IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Model-based Transrating of H.264 Coded Video

Naama Hait and David Malah

CCIT Report #713
December 2008

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #713 December 2008

Model-based Transrating of H.264 Coded Video

Naama Hait and David Malah

Abstract

This paper presents a model-based transrating (bit-rate reduction) system for H.264 coded video via requantization. In works related to previous standards, optimal requantization step-sizes were obtained via Lagrangian optimization that minimizes the distortion subject to a rate constraint. Due to H.264 advanced coding features, the choices of quantization step-size and coding modes are dependent and the rate control becomes computationally expensive. Therefore, optimal requantization algorithms developed for previous standards cannot be applied as is. Hence, previous works on transrating in H.264 focused on changing the input coding decisions rather on rate control, while requantization was addressed by a simple one-pass algorithm.

Here we propose new model-based optimal requantization algorithms for transrating of H.264 coded video. The optimal requantization goal is to achieve the target bit rate with minimal effect on video quality. Incorporation of the proposed models serves two goals. For intra-coded frames, a novel closed-loop statistical estimator that overcomes spatial neighbors dependencies is developed. For inter-coded frames, the proposed macroblock-level models reduce the computational burden of the optimization. Overall, as compared to re-encoding (cascaded decoder-encoder), the proposed system reduces the computational complexity by a factor of about 4, at an average PSNR loss of only 0.4[dB] for transrating CIF/SIF sequences from 2[Mbps] to 1[Mbps]. In comparison with a simple one-pass requantization, the proposed algorithm achieves better performance (an average PSNR gain of 0.45[dB]), at the cost of just twice the complexity.

Index Terms

Bit rate control, H.264 video coder, requantization, transrating.

I. Introduction

Video services and multimedia applications use pre-encoded video in different formats for storage and transmission. As various user types require different formats and bit rates, a single copy of the encoded video cannot satisfy all users. One could store many copies of the video in the server, each encoded at a different format or bit rate, and send the bitstream that best matches the requirements of the user. However, such a server would suffer from very high storage costs and the chosen bitstream may not meet the exact user requirements. Therefore, servers typically

This work was supported in part by STRIMM consortium under the MAGNET program of the Ministry of Trade and Industry via the Samuel Neaman Institute.

store a single copy, pre-encoded at a high quality, and convert (or transcode) it on-line to match user-specific requirements. Transrating, which refers to bit rate reduction within the same video format, can be achieved by a number of methods, such as frame rate reduction, spatial resolution reduction and requantization of the transform coefficients. In this paper, we examine model-based transrating via requantization of the transform coefficients, for the state of the art H.264 video coder.

Optimal requantization for MPEG-2 encoded video was suggested in [1] by minimizing the frame's distortion subject to its target bit rate. In that work, the optimization procedure became an expensive exhaustive search since it evaluated the rates and the distortions for each picture region (e.g. a macroblock) at multiple requantization steps exhaustively, with no models. Previous works that did use analytic models for optimal bit allocation [2], [3], aimed at encoding the *original input* video, using earlier video coding standards.

H.264 is currently the state of the art video coding standard. Its advanced coding features offer an improvement in the coding efficiency by a factor of about two over MPEG-2 [4], at the expense of higher complexity. As the choices of quantization step-size and coding modes are dependent, the rate control becomes computationally expensive. Therefore, previous works on transrating in H.264 [5], [6], [7] focus on changing the input coding decisions (intra prediction modes and motion) rather on the rate control, and requantization is addressed by a simple one-pass algorithm [5].

In this paper, new model-based optimal requantization algorithms for transrating of H.264 coded video are developed and examined. The models incorporated in this work relate the rate and the distortion to the fraction of zeroed quantized transform coefficients, ρ [8], rather than to the step-size itself. At first, frame-level bit allocation is determined by minimizing the overall distortion over a group of frames, such that the target average bit rate is achieved. To keep a smooth constant video quality, the frame distortions are equalized. This step follows by requantization of the intra and inter frames, separately.

For *intra-coded* frames, requantization gets complicated because of the spatial prediction used in H.264 for these frames, which introduces dependencies between neighboring residual blocks. Due to these dependencies, the residual coefficients to be requantized are not available when needed for requantization step-size selection. Therefore, the estimation of the relation between ρ and the requantization step-size becomes a challenging task. To this end, we propose a novel closed-loop statistical estimator, which outperforms the simple open-loop estimator.

For *inter-coded* frames, we propose to solve an optimal nonuniform requantization problem. The requantization step-size for each macroblock is chosen such that the overall frame distortion is minimized subject to a rate constraint and a limitation of the change in the requantization step-size in consecutive macroblocks that helps to improve the subjective quality. To solve that regularized optimization problem, we suggest to extend the Lagrangian optimization (see [1]) by an inner loop that applies dynamic programming. To reduce the computational burden of the optimization, we use rate-distortion models at the macroblock level. As the models suggested in the literature are not suitable for macroblock level coding in H.264, we develope macroblock level rate-distortion models adapted to H.264 requantization. Since the recommended software encoder [9] eliminates very sparse blocks, we also examine the option of extending the optimal requantization by selective coeffcient elimination. In addition, we incorporated some HVS based considerations in the system design to gain a higher perceptual quality, as a secondary focus of the work. Partial details and preliminary results were reported in [10], [11], dealing with transrating of intra-coded and inter-coded frames, respectively. This paper describes in full the complete proposed transrating system, including the final algorithms and overall system performance evaluation.

The following subsection, I-A, provides a short overview of existing ρ-domain models. Subsection I-B discusses the chosen transrating architectures for intra-coded frames and inter-coded frames. We assume here that the reader is familiar with the basics of the H.264 coder. Further details on the H.264 standard can be found in [4], [12].

A. ρ-Domain Rate-Distortion Models

Different models in the literature suggest different relations for rate vs. quantization step-size. In [8], [13], the ρ-domain source model is suggested, where ρ is the fraction of zero coefficients among the quantized transformed coefficients in a frame. The model assumes that there is a strong linear relation between ρ and the actual frame's bit rate: coarser quantization step-sizes generate more zero coefficients (and hence increase ρ) while decreasing the rate (where the rate here refers to the bits spent on coding the transform coefficients). Therefore, the suggested $rate - \rho$ relation is [8], [13],

$$R(\rho) = \theta \cdot (1 - \rho) \tag{1}$$

where R is the rate and θ is a parameter determining the slope. According to this equation, for $\rho = 1$ all the quantized coefficients are zeroed and thus the coding rate should approach zero. It is also argued in [8], [13] that the rate-ρ model is more robust than a rate- quantization-step model: the observed rate-ρ curves for both I and P frames share a very similar pattern, whereas the rate- quantization step-size curves change between different frame types.

