui Yang and Xiang Yu

Dept. of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
ehyang@uwaterloo.ca x23yu@bbcr.uwaterloo.ca

ABSTRACT

This paper presents a framework for jointly designing motion compensation, quantization and entropy coding in a hybrid video coding structure to minimize a rate distortion cost. Given motion compensation, a soft decision-based quantization algorithm is first designed to reduce the rate distortion cost by adapting quantization outputs to the baseline entropy coding method in the newest standard H.264. Motion compensation is then optimized by searching for a prediction to further reduce the rate distortion cost based on given quantization outputs. By alternating these two steps, an iterative method is then proposed. The proposed algorithms have been implemented based on the reference encoder of H.264 with complete baseline decoder compatibility. Comparative studies show that the baseline-based iterative optimization method achieves coding performance comparable or sometimes superior to that afforded by the main profile encoder.

1. INTRODUCTION

Recently, rate-distortion (RD) theory has proved a great success in video coding applications [2, 4]. As the newest video coding standard, H.264/MPEG4 part10 achieves superior RD performance over earlier standards such as MPEG2, H.263, etc. by using RD optimization methods for motion compensation and mode selection[2].

The H.264 standard has a hybrid coding structure consisting of four functional elements, i.e., prediction, transform, quantization and entropy coding. Ideally, an objective function for RD optimization shall always be based on the final reproduction distortion and the entire bit rate. However, the computational cost for minimizing the actual RD cost is often too high for today's silicon technology[4]. It

involves in a search in a product space, which is often extremely large. Moreover, for a hybrid coding structure with hard decision quantization the computational cost to evaluate the actual RD cost is very high since it requires going through the entire coding procedure.

There have been many studies on this RD optimality vs. complexity issue[2, 4]. The most successful one that has won its way into the H.264 codec is to ignore the effect of residual coding on the prediction, which Wiegand et. al. [2] proposed and implemented in the H.264 reference encoder. Specifically, motion compensation is optimized based on the prediction error only, i.e.,

$$\arg \min_{\mathbf{v}} J_p = \|\mathbf{x} - \mathbf{p}(\mathbf{v})\| + \lambda \cdot R(\mathbf{v}), \quad (1)$$

where $\mathbf{x}$ is the pixel data, $\mathbf{p}$ is its prediction, $\mathbf{v}$ is the corresponding motion vector, and R is the coding rate. This method works fairly well as reported in [2]. However, the fact that it discards all the bit and distortion for the residual coding part suggests that there is still space for further optimization.

In this paper, we propose a framework for jointly designing motion compensation, quantization and entropy coding in the H.264 standard based on the actual RD cost. First an optimal soft decision quantization algorithm is designed to adapt quantization outputs corresponding to given motion prediction to the entropy coding method. Then, the motion prediction is updated by searching for a new prediction to further reduce the RD cost while fixing quantization outputs. By alternating these two steps, both motion compensation and residual coding are gradually improved in the sense of reducing the actual RD cost. Moreover, as to be presented later, the computational complexity is well managed to be affordable for applications such as digital broadcasting or multimedia distributions (optical disks, etc.).

H.264 supports two methods for coding the residual[5]: the baseline context adaptive variable length coding(CAVLC) and the main profile context adaptive binary arithmetic coding(CABAC). CABAC significantly outperforms CAVLC on the coding efficiency over a range of acceptable video

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants RGPIN203035-98 and RGPIN203035-02 and under Collaborative Research and Development Grant, by the Premier's Research Excellence Award, by the Canada Foundation for Innovation, by the Ontario Distinguished Researcher Award, and by the Canada Research Chairs Program.

quality of PSNR from 30dB to 38dB[7]. The soft decision quantization algorithm in this paper is designed based on the baseline CAVLC method, while CABAC is used for comparison. The idea of trading off a little extra distortion for a better coding efficiency has already been used in the H.264 reference codec. However, it is in an ad hoc way, e.g., a whole block is discarded under certain conditions. The soft decision quantization design here provides a systematic method for adapting quantization outputs to entropy coding for a better RD trade off.

