Rate-Distortion Optimized Multipass Video Encoding with Application to MPEG-4

Nikolaus Färber, Hussain Mohammed, and Herbert Thoma
Fraunhofer Institute for Integrierted Circuits
Am Wolfsmantel 33
91058 Erlangen, Germany
{fae, hussainmh, tma}@iis.fraunhofer.de

Abstract— **In this paper we propose a rate-distortion optimized multipass algorithm for video encoding. The problem of dependent quantization in a hybrid coding framework is addressed by constructing a trellis from the exponentially growing search tree which is then searched using a Lagrangian cost function. To avoid exponential growth, states with similarly distorted reference frames are merged based on the assumption that similar rate-distortion behavior will result for future frames. The goal is to select an optimal quantizer for each frame such that the overall R-D performance is optimized. The algorithm is applied to MPEG-4 SP and typical average PSNR gains of 0.3 dB are achieved. A comparison with the DivX 5.2 codec shows PSNR gains of up to 2 dB.**

I. INTRODUCTION

The hybrid coding framework employed in all current video coding standards, such as MPEG-2, MPEG-4 or H.264/AVC, makes it very difficult to apply the optimization over time, i.e. considering several subsequent frames jointly. The fact that decisions in the current frame have significant influence on the Rate-Distortion (*R-D*) performance of future frames results in a dependent coding framework with an exponentially growing search space. Hence, *R-D* optimization is typically done on a frame-by-frame basis [1,2].

One approach to consider not only the current frame but the overall characteristic of a sequence is multipass encoding. In a first encoding pass, statistics of the sequence are collected which are then analyzed to optimize the second pass. The results from the second pass may then be used for a third pass, and so forth. Even though multipass codecs usually help to distribute the available bits more intelligently across the sequence, they are usually not *R-D* optimized.

In this paper we combine multipass encoding and *R-D* optimization for video encoding. The goal is to select an optimal quantization parameter (*Q*) for each frame such that the overall *R-D* performance is optimized. Like for all multipass codecs, a higher variation in bit rate and increased complexity has to be accepted. However, many applications, ranging from MMS (Multimedia Messaging Service) to DVD (Digital Versatile Disk) can accept these constraints because encoding is done off-line and only once.

This paper is structured as follows. In Section II we formulate the problem and describe the algorithm in Section III. In Section IV we provide coding results applying the algorithm to MPEG-4 SP. In this section we also compare our results to DivX 5.2 as another common multipass codec.

II. PROBLEM FORMULATION

Because we target the problem of dependent encoding of predicted frames (P-frames) we ignore bi-directional predicted frames (B-frames) and do not enforce Intra-frames (I-frames) during encoding. I.e. we focus on the encoding pattern IPPP.

In the following we denote the rate and distortion in frame *i* as $R(Q1,Q2,...,Qi)$ and $D(Q1,Q2,...,Qi)$ respectively, where *Qi* denotes the quantizer used for frame *i*. *R* corresponds to the number of bits in frame *i* and *D* is measured as the mean squared error (MSE). Note that all previous *Q*s need to be specified because of the dependency. As an example, in frame #2, we have to be aware that $R(1,Q2) \neq R(10,Q2)$ and $D(1,Q2) \neq D(10,Q2)$ because the quality of the first frame ($Q1$=1 vs. $Q1$=10) has a significant influence on the *R-D* performance of the 2nd frame. If we want to refer to *R* and *D* jointly or refer to the encoding state we use the notation $(Q1,Q2,...,Qi)$.

Our optimization task involves the optimal selection of quantizers *Qi** for each frame *i*, under a rate constrained R_{max}, i.e.,

$$\min_{Q1,Q2,...QN}[D(Q1)+D(Q1,Q2)+...+D(Q1,Q2,...QN)]$$

under the constraint that

$$R(Q1)+R(Q1,Q2)+...+R(Q1,Q2,...,QN)<R_{max}$$

As shown in the literature (e.g. [3]) this constrained optimization problem can be solved by the equivalent unconstrained problem

$$\min_{Q1,Q2,...QN}[J(Q1)+J(Q1,Q2)+...+J(Q1,Q2,...,QN)] \quad (1)$$

where

$$J(Q1,Q2,...QN) = D(Q1,Q2,...QN) + \lambda R(Q1,Q2,...QN)$$

is the Lagrangian cost and $\lambda \geq 0$ the Lagrange multiplier which is used to select the desired operating point, i.e. the trade-off between rate and distortion.