The distortion too is more conveniently described in the ρ-domain than in the quantization step-size domain as it is defined within a finite range, $0 \leq \rho \leq 1$, and follow a more robust and regular behavior. In [3], an exponential-linear model for the MSE distortion in the ρ-domain was suggested as

$$D(\rho) = \sigma^2 \cdot e^{-\alpha \cdot (1-\rho)} \tag{2}$$

where σ^2 is the variance of the transformed coefficients and $\alpha > 0$ is a model parameter. Again, as $\rho \rightarrow 1$ and all the quantized coefficients are zeroed, the distortion approaches the σ^2 bound.

These models were derived for describing the rate and the distortion at the frame level, and were found quite accurate in [8], [3], [13], when tested for standards such as MPEG-2 and H.263, and were also used in [14], [15] for H.264. However, we found that for H.264 requantization at the macroblock level, these models are not good descriptors of the empirical data. Therefore, in subsection IV-B, we suggest different ρ-domain models, specifically adapted for H.264 requantization.

B. Architectures for Transrating of Coded Video

In this subsection we outline four transrating architectures that provide different compromises between quality and computational complexity. The spatial prediction introduced in H.264 intra frames requires distinguishing between the transrating approaches for intra-coded frames and inter-coded frames, as explained in the sequel.

A naive and straightforward transrating architecture is re-encoding [16], [17], where a decoder and encoder are cascaded. The input bit stream is fully decoded to obtain the reconstructed sequence and then re-encoded at the target output bit rate using new coding decisions. This architecture has the highest computational complexity among transrating architectures, as it makes new coding decisions, which also involve performing motion estimation (ME).

The architecture with the lowest computational complexity for requantization is the open-loop transrater [18], [16], [19], [17]. The residual's transform coefficients are dequantized and then requantized at a coarser step-size to meet the target bit rate. Following this scheme, expensive operations such as motion estimation (ME) and transforms are avoided and there is no need for a frame-store. However, open loop transraters are subject to a drift error that degrades the video's quality [19], [17]. The drift error is caused when the decoder and the encoder are not using the same reference signal for prediction.

In between these two extremes, there are architectures that reduce the computational complexity as compared to re-encoding, without introducing a drift error. In the *full decoder - guided encoder* (FD-GE) architecture [16], [17], [20], the input bit stream is fully decoded and then encoded by reusing the input coding decisions (e.g., motion vectors and intra prediction modes) to reduce the encoder's complexity. This transcoder does not suffer from drift error as the decoder-loop and the encoder-loop are independent and the residual is recomputed at the encoder.

The spatial prediction in intra frames use previously decoded neighbor pixels in the same frame to predict the current block pixels. Therefore, any mismatch between the transcoder and the encoder/decoder introduces a drift error that propagates throughout the frame [21]. Since some of the operations are not linear (due to rounding and clipping), this drift cannot be fully compensated. Therefore, to avoid the drift error, intra frames should be fully decoded into images in the pixel domain and then encoded [21], using the FD-GE architecture. The *guided encoding* allows either to reuse the input intra prediction modes or to selectively modify them, as will be discussed in subsection III-B. The selection of the requantization step-size for intra frames is discussed in subsection III-A.

A simplified FD-GE architecture for the case in which input coding decisions are reused, is the *partial decoder - partial encoder* (PD-PE) architecture [22], [19], [16], [17], [20]. The partial decoding reconstructs just the residual signal in the pixel domain, rather than reconstructing the fully decoded picture. It performs a closed-loop correction to compensate for the drift error, by applying the motion compensation (MC) once (in the joint transrater loop) instead of twice (during both decoding and encoding).

For inter-coded frames, it is customary to assume that the motion compensation is linear and that rounding and clipping operations can be neglected. Since the MC prediction is temporal, the drift error for inter-coded blocks using the PD-PE architecture is very small and it takes a number of frames before the accumulated error is noticeable. Therefore, we use the PD-PE architecture for transrating inter-coded frames.

H.264 defines an in-loop deblocking filter, which may be applied on the fully decoded pictures in the pixel domain. We assume that the filter is disabled, so the pictures need not be fully decoded and the PD-PE architecture can be applied [21]. Still, in section V we discuss the case of an input sequence for which the deblocking filter was enabled, proposing a modification that allows using our algorithm for such an input as well.

Intra-coded blocks inside inter frames are transrated using the PD-PE architecture too (with the appropriate changes, e.g., the MC block is replaced by the spatial predictor, etc.) though this is not the recommended architecture for them. Therefore, transrating inter frames with many intra-coded blocks using PD-PE architecture do cause some drift, but these cases are rather infrequent. The rate control algorithm handles these blocks as if they were inter-coded blocks. A block diagram of the proposed transrating system is depicted in Fig. 1.