The paper is organized as follows. A formal description of the RD optimization problem is presented in Section 2. An iterative solution is then proposed based on the soft decision quantization design and motion compensation optimization algorithms. Experimental results are presented in Section 4. Conclusions are drawn in Section 5.

2. FORMAL PROBLEM DEFINITION

The objective function for the RD optimization of the hybrid coding structure in H.264 is formed based on the final reproduction distortion and entire coding rate as follows,

$$J(m, \mathbf{v}, \mathbf{v}) = D(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \cdot (R(\mathbf{u}) + R(m) + R(\mathbf{v})), \quad (2)$$

where $\mathbf{x}$ is the original signal, $\hat{\mathbf{x}} = \mathbf{p}(m, \mathbf{v}) + \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u}))$ is its reconstruction, and $R(\mathbf{u})$, $R(m)$, and $R(\mathbf{v})$ are the bit rate for coding the quantized coefficients $\mathbf{u}$, prediction modes m , and motion vectors $\mathbf{v}$, respectively. Here $\mathbf{p}$ stands for prediction, $\mathbf{T}(\cdot)$ represents the DCT transform, and $\mathbf{Q}(\cdot)$ describes a scalar quantizer. Correspondingly, the optimization problem is,

$$\min_{m, \mathbf{v}, \mathbf{u}} J(m, \mathbf{v}, \mathbf{u}) = \|\mathbf{x} - \mathbf{p}(m, \mathbf{v}) - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u}))\|^2 + \lambda \cdot (R(\mathbf{u}) + R(m) + R(\mathbf{v})). \quad (3)$$

Optimization over m and $\mathbf{v}$ provides the optimal mode and motion prediction. For the conventional hard-decision quantizer, $\mathbf{u}$ is determined by $\mathbf{p}$, i.e., $\mathbf{u} = \mathbf{Q}(\mathbf{T}(\mathbf{x} - \mathbf{p}))$. The consideration of the somehow hidden parameter $\mathbf{u}$ as an optimization variable is the key point in the joint design. It leads to the so-called soft decision quantization design, which has displayed a significant gain in our work on optimizing the JPEG standard[3].

3. ALGORITHM DESIGN

The problem of (3) leads to a joint design of motion compensation, quantization and entropy coding in the hybrid video coding structure. To the best of our knowledge, this joint design has not been studied in the literature. To make the problem tractable, we propose an iterative method to optimize motion compensation and residual coding alternately. In the following, an optimal soft decision quantizer

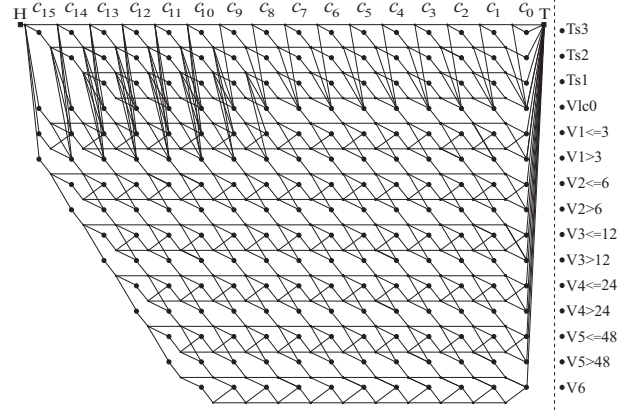


Fig. 1. The graph for quantizing $\mathbf{c}$. Each transition determines a $(run, level)$ pair. A path represents a possible sequence as the quantization output $\mathbf{u}$. Listed on the right are the states, named after the trailing one rule and 7 codes for coding levels. Transitions are built based on the run-length code and the table switching rules in CAVLC.

is first designed based on the baseline CAVLC method using a graph structure. Motion prediction is then updated based on the actual RD cost. By alternating these two steps, we get the iterative method, which solves the problem of (3) based on the actual RD cost of (2).