0-7803-9134-9/05/$20.00 ©2005 IEEE

The problem of solving (1) lies in the exponential growth of the search range. If M different quantizers are considered for each frame then M^N combinations have to be evaluated for a sequence of N frames. In particular the data collection phase, i.e. encoding the sequence M^N times, results in prohibitive complexity.

III. RD-Optimized Multipass Algorithm

Before the description of the proposed multipass algorithm, an intuitive explanation of the desired goal shall be discussed. As an example, consider a cut to an almost static scene. If the 1st frame after the scene cut is encoded with very high quality then all following frames can take advantage of this encoding decision since they can simply copy from this high quality frame for a long time. Hence, a very low Q should be selected. However, if the scene after the cut contains complex motion and detailed texture, the situation is different. Then bits spent in the 1st frame will not have an equally positive effect on future frames. The algorithm we propose will automatically detect such situations by evaluating the long-term *R-D* trade-off and thus select the optimal Q for each frame.

The *R-D* optimized multipass algorithm (RDM) comprises three steps. The first step is the data collection in which the sequence is encoded several times using fixed Qs. In the second step, a trellis is constructed from the collected *R-D* data and the optimal path (i.e. Qi^*) is selected using a Lagrangian cost. This is the actual optimization step. Then, in the last step, the final encoding is performed using the obtained Qi^* for each frame as obtained from the optimization step.

Since the final encoding step is obvious, only steps #1 and #2 are explained in more detail in the following. For illustration we use a simple ("sand-box") example where N=3 frames are encoded using M=3 different quantizer values, $Qi \in \{2,4,6\}$.

A. Data Collection Step

In this step, the sequence is encoded M times using a fixed quantizer $Qi \in \{q1,q2,...qM\}$ for all N frames. Hence, in the first pass the data points *(q1)*, *(q1,q1)*, … are generated and in the Mth pass the data points *(qM)*, *(qM,qM)*,… . This corresponds to a fixed quantizer encoding and sets the baseline for the multipass algorithm. By optimizing the quantizers we expect to be better than any of the M fixed quantizer encodings.

Since we want to allow a change in Q, we need to measure additional data. In order to decide whether a change to a new Q is better than staying with the fixed Q, we need to know what rate and distortion this change would result in. Hence, while doing the fixed quantizer runs, we encode the residual also with the other Qs and measure R and D. However, to avoid exponential growth, the final encoding is always done with the fixed Q for that pass (note that motion estimation, which is most complex in encoding, has to be done only once). In summary, all collected data points can be described as

(a,a,...,a,b), with $(a,b) \in \{q1,q2,...qM\}$

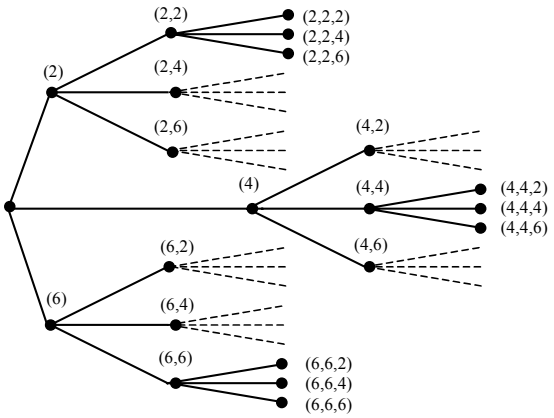

Figure 1. Subset of measured data points

Fig.1 illustrates the complete exponential search tree of the above mentioned example with the subset of actually measured data indicated as bold dots. Note that the complexity has been reduced from M^N to M^2.