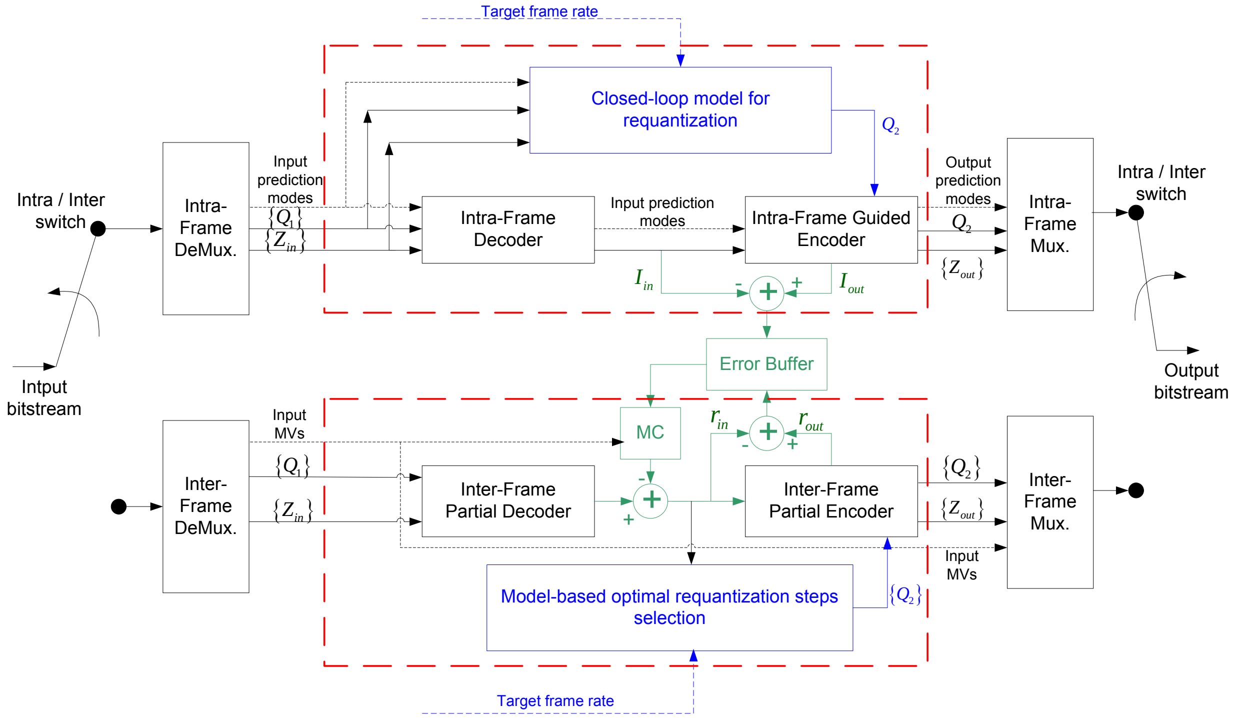

Fig. 1. Block diagram of the proposed transrating system. For each frame, the input bitstream is first parsed to read the input quantized coefficients indices, $\{Z_{in}\}$, the input quantization steps, $\{Q_1\}$, and the input prediction modes / motion vectors (MVs). Intra-coded frames are transrated using a FD-GE architecture (top block enclosed in a red dashed line). The guided encoder outputs are the output quantized coefficients indices, $\{Z_{out}\}$, the requantization step, Q_2, and the output intra prediction modes, all of which are entropy encoded and written in the output bitstream. The requantization step Q_2 is found using the closed-loop model for requantization, denoted in blue. The transrating error is saved in the error buffer (denoted in green), as part of a closed-loop correction scheme. Inter-coded frames are transrated using a PD-PE architecture (bottom block enclosed in a red dashed line). The partial decoder reconstructs the residual in the pixel domain, and then performs a closed-loop compensation, to account for the transrating errors introduced in the previous frames (denoted in green). The corrected residual, r_{in}, is fed into the model-based optimal requantization steps selection algorithm (denoted in blue), to find the optimal requantization steps, $\{Q_2\}$. The corrected residual, r_{in} is subtracted from the transrated residual r_{out} to form the transrating error, saved in the error buffer.

The remainder of the paper is organized as follows. Section II describes the use of ρ-domain rate-distortion models for bit allocation among transrated video frames in a Group of Pictures (GOP). The algorithm for transrating of intra-coded frames is described in section III, where the main mean for bit rate reduction is model-based uniform requantization (in subsection III-A) and a secondary mean is modification of the prediction modes (in subsection III-B). The algorithm for transrating of inter-coded frames is presented in section IV, using model-based optimal nonuniform requantization. The optimization algorithm is described in subsection IV-A, and new macroblock-level models in subsection IV-B. Section V summarizes the main simulation results and section VI concludes the paper.

II. Model-based optimal GOP-Level Bit Allocation

To achieve the bit rate reduction, we apply rate control algorithms at two levels. The coarser level determines the bit allocation to frames in a GOP, and is discussed in this subsection. The finer level allocates the bits to each frame encoding units (e.g., macroblocks) to achieve the frame target rate, and will be discussed in subsections III-A and IV-A for intra and inter frames, respectively.

The encoded bitstream describes two types of data. The 'texture bits' describe coding the quantized residual transform coefficients, whereas the 'overhead bits' describe the coding modes, MB types, etc. When the input coding modes are reused, most of the overhead bit count remains. Therefore, we assume that the change in the overhead bits due to transrating is negligible. To reduce the bit rate at an average transrating factor $BRfactor$, one could reduce each frame's bit rate by the $BRfactor$ factor. But, in H.264 the overhead bits are not negligible and therefore such a simple frame-level bit allocation is not suitable as it may leave too few texture bits for coding the residual.

Thus, we would like to find the optimal texture-bits allocation to the frames of that GOP. That is, to minimize the overall GOP distortion subject to the average rate constraint. This optimization problem was solved in [3] analytically by using the ρ-domain rate-distortion models. The authors of [23], [2], [24] suggested equalizing the frames distortions since subjectively the overall sequence distortion is more tolerable when all frames suffer similar distortion. In, [2], [24] the texture bits were not optimally allocated. Rather, each frame's target distortion was set as the average distortion of the previously encoded frames, and then its target rate was extracted using the ρ-domain rate-distortion models. In [25], a new optimal bit allocation problem was analytically solved for each encoded frame. For each frame, the target bit rate was calculated such that all the remaining frames in the GOP would have an equal distortion subject to the rate constraint, using a modified distortion model in the ρ-domain. Assuming that a GOP delay is tolerable, we propose to analytically solve a single optimal bit allocation problem per GOP, prior to its transrating. We minimize and equalize the transrating distortion over all the frames of that GOP, and the optimization problem formulation becomes:

$$\begin{aligned}
&\min_{\{R_k\}} \sum_{k=1}^{N} D_k(\rho_k) \qquad (3)\\
&subject \quad to: \\
&\sum_{k=1}^{N} R_k(\rho_k) \leq R_{GOP,target} \\
&D_1(\rho_1) = D_2(\rho_2) = ... = D_N(\rho_N)
\end{aligned}$$

where N is the number of frames in the GOP, R_k and D_k are the rate and the distortion of frame #k where $1 \leq k \leq N$ and $R_{GOP,target}$ is the target rate for the N frames together. We use the ρ-domain models (1) and (2) to obtain an analytic solution (using Lagrangian parameters to convert the constrained problem into an unconstrained problem):

$$R_k = \xi_k \cdot [ln(\sigma_k^2) - \frac{\sum_{l=1}^{N} \xi_l \cdot ln(\sigma_l^2) - R_{GOP,target}}{\sum_{l=1}^{N} \xi_l}] \tag{4}$$

$$D_k = exp(\frac{\sum_{l=1}^{N} \xi_l \cdot ln(\sigma_l^2) - R_{GOP,target}}{\sum_{l=1}^{N} \xi_l}) \tag{5}$$

where the resulting D_k is a constant (independent of the frame number k) and $\xi_k = \frac{\theta_k}{\alpha_k}$. This solution allocates more texture bits for the intra-coded frame (as compared to the allocation that does not pose the equal distortion constraint) to keep an equal distortion over all the frames.