3.1. Graph-based soft decision quantization

A simple example is helpful to develop an intuition for the soft decision quantization. Consider a step size $q = 5$, inputs $\mathbf{c} = (84, 0, -8, 17, 0, -11, -8, 1)$ and the CAVLC method. Denote $\bar{\mathbf{c}} = \mathbf{c}/q = (16.8, 0, -1.6, 3.4, 0, -2.2, -1.6, 0.2)$. The quantization output given by the deterministic function $\text{round}(\cdot)$ is,

$$\mathbf{u} = (17, 0, -2, 3, 0, -2, -2, 0), R_{\text{CAVLC}}(\mathbf{u}) = 45\text{bits}.$$

On the other hand, a soft-decision based output can be,

$$\mathbf{u}' = (17, 0, -2, 4, 0, -2, -1, 0), R_{\text{CAVLC}}(\mathbf{u}') = 27\text{bits}.$$

Note that the quantization for a value of -1.6 may be either -2 or -1 , as $(\bar{c}_3 = -1.6, u'_3 = -2)$ and $(\bar{c}_7 = -1.6, u'_7 = -1)$. The soft decision trades off a little more distortion for a significant rate reduction for using CAVLC. Given $m, \mathbf{v}$ and the residual input of $\mathbf{x} - \mathbf{p}$, the soft decision quantizer works as

$$\mathbf{u} = \arg \min_{\mathbf{u}} J, \quad (4)$$

which is equivalent to,

$$\mathbf{u} = \arg \min_{\mathbf{u}} \|\mathbf{x} - \mathbf{p} - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u}))\|^2 + \lambda \cdot R(\mathbf{u}). \quad (5)$$

We now address the computation issue for the distortion term as it contains the inverse DCT transform. After some calculations, we have

$$\begin{aligned} D &= \|\mathbf{x} - \mathbf{p} - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u}))\|^2 \\ &= \|(\mathbf{c} - \mathbf{u} \cdot 2^{q_{per}}) \otimes \mathbf{a}[q_{rem}]/2^6\|^2, \end{aligned} \quad (6)$$

where $\mathbf{c}$ and $\mathbf{a}$ can be easily computed before the quantization process. Thanks to the fact that the DCT transform maintains the Euclidean distance, the distortion is now conveniently calculated in the DCT domain. The evaluation of (6) only consumes 2 integer multiplications per coefficient.

The problem of (5) is solved based on a graph structure as shown in Fig. 1. The graph is constructed to represent the vector space of the quantization output $\mathbf{u}$ for input $\mathbf{c}$, based on coding features of run length code, variable length codes selection, the trailing ones rule, etc(see [3] for details). The optimal soft decision quantization design then becomes a problem to search for a path in the graph for the minimal RD cost. It is not hard to see that the above graph design allows an additive computation of the RD cost in (5). The Viterbi algorithm is then used to do the search.

The proposed soft decision quantization somehow shares the same spirit with an entropy-constrained quantization design. In terms of using the Lagrangian method, the difference is that the soft decision quantization here is based on a fixed slope scheme[1], where there is not an explicit constraint on either the rate or the distortion. The fixed-slope method is generally superior to a fixed-rate or fixed-distortion method by its lower coding complexity[1].

3.2. Motion compensation optimization

For soft decision quantization, there is no deterministic relationship between $\mathbf{p}$ and $\mathbf{u}$ any more. For a given $\mathbf{u}$, it is thus applicable to find a new prediction $\mathbf{p}$ to minimize the actual RD cost, i.e., to update the motion vector $\mathbf{v}$ by,

$$\mathbf{v} = \arg \min_{\mathbf{v}} \|\mathbf{x} - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u})) - \mathbf{p}(m, \mathbf{v})\|^2 + \lambda \cdot R(\mathbf{v}),$$

because m and $\mathbf{u}$ are now considered fixed. Define $\mathbf{x}' = \mathbf{x} - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u}))$. We have

$$\mathbf{v} = \arg \min_{\mathbf{v}} \|\mathbf{x}' - \mathbf{p}(\mathbf{v})\|^2 + \lambda \cdot R(\mathbf{v}). \quad (7)$$

The beauty in (7) is that it can be solved using the same algorithm that has been developed in H.264 reference encoder for minimizing (1). For a given m and $\mathbf{u}$, $\mathbf{x}'$ can be easily computed. The optimization is then conducted as to search for a motion vector to match a 'new' block $\mathbf{x}'$. There is hardly any increase of computation, while the optimality over the actual RD cost has been achieved.