B. Optimization Step

The optimization step consists of two sub-steps. In the first sub-step, a trellis is constructed from the collected data points. This sub-step is described in more detail below and constitutes the main contribution of this work. In the second sub-step, the generated trellis is searched using the Viterbi algorithm and a Lagrangian cost. This step involves standard optimization techniques and is not described in detail. For more information and references to further literature see [3] and [1].

The trellis we construct has M *main branches* which correspond to the fixed Q encodings. The trellis grows from left to right as more frames are encoded and the corresponding states along the main branch are denoted *(a,a,...,a)*. Each branch connecting two states is characterized by the Q used to get to the next state and the resulting R and D for that frame.

The additional data points *(a,a,...,a,b)*, $a \neq b$, are also added to the trellis. Hence, from each state on a main branch we have M-1 *transition branches* leaving this main branch. This is illustrated in Fig. 2 for the above example. Note that the constructed trellis is identical to the sub-tree illustrated in Fig. 1 – just arranged differently.

From Fig. 2 it can be seen that all transition branches are "dead ends", i.e. they are not connected to any other state. Hence, the trellis in its current form cannot be used for optimization since no change in quantizer is possible.

At this point we introduce a key assumption that allows to construct an approximated trellis. We assume that the R-D performance of future frames does not depend so much on the exact sequence of the encoding but mainly on the quality of the previously reconstructed reference frame. In other words, if two different encoding paths result in similar D, i.e. $D(Q1,Q2,...Qn) \approx D(Q1',Q2',...Qn')$, then the *R-D* behavior in the following frames $k > n$ will also be similar.

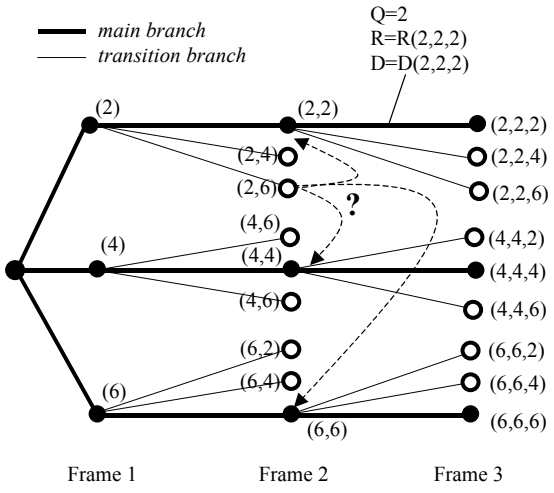

Figure 2. Sub-tree rearranged as trellis

Even though this assumption can be justified intuitively, we are aware that it may not hold in general. Hence, we do not claim true optimality but restrict our claim to optimality under the given assumption.

Based on the *similar distortion assumption* we merge states with similar distortion in order to obtain a connected trellis. The states from the main branches are used as reference points and the transition branches leaving the main branch are merged to the closest state on one of the main branches – using the distortion as a similarity measure. This is illustrated in Fig. 2 for the state (2,6) by the dotted arrows. $D(2,6)$ is compared to $D(2,2)$, $D(4,4)$ and $D(6,6)$ and the closest value determines where the transition branch is merged to. This merging operation is repeated for each transition branch.

The resulting trellis is illustrated in Fig. 3. Note that a pair of states may actually be connected by more than one branch. These branches correspond to different Qs and have different associated *R-D* values. In general, the structure of the trellis is signal dependent and irregular. Each branch has associated Q, R, and D values.

In the final optimization step a Viterbi algorithm is used to find the optimum path through the constructed trellis. As an optimization criterion, the accumulated Lagrangian cost until frame i is used. Note that the Viterbi algorithm has to be run with a certain Lagrange multiplier λ. Hence, if a certain rate constraint R_{max} has to be achieved using a convex hull search for the appropriate λ, then the Viterbi algorithm has to be applied several times. However, compared to the data collection step, this results only in a minor complexity increase. After finding the minimum J, back tracking is used to find the optimal path, i.e. the quantizer values Qi^*.