The model parameters are adaptively extracted from the coded input for each frame. At the end of each frame's encoding, the deficit or surplus is uniformly distributed among the remaining frames in the GOP.

III. Intra Frames Transrating

In subsection I-B, we concluded that the spatial prediction introduced in intra-coded frames require a full decoding and guided encoding architecture (FD-GE) in order to avoid a drift error. The main mean for bit rate reduction for intra frames is via transform coefficients requantization (discussed in subsection III-A). A secondary mean is via modification of the prediction modes, to increase the coding efficiency (discussed in subsection III-B).

A. Model-Based Uniform Requantization

For intra-coded frames, we propose using uniform requantization for two reasons. One is that the typical bit budget for intra frames is sufficiently high (as compared to inter frames) to allow a frame-level rate control. The other reason is that the spatial prediction introduces block dependencies that extremely increase the computational complexity and memory requirements of solving an optimal nonuniform requantization problem. Due to these dependencies, the residual coefficients to be requantized are not available when needed for the requantization step-size selection. The uniform requantization step-size is found using two ρ-domain models: the linear rate-ρ model and a new $\rho - Q_2$ model, where Q_2 is the requantization step-size. The evaluation of the linear rate-ρ model is fairly simple and is described in subsection III-A1. Most of the effort is aimed at estimating the $\rho - Q_2$ model. Subsection III-A1 reviews the open-loop approach for evaluating the $\rho - Q_2$ relation and explains its shortcomings. Subsection III-A2 proposes a closed-loop statistical estimator for the $\rho - Q_2$ relation. It overcomes the block dependency problem by modeling the correction signal of the requantizated residual.

1) Open-loop approach for requantization step-size selection: We use the linear rate-ρ model (1) to set a uniform requantization step-size for an I-frame. The model parameter θ is estimated using the input rate-ρ point, $(\rho_{in}, R_{in}^{texture})$ and an anchor point at $(1, 0)$, see Fig. 2(a). Given the target rate for that frame, $R_{target}^{texture}$, we extract the expected fraction of zeros by

$$\rho_{target} = 1 - R_{target}^{texture}/\theta \tag{6}$$

The next step is to estimate the relation between ρ and the requantization step-size Q_2 as a $\rho = f(Q_2)$ lookup table, to be discussed in section III-A2. Then, the target step is found by

$$Q_{2,target} = f^{-1}(\rho_{target}) \tag{7}$$

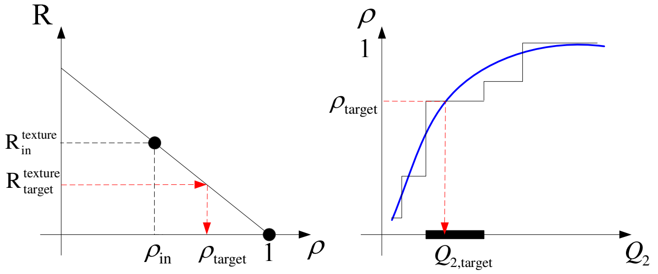

Fig. 2. Uniform requantization using a rate-ρ model. Left: rate-ρ relation, the dark circles are at $(\rho_{in}, R_{in}^{texture})$ and $(1, 0)$, from which θ is estimated. Right: $\rho(Q_2)$ relation, blue smooth curve: closed-loop estimator, black staircase curve: open-loop estimator. Given $R_{target}^{texture}$, we extract ρ_{target} and then find the corresponding $Q_{2,target}$ using the closed-loop $\rho(Q_2)$ estimator. Using the open-loop $\rho(Q_2)$ estimator, there is an uncertainty interval regarding $Q_{2,target}$ choice, as illustrated by the thick black line.

Due to spatial prediction, requantization of the prediction residual at one block changes the residual in neighboring casual blocks (where casual neighbors are the previous blocks processed according to a raster scan order). To avoid a drift error, intra frames are fully decoded into pictures in the pixel domain, and then encoded. But, estimating the $\rho(Q_2)$ relation this way requires multiple encoding of the picture at different Q_2 steps, which is not practical.

The simplest $\rho(Q_2)$ estimator is the *open-loop estimator*, evaluated from the output of the scheme depicted in Fig. 3. The input quantized indices, Z_{in}, are dequantized using the input quantization step-size, Q_1, to yield the residual transform coefficients Y. When Y is requantized, using a quantizer with step-size Q_2 and deadzone Δz, the output indices are derived by

$$Z_{out} = sign(Y) \cdot \lfloor \frac{|Y|}{Q_2} + \Delta z \rfloor \tag{8}$$

Therefore, all transform coefficients that fall in the interval $[-t(Q_2), t(Q_2)]$ are requantized to zero, where $t(Q_2) = (1 - \Delta z)Q_2$. For intra frame, $\Delta z = \frac{1}{3}$ and theredore $t(Q_2) = \frac{2}{3}Q_2$. This process is repeated for each Q_2 step-size, to derive the $\rho(Q_2)$ relation.

This open-loop $\rho(Q_2)$ estimator cannot track the changes in the residual and therefore it has two disadvantages: One is that it is not accurate enough at moderate to coarse requantization, where large changes in residual intensity cause a large drift error. The other is its staircase characteristic, see staircase curve in Fig. 2(b). Given a target ρ

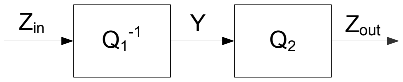

Fig. 3. Open-loop requantization scheme.

value, the estimator may encounter an uncertainty as to which requantization step-size to choose, which is illustrated by the thick black line in Fig. 2(b), denoting the uncertainty interval.

2) Closed-loop estimation of $\rho(Q_2)$*:* As noted earlier, since the residual coefficients to be requantized are not available in advance of setting Q_2, the estimation of $\rho(Q_2)$ is not trivial. To estimate $\rho(Q_2)$ more accurately than the open-loop estimator, we propose [11] to model the process that the input coefficients Y undergo to become the residual coefficients to be requantized. To this end, we need not estimate the value of every single coefficient, but rather the statistical distribution of the coefficients. We start by describing the model's scheme and continue by providing a statistical description of the residual coefficients to be requantized.