3.3. The joint optimization algorithm

Based on the above discussion of soft-decision quantization and motion vector updating, we now summarize our method to solve the joint optimization problem of (3).

1. Given mode m , initialize $\mathbf{v}$ by minimizing (1).
2. Soft-decision quantization. For given mode m and motion vector $\mathbf{v}$, compute $\mathbf{u}$ by solving (5).

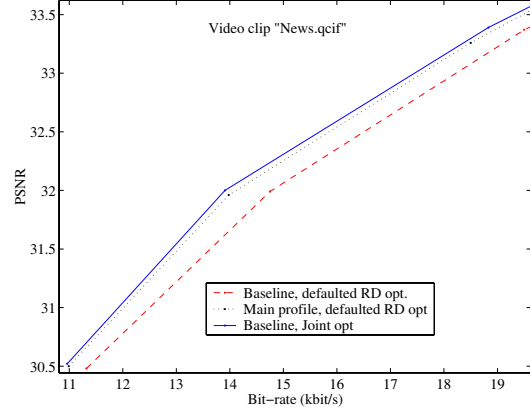


Fig. 2. RD performance for the video clip “News.qcif”.

3. Motion compensation optimization. For given m and $\mathbf{u}$, update $\mathbf{v}$ by solving (7).
4. Choose the optimal m by
$$m = \arg \min_m \|\mathbf{x} - \mathbf{T}^{-1}(\mathbf{Q}^{-1}(\mathbf{u})) - \mathbf{p}\|^2 + \lambda \cdot (R_{\text{VLC}}(\mathbf{v}) + R_{\text{VLC}}(m) + R_{\text{CAVLC}}(\mathbf{u})).$$
5. Repeat step 2, 3, and 4 until the decrease of J reaches a given threshold.

The proposed algorithm is guaranteed to converge because the actual RD cost is decreasing at each iteration step. From the fixed-slope optimization point of view, we shall fix the Lagrangian multiplier λ and optimize over the quantization step size to further reduce the RD cost. However, users would prefer using the quantization step size q_p to control the codec. For simplicity we give up the minor gain and initialize λ from q_p by using an equation of $\lambda = 0.85 \cdot 2^{(q_p-12)/3}$, which was discovered in empirical studies of [2, 6]. Experiment results show that this formula works well for our optimization algorithm.

The computational complexity of the proposed iterative algorithm comes from three parts. Motion compensation and mode selection hardly cause any increase compared to the method by Wiegand et. al. [2]. The main extra computational cost results from the soft-decision quantization, which is basically a path search in a graph shown in Fig. 1 using the Viterbi algorithm. It turns out that the computational cost for one iteration is just slightly more than the coding process in the reference code.

4. EXPERIMENTAL RESULTS

Simulations have been conducted over a range of typical video sequences. The optimization algorithm is implemented based on the H.264 reference software Jm82. Only the first frame is intra coded (I-frame), while all the subsequent frames use temporal prediction (P-frame). B frame is not used since

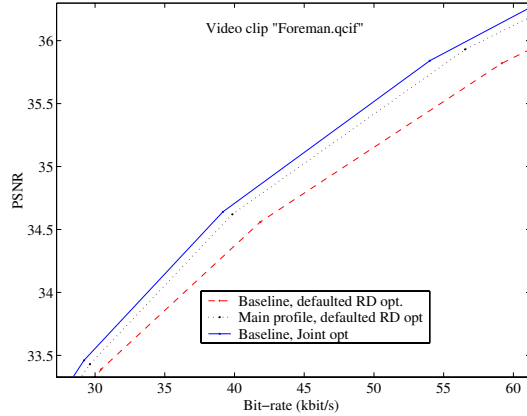


Fig. 3. RD performance for the video clip “Foreman.qcif”.