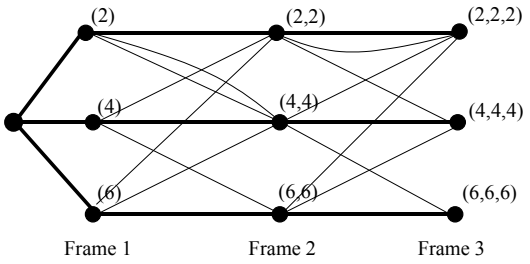

Figure 3. Resulting trellis after merging transition branches

IV. Implementation Results

The RDM algorithm can be applied for any hybrid video codec and is assessed in the following for MPEG-4 Simple Profile (SP).

In a first experiment we use the first N=3 frames of the Foreman sequence (QCIF, 15 fps) and allow M=8 quantizer parameters for each frame. For this limited example, it is still possible to encode all $M^N = 512$ combinations which are illustrated as *R-D* points in Fig. 4. The points enclosed in a box are selected by the RDM algorithm for various values of λ. As can be seen, the selected points are close to the convex hull of the data set and therefore close to the optimal solutions.

In a second experiment, four standard test sequences (Foreman, Paris, Football, Tempete) and three additional sequences (CNN, Spiderman, Red-October) are encoded in QCIF resolution at 15 fps. The three additional sequences include scene cuts and scenes with different complexity where multipass algorithms show their actual strength. We use the Fraunhofer-IIS MPEG-4 SP video codec for encoding which is also used in MPEG-4 and 3GPP as the baseline codec for evaluation [4, 5]. It exploits *R-D* optimized mode decision and has shown to provide state of the art *R-D* performance [6].

Fig. 5 illustrates typical results for two selected test sequences where the results for RDM ('*') are compared to the case where a fixed quantizer is used for the whole sequence ('o'). Standard sequences (e.g. Foreman, Fig. 5 left) provide little opportunity for multipass optimization since a single scene with time invariant complexity is encoded. Hence, the multipass results are almost identical to fixed quantizer results.

For sequences with scene cuts and time variant complexity, multipass encoding proves its usefulness. One such sequence is the CNN sequence, where an average PSNR gain of 0.74 dB can be achieved (see Fig. 5 right). It should be noted that the subjective quality is often improved more than indicated by the PSNR gains. Especially during scene cuts and fades, RDM provides improved subjective quality by spending bits more intelligently, i.e., where they are useful in the future encoding process. Hence, we found that minimizing MSE over the whole sequence has a positive effect on subjective quality.

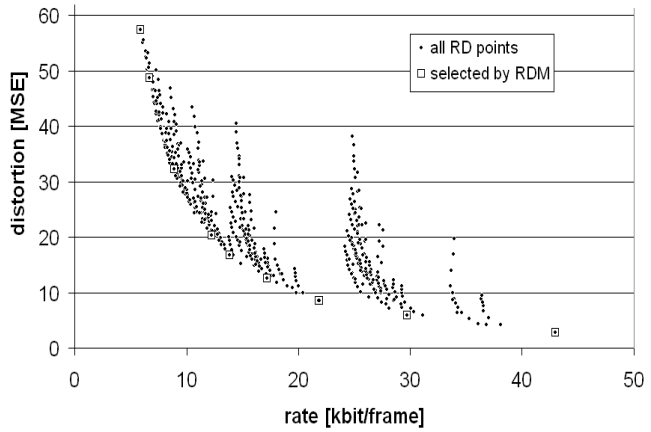

Figure 3. Complete set of *R-D* points for Foreman (*N*=3, *M*=8) and *R-D* points selected by RDM (boxes).