Closed-loop residual modeling architecture

We propose to estimate $\rho(Q_2)$ using a model that is based on a closed-loop residual architecture in the transform domain, as depicted in Fig. 4. The closed-loop estimator statistically models the required correction of the requantized residual coefficients, thereby overcoming the dependency problem. The scheme in Fig. 4 is merely used in order to model the distribution of residual coefficients to be requantized, from which ρ is estimated. During actual transrating, we fully decode the picture, estimate the $\rho(Q_2)$ relation using this model, estimate the linear rate$-\rho$ model (as described in subsection III-A1), choose Q_2 that meets the target rate (as illustrated in Fig. 2) and then encode the picture once (by performing spatial prediction, transforming the obtained residual and requantizing) using the chosen Q_2.

Instead of evaluating $\rho(Q_2)$ based on Y, the closed loop $\rho(Q_2)$ estimator evaluates how many of the corrected transform coefficients W (see Fig. 4) fall in the deadzone interval. The corrected residual is defined as $W \triangleq Y - C$, where C is the correction signal in the transform-domain. This signal is formed by feeding the transform-domain transrating error ε, into the transform-domain spatial-predictor (performs the equivalent operation to spatial prediction in the transform-domain [26]). Due to some nonlinearities (rounding and clipping operations), the transrating error ε cannot be defined simply as the requantization error. Rather, it is defined as the transform of the difference between the decoded output and input images, where the output image is decoded using the requantized indices $Z_{out} = Q_2(W)$.

In order to evaluate $\rho(Q_2)$ from W, we first characterize the statistical distributions of Y and C, and then find how W is distributed. Since the input transform coefficients Y have values that are multiples of the input

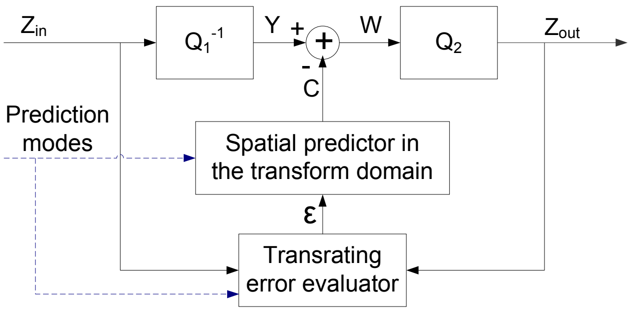
Z_{in}
Q1-1
Y
+
W
Q2
Zout
-
C
Prediction modes
Spatial predictor in the transform domain
ε
Transrating error evaluator

Fig. 4. A closed-loop modeling scheme for estimating $\rho(Q_2)$. The transrating error ε is fed into the predictor to yield the correction signal C. Then, $\rho(Q_2)$ is estimated based on $W \triangleq Y - C$.

quantization step-size Q_1, their distribution is discrete, and given as:

$$p_Y(y) = \sum_{l=-L}^{L} p_l \cdot \delta(y - lQ_1) \tag{9}$$

where $\delta(y)$ is the unit impulse function, L is the smallest integer such that $|Y| \leq LQ_1$, and $\{p_l\}_{l=-L}^{L}$ are extracted from the input coefficients.

The correction signal C is modeled as a continuous distribution. Since this signal can not be explicitly extracted from the input stream, most of the effort is aimed at its characterization and its statistical modeling. Once the distribution of C is obtained, the next step is to find the distribution of $W = Y - C = Y + (-C)$. A schematic illustration of the distribution of W is depicted in Fig. 5. Since we cannot assume that C is independent of Y, we use the joint probability of $(Y, -C)$:

$$p_{Y,-C}(y, c) = p_{-C|Y}(c|y) \cdot p_Y(y) \tag{10}$$

to calculate the cumulative distribution of W:

$$\begin{aligned} Pr.(W \leq w_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{w_0 - y} p_{Y,-C}(y, c) dc dy = \\ = \sum_{l=-L}^{L} p_l \cdot \int_{-\infty}^{w_0 - lQ_1} p_{-C|Y}(c|Y = lQ_1) dc \end{aligned} \tag{11}$$

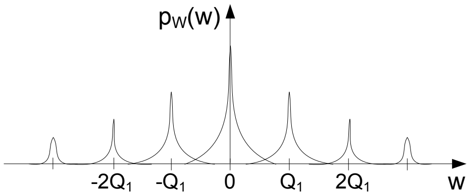

Fig. 5. Schematic illustration of the probability distribution of W.

Therefore, the closed-loop $\rho(Q_2)$ evaluation is given by:

$$\rho(Q_2) = Pr.(|W| \leq t(Q_2)) = \sum_{l=-L}^{L} p_l \cdot \phi(l|Y) \tag{12}$$

where

$$\phi(l|Y) = \int_{-t(Q_2)-lQ_1}^{t(Q_2)-lQ_1} p_{-C|Y}(c|Y = lQ_1)dc \tag{13}$$

Lacking a known model for the correlation between Y and C, we are left with the unfeasible task of modeling $\phi(l|Y)$, for every possible value of Y (corresponding to $|l| \leq L$). From observations, we found that a reasonable approximation can be obtained by distinguishing between zero and non-zero inputs. That is, to model $\phi(0|Y = 0)$ and $\phi(l|Y \neq 0)$ separately. In that case, the model in (14) for $\rho(Q_2)$ is simpler than substituting (13) into (12), as there are two possible input dependencies instead of $2L + 1$. To complete the evaluation of $\rho(Q_2)$, we now address the evaluation of $\phi(0|Y = 0)$ and $\phi(l|Y \neq 0)$, by characterizing the correction signal C and modeling its distribution.

$$\rho(Q_2) = p_0 \cdot \phi(0|Y = 0) + \sum_{l=-L, l\neq 0}^{L} p_l \cdot \phi(l|Y \neq 0) \tag{14}$$

Correction signal characterization

To ease its statistical modeling, the correction signal C is partitioned into homogenous *data groups* that share the same characteristics, according to three partitioning criterions.

The first partition of the data is according to its spatial prediction modes that spectrally shape the white error ε.