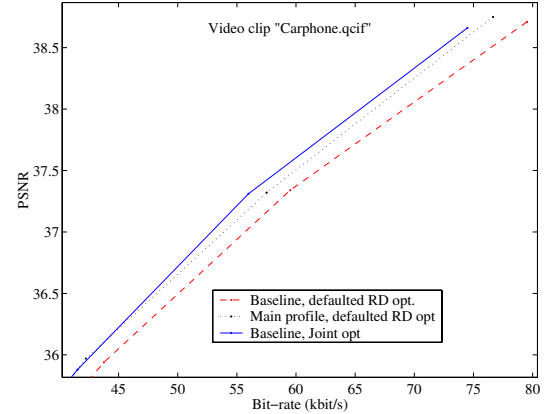


Fig. 4. RD performance for the video clip “Carphone.qcif”.

we target baseline decoder compatibility. The range for full-pixel motion compensation is ± 32 . The iteration is stopped when the RD cost decrease is less than 1%. By using the solution from (1) as initialization, it is observed to stop after 2 loops in most cases.

Figures 2, 3 and 4 show the RD curves for the luminance component of various video clips. The dashed line accords to a baseline codec using CAVLC; the dotted line stands for a codec using main profile CABAC. The defaulted RD optimization in the codec is enabled in both cases. The solid line shows the result of the proposed joint optimization method based on the baseline CAVLC. Experimental results show that the joint optimization method successfully improves the coding efficiency of the baseline codec to the level of the codec using main profile method. Theoretically, the advantage of the main profile CABAC over the baseline CAVLC comes from its adaptability to the symbol statistics and its ability to use a noninteger code length. The fundamental 1bit/symbol limit on the variable length code leads to a poor coding performance for CAVLC when the symbol probability is large, e.g., greater than 0.5. It is shown that this fundamental deficit of CAVLC to CABAC has been well compensated when we tune up the whole system with the joint optimization.

5. CONCLUSIONS

In this paper, we proposed a framework for joint optimization of motion compensation, quantization and entropy coding in a hybrid video coding structure. The proposed method has been applied to improve the coding efficiency of the baseline codec for the H.264 standard.

Experimental results show that the baseline-based optimization method achieves very close coding performance at average with that of an encoder using main profile CABAC. Because optimization is conducted only to the encoding process, the decoder enjoys both the computational efficiency

of the baseline CAVLC and the coding efficiency, resulting in a faster and cheaper access to the video content, yet with similar quality. It shows a great potential to open the market for baseline products.

In addition, the joint optimization framework is by no means restricted to the baseline entropy coding method. Information theoretical analysis shows that a similar gain can be expected by designing soft decision quantization based on the main profile CABAC. We are currently working on this. Results will be delivered soon.

6. REFERENCES

- [1] E.-h. Yang, Z. Zhang, and T. Berger, “Fixed-slope Universal Lossy Data Compression,” *IEEE Transactions on Information Theory*, Vol. 43, No. 5, pp.1465-1476, September 1997.
- [2] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, “Rate-Constrained Coder Control and Comparison of Video Coding Standards”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp.688-703, July 2003.
- [3] E.-h. Yang and L. Wang, “Joint Optimization of Run-length Coding, Huffman Coding and Quantization table with Complete baseline JPEG Decoder Compatibility,” to be published.
- [4] G. J. Sullivan and T. Wiegand, “Rate-Distortion Optimization for Video Compression,” *IEEE Signal Processing Magazine*, pp.74-90, November 1998.
- [5] T. Wiegand, G. J. Sullivan and A. Luthra, “Draft ITU-T Rec. H.264/ISO/IEC 14496-10 AVC,” Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, Doc. JVT-G050r1, 2003.
- [6] T. Wiegand and B. Girod, “Lagrangian Multiplier Selection in Hybrid Video Coder Control,” in *Proceedings of ICIP’2001*, pp.542-545, October 2001.
- [7] D. Marpe, H. Schwarz, and T. Wiegand, “Context-based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard”, *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 620-636, July 2003