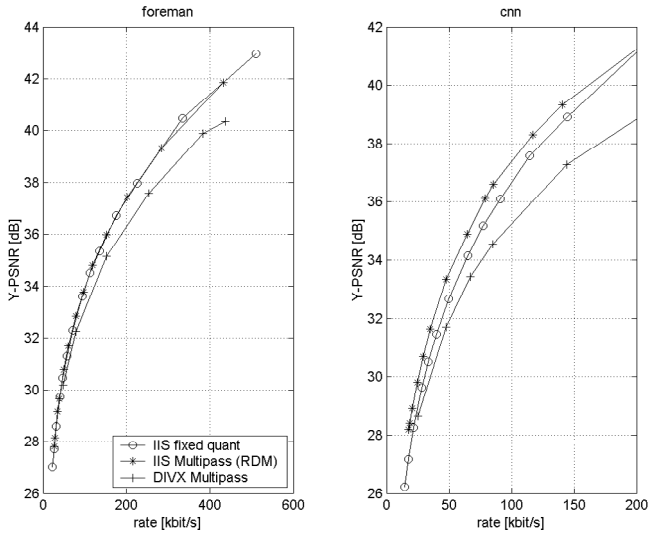

Figure 5. *R-D*-Performance of fixed quant vs. multipass and IIS vs. DivX for two test sequences (Foreman, CNN)

Fig. 5 also compares RDM results ('*') with the DivX 5.2 codec ('+'). Note that both codecs use identical coding tools and can be decoded with the same MPEG-4 SP compliant decoder. Furthermore, both encoders use multipass encoding and *R-D* optimized mode decision. The average PSNR gain is 0.89 dB and 1.92 dB for the foreman and CNN sequence, respectively. Further results for all test sequences are summarized in Table 1. The typical gain for sequences with time variant complexity (CNN, Spiderman, Red October) is 0.47 dB.

TABLE 1.
AVERAGE PSNR GAIN [DB] FOR ALL TEST SEQUENCES

Seq. name	# frames (*N*)	IIS RDM vs. IIS fixed Q	IIS RDM vs. DivX 5.2
Foreman	150	0.09	0.89
Paris	150	0.26	1.17
Football	125	0.02	0.63
Tempete	125	0.01	0.53
CNN	1000	0.74	1.92
Spiderman	1000	0.33	0.66
Red October	1000	0.34	0.65
average	NA	0.26	0.92

It should be noted that the DivX codec may not use MSE as the only optimization criteria. Hence, the presented comparison must be treated with caution. However, we found that high PSNR gains correlate very well with subjective quality. Furthermore, we believe that PSNR is working well as an objective quality measure as long as it is used with the same codec and sequence – which is both the case here.

V. CONCLUSIONS

In this paper, we propose the RDM algorithm for video encoding that provides improved *R-D* performance by multipass encoding. We show how the exponential search tree can be converted to a linear trellis which makes *R-D* optimization feasible. This is achieved by merging states with similar distortion under the assumption that similar *R-D* performance will result in future frames. The application of RDM to MPEG-4 SP shows a typical average PSNR gain of 0.3 dB. Compared to DivX 5.2 a gain of up to 2 dB is achieved.

Future work will investigate the application of RDM to H.264/AVC where the use of multiframe motion compensation requires an extension of the similar distortion assumption. Other issues that need to be investigated further are the optimization of quantizers for individual macroblocks (in contrast to the whole frame) and strategies for reducing the complexity of the data collection phase.

REFERENCES

[1] A. Ortega, K. Ramchandran, and M. Vetterli, "Bit Allocation for Dependent Quantization with Applications to Multiresolution and MPEG Video Coders," IEEE Transactions on Image Processing, vol. 3, No. 5, Sep. 1994.

[2] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression", IEEE Signal Processing magazine, pp. 74-90, Nov. 1998.

[3] A. Ortega and K. Ramchandran, "Rate-Distortion Methods for Image and Video Compression", IEEE Signal Processing magazine, pp. 23-50, Nov. 1998.

[4] MPEG Doc. N6231, "Report of The Formal Verification Tests on AVC", Waikoloa, Dec. 2003.

[5] 3GPP Doc. S4-030718, "Test Material and Reference Results for Video Codec Candidate Qualification Criteria", Tampere, Nov. 2003.

[6] http://www.iis.fraunhofer.de/amm/download/wp_iismpeg4videosoftware.pdf