The second partition distinguishes the affected coefficients from the unaffected coefficients. Affected coefficients are those coefficients that are changed as a result of spatial prediction; whereas unaffected coefficients have a zero correction signal. For example, DC prediction affects just one transform coefficient out of a 4x4 ICT block. This classification is predefined for each prediction mode by an "affected coefficients mask" whose shape is characterized by the prediction mode type, see Fig. 6. The advantage of the affected/unaffected coefficients classification is that the $\rho(Q_2)$ relation for the unaffected coefficients can be evaluated as in the simple case of an open-loop estimator, thereby reducing the complexity of evaluating the $\rho(Q_2)$ relation.

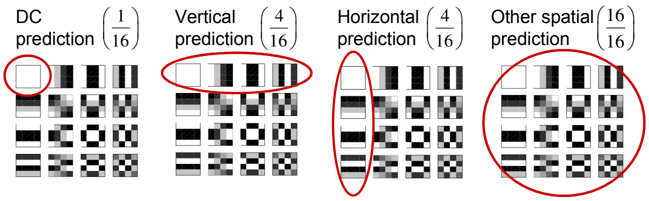
DC prediction $\left(\frac{1}{16}\right)$
Vertical prediction $\left(\frac{4}{16}\right)$
Horizontal prediction $\left(\frac{4}{16}\right)$
Other spatial prediction $\left(\frac{16}{16}\right)$

Fig. 6. Illustration of the location of the affected/unaffected transform coefficients using their ICT basis images. The classification is done according to the prediction modes. The affected coefficients basis images are encircled in red, and their fraction is denoted in parenthesis.

The third partition distinguishes between the corrections applied to zero/non-zero input coefficients. Next, a probability distribution is fitted to each data group allowing evaluation of its $\rho(Q_2)$ relation according to (14).

Correction signal modeling using a Γ distribution

To evaluate (14) for each data group, a statistical description of $\phi(0|Y = 0)$ and $\phi(l|Y \neq 0)$ is required. To study this issue, we evaluated the correction signal C offline, according to the scheme of Fig. 4, and performed the partitioning into data groups. We then found that the Γ distribution is a good descriptor of each of the correction signal partitions. The probability density function for the two-sided Γ distribution is defined as [27]:

$$p_X(x;\beta) = \frac{1}{2\sqrt{\pi}}\sqrt{\frac{\beta}{|x|}} \cdot exp\{-\beta|x|\} \tag{15}$$

where $\beta > 0$ is a scale parameter, whose decrease results in a wider distribution. The Γ cumulative distribution function is defined by (16), where $\Gamma(a, 0.5) \triangleq \int_0^a t^{-0.5}exp(-t)dt$.

$$Pr.(X \leq x;\beta) = \frac{1}{2} + sgn(x)\frac{1}{2\sqrt{\pi}}\Gamma(\beta|x|, 0.5) \tag{16}$$

For each prediction mode, a ML estimator was applied to find the scale parameter β for the affected correction coefficients, while distinguishing $\beta^{C|Y=0}$ from $\beta^{C|Y\neq 0}$ for the zero/non-zero input coefficients, respectively. Using (16) and these estimated parameters, the functions $\phi(0|Y = 0)$ and $\phi(l|Y \neq 0)$ take the form of (17), and $\rho(Q_2)$ can be evaluated for each data-group by substituting (17) into (14). Then, all data-groups $\rho(Q_2)$ relations are linearly weighted (according to their size) to obtain the frame level relation.

$$\begin{aligned} \phi(0|Y = 0) &= Pr.(|C| \leq t(Q_2); \beta^{C|Y=0}) \\ \phi(l|Y \neq 0) &= Pr.(|C + lQ_1| \leq t(Q_2); \beta^{C|Y\neq 0}) \end{aligned} \tag{17}$$

As stated earlier, in a real-time scenario, the scheme of Fig. 4 is not implemented. Therefore, the correction signal C is not available and the ML estimator for β cannot be used. Observations show that the value of β monotonically decreases with Q_2, as coarser requantization generates a transrating error ε with a wider dynamic range (here, measured by $||\varepsilon||_1$), which in turn generates a correction signal with a wider dynamic range when fed back to the predictor. However, the great variability in the $\beta - Q_2$ relation over different data-groups complicates its modeling. Therefore, we suggest to decompose this relation into two separate models: β vs. $||\varepsilon||_1$ and $||\varepsilon||_1$ vs. Q_2, as illustrated in Fig. 7. The β vs. $||\varepsilon||_1$ relation is modeled by $\beta = \beta_0/||\varepsilon||_1$. When the transrating error is zero, a correction signal is not generated, hence $\beta \to \infty$. The $||\varepsilon||_1$ vs. Q_2 relation was empirically fitted using the monotonically increasing function $||\varepsilon||_1 = a_1 \cdot (ln(Q_2))^2 + a_2$, whose parameters a_1, a_2 are functions of the input "initial conditions", Q_1 and $||Y||_2$.

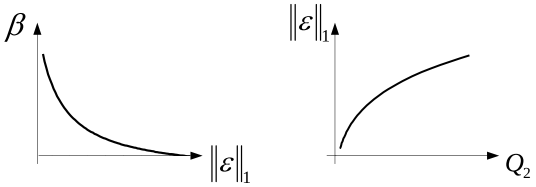

Fig. 7. Decomposition of the β vs. Q_2 relation, using $||\varepsilon||_1$.

To summarize, the modeling steps are as follows:

1) Segment the transform coefficients into data groups (according to the prediction modes, affected/unaffected coefficients, and zero/non-zero input coefficients).

2) For each data group, evaluate the β distribution parameter from the input data in two stages:
   a) Model the $||\varepsilon||_1$ vs. Q_2 relation (fit the parameters a_1, a_2).
   b) Model the β vs. $||\varepsilon||_1$ relation (fit the parameter β_0).

   Substitute (17) into (14) to evaluate the $\rho(Q_2)$ relation for that data group.
3) Linearly weight the obtained $\rho(Q_2)$ relations for the different data parts according to their relative size to get the frame level $\rho(Q_2)$ relation.

If the input frame is not uniformly quantized during the first encoding, an additional data partition according to the initial quantization step is added to the data groups segmentation. Subsection V-B1 compares the $\rho(Q_2)$ evaluation using the proposed model to the true data and the open-loop estimator.

B. Modification of Prediction Modes

The proposed architecture used for transrating intra-coded frames (see subsection I-B) requires full decoding and encoding in order to avoid a drift error. Although we have to fully decode the frame, we need *not* fully encode it by means of a computationally expensive full prediction modes search. Rather, we perform a *guided encoding*, which uses already encoded information from the input bitstream. One option is to reuse the input prediction modes. The other option is to selectively modify the input prediction modes where the coding efficiency is expected to improve.

Spatial prediction in intra-coded frames significantly increases the coding efficiency when the coding modes are appropriately selected. As the bit rate is reduced, the quality is degraded and fine details are less likely to be preserved. The observed trend regarding the encoder's intra coding decisions shows that as the bit rate is reduced, larger prediction blocks are chosen (more 16x16 partitions) and the frequency of ”simple” modes (horizontal, vertical and DC prediction) increases at the expense of the more complex ”diagonal” modes for the remaining 4x4 partitions. However, for some blocks, ”complex” modes usage significantly improves the coding efficiency, so these modes cannot be completely discarded from the search.

A previous work [28], considered the modification of prediction modes originally coded as 4x4, as most of the coding gain is expected due to these modes modification. That work used the number of bits spent on coding the original MB as a prior to discern the smooth from the highly detailed MBs. Based on that classification, smooth MBs were examined for 16x16 prediction whereas highly detailed MBs were examined for 4x4 predictions. The decision whether or not to change the mode, in that work, was based solely on the distortion. Such an approach may yield large rate deviations, as the best mode selection is correlated with its rate-distortion cost at the current bit rate working point.

We suggest choosing the best new modes, while considering both the input prior and the Human Visual System (HVS) characteristics. The input bit consumption is used as the input prior and the distortion is weighted according to the HVS characteristics, both explained in the sequel.

To better understand our mode decision process, we first outline how the mode is chosen in the H.264 encoder. Let us denote by d_i and r_i the transrating distortion and the number of bits spent for block i. Using the Lagrangian parameter λ as defined by the H.264 rate-distortion function [9]: $\lambda(QP) = 0.85 \cdot 2^{\frac{QP-12}{3}}$, where QP is the quantization parameter, the best mode m_i^* is chosen by:

$$m_i^* = \underset{m}{\operatorname{argmin}}\{d_i(m, QP) + \lambda(QP) \cdot r_i(m, QP)\} \tag{18}$$

Our best mode choice is given by:

$$m_i^* = \arg\min_{m \in M}\{d_i(m, QP) + \lambda(QP) \cdot f_{HVS}(b_i) \cdot r_i(m, QP)\} \tag{19}$$

where M is the subset of modes found using the input prior and $f_{HVS}(b_i)$ is the perceptual weight given to block b_i, as we explain next.

1) Input prior: We suggest to use the input prediction mode to narrow down the number of searched modes. For MBs initially encoded at a 16x16 prediction and for the chrominance components, the input mode is reused so no new modes are searched for. For MBs initially encoded as 4x4, we determine the subset M of modes that are searched for, by classifying the picture macroblocks into three groups. The classification is done according to their input bit consumption, as depicted in Fig. 8, where N_B is the number of macroblocks in the frame.

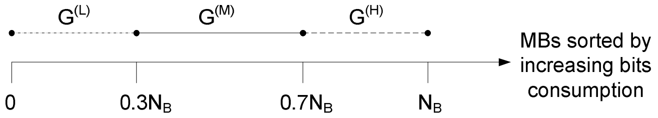

Fig. 8. Macroblocks classification to G^L, G^M, G^H groups according to the input bits consumption.

The searched modes groups are defined as follows:

- G^L group (the lowest 30% input bits consumption) - blocks are assumed to be relatively smooth and are therefore candidates for a 16x16 prediction. $M = \{\text{input mode, all 16x16 modes}\}$
- G^H group (the highest 30% input bits consumption) - blocks are assumed to be highly detailed. Since these constitute only 30% out the macroblocks, but expected to increase the coding efficiency if the best matched modes are chosen, we examine all 4x4 modes for this group. $M = \{\text{all 4x4 modes}\}$
- G^M group. $M = \{\text{input mode, 4x4 DC mode}\}$

2) HVS characteristics considerations: Psychovisual studies have led to the concept of a perceptual three component image model [29]: texture regions, smooth regions, and edges. In [30], the authors suggest to modify the block's distortion value according to its perceptual importance, using 6 different perceptual groups, where each has a different f factor. The distortion is weighted by the $1/f$ factors and is plugged into the rate distortion cost function. We follow this idea but segment the image into the three perceptual groups of {texture regions, smooth regions, and edges}. First, we calculate the variance of the block coefficients, where the DC term and the first two AC coefficients are not taken into account to avoid slow intensity changes detection. The variances map is translated into low and high activity blocks using an adaptive threshold. Morphological operations are then used to detect the edges and smooth regions and form the segmented picture. Since artifacts are most apparent at smooth regions and less noticeable at textured regions, we set $f_{texture} > 1$, $f_{smooth} < 1$, and $f_{edge} = 1$. The specific parameter values are given in subsection V-B2.

IV. Inter Frames Transrating

In subsection I-B, we defined the closed-loop residual correction architecture for inter frames, which also reuses the input motion decisions. Since the typical bit budget for inter frames is low (as compared to intra frames), the rate control should be accurate in order to meet the target bit rate. Therefore, we propose an optimal non-uniform requantization (subsection IV-A). To reduce the computational load, we suggest using new macroblock level models, adapted to H.264 requantization (subsection IV-B).

A. Optimal Requantization

1) Introduction: In previous standards, like MPEG-2, the optimal requantization problem is defined as finding a set of optimal new step-sizes, where optimality is in the sense of minimizing the total distortion, subject to a given bit-rate constraint:

$$\min_{\{QP_i\}} \quad D, \quad subject \quad to \quad R \leq R_{target} \tag{20}$$

where,

$$D = \sum_{i=1}^{N_B} d_i(QP_i), R = \sum_{i=1}^{N_B} r_i(QP_i) \tag{21}$$

with, N_B - number of macroblocks in the frame, QP_i - quantization parameter for the i-th macroblock, d_i - distortion caused to the i-th macroblock, r_i - number of bits produced by the i-th requantized macroblock.

A common approach [1] is to convert the constrained optimization problem to an unconstrained one:

$$\min_{\{QP_i\}} \quad J, \quad J = D + \lambda(R - R_{target}) \tag{22}$$

where λ is the Lagrangian parameter. The main advantage of solving the unconstrained problem is that the cost J can be broken into a sum of independent costs for each macroblock. Given a λ value, the set of quantization steps $\{QP_i^*\}_{i=1}^{N_B}$ that minimizes the set of independent costs is found and the corresponding average rate is calculated by $\sum_{i=1}^{N_B} r_i(QP_i^*)$. Then, the λ parameter is altered, using for instance, bisection iterations, until an average rate that is close enough to the target is obtained.

In [31], [30], [24], it is argued that avoiding large fluctuations in the quantization step-size throughout the frame results in better subjective quality, as the overall perceived frame's quality appears constant and blocking artifacts are reduced. In addition, the H.264 standard encodes the quantization parameter differentially, that is, it encodes $\Delta QP = QP - QP_{Prev}$, where QP, QP_{Prev} are the quantization parameters of the current and the previous encoded macroblock according to a raster scan order. Moreover, the cost in bits of the ΔQP transition increases with its absolute value. As a result, many rate control algorithms for H.264 limit $|\Delta QP|$ to take small values (typically, up to 2).

2) Optimization: Following the assumption that the change in the overhead bits due to transrating is negligible (see section II), we define the optimization problem in terms of the texture bits:

$$\min_{\{QP_i\}} \quad J, \tag{23}$$
$$J = D + \lambda(R^{texture} - R^{texture}_{target})$$

In addition, we propose to regulate the changes in QP to achieve better subjective quality by adding a regularization term $\mu \sum_{i=2}^{N_B} cost(\Delta QP_i)$, that accounts for the cost in bits of coding ΔQP (as defined in the standard [9]). As the weight parameter μ translates the regularization term measured in bits to distortion units and we do not try to achieve an exact bit target for coding ΔQP, we choose to set $\mu = \lambda$, so that it has the same units, simplifing the solution:

$$\min_{\{QP_i\}} \quad J, \tag{24}$$
$$J = D + \lambda(R^{texture} - R^{texture}_{target}) + \lambda \sum_{i=2}^{N_B} cost(\Delta QP_i)$$

Since the choices of quantization step-sizes for different macroblocks are no longer independent, the whole set of quantization step-sizes $\{QP_i^*\}$ should be found at once. Therefore, we propose to extend each Lagrangian iteration with a dynamic programming stage. The external Lagrangian iterations change the Lagrangian parameter λ to improve the rate guess. At each examined value of λ, the dynamic programming algorithm finds an optimal QP path by solving (24), as will be explained next. The results showed that the above algorithm rarely chooses $|\Delta QP|$ values bigger than 3. As there is no practical need for larger $|\Delta QP|$, we limit the allowed transition to $|\Delta QP| \leq 3$.

The optimization problem is then defined by:

$$\min_{\{QP_i\}} D, \tag{25}$$
$$subject \quad to$$
$$R^{texture} \leq R^{texture}_{target} \quad and \quad |\Delta QP| \leq 3$$

At each examined value of λ, the constrained dynamic programming algorithm finds an optimal QP path by solving:

$$\min_{\{QP_i\}} J \quad subject \quad to \quad |\Delta QP| \leq 3 \tag{26}$$

where $J = D + \lambda(R^{texture} - R^{texture}_{target}) + \lambda \sum_{i=2}^{N_B} cost(\Delta QP_i)$.

The dynamic programming algorithm is defined over the set of states $\{(QP, i)\}$, where i is the macroblock index and QP is the quantization index, see Fig.9. Each state (QP, i) has its cost-value $j_i(QP) = d_i(QP) + \lambda r_i(QP)$ and the total frame's cost along a path is $J = \sum_{i=1}^{N_B} j_i(QP) + \lambda \sum_{i=2}^{N_B} cost(\Delta QP_i)$.

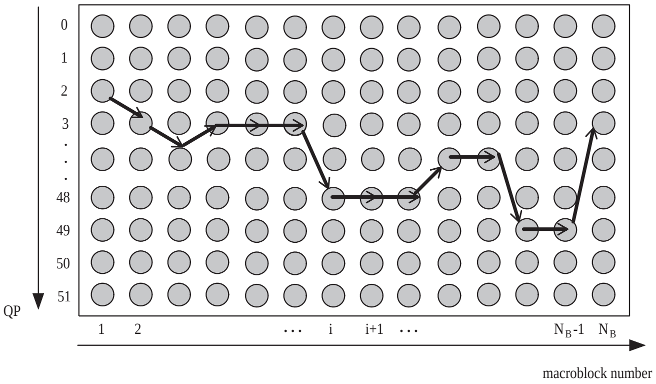

Fig. 9. Dynamic programming path illustration. Horizontal axis: macroblock number, vertical axis: the quantization parameter QP. Each circle denotes a state, and each column corresponds to a macroblock stage. The arrows show a path example, where the change in QP from one macroblock to the next is within ± 3 units.

The optimal path up to state (QP, i) is the path that has the minimal accumulated cost, $V_i(QP^*)$, over all possible paths that end at that state. Because $|\Delta QP| \leq 3$, there are at most 7 possible paths that end at the previous macroblock (#i-1) and that can be continued to the current state (QP, i). We choose among these by minimizing the value function of the current state:

$$V_i(QP) = V_{i-1}(QP_{Prev}) + j_i(QP) + \lambda cost(QP_{Prev}, QP) \tag{27}$$

where $QP_{Prev} - QP \in \{-3, -2, -1, 0, 1, 2, 3\}$. It is the sum of the cost of the path until the previous macroblock $V_{i-1}(QP_{Prev})$, plus the cost of the current state $j_i(QP)$, plus the cost of moving from state $(QP_{Prev}, i-1)$ to (QP, i). Or, in other words, the best path up to state (QP, i) is continued from state $(QP^*_{Prev}, i-1)$, where

$$QP^*_{Prev} = \arg\min_{QP_{Prev}} \{V_{i-1}(QP_{Prev}) + \lambda cost(QP_{Prev}, QP)\} \tag{28}$$

The corresponding value function update is then:

$$V_i(QP) = V_{i-1}(QP^*_{Prev}) + j_i(QP) + \lambda cost(QP^*_{Prev}, QP) \tag{29}$$

At each stage i of the dynamic programming algorithm (from the first to the last macroblock), the best paths for all (QP, i) states are found and kept as lists of pointers, along with their values. When the algorithm reaches the last stage ($i = N_B$), the best path found is the optimal path over the entire frame:

$$BestPathEnd = \operatorname*{argmin}_{QP} V_{N_B}(QP) \tag{30}$$

The algorithm then traces back the best frame-path using the chosen list of pointers, to obtain the optimal path: $\{QP^*_i\}_{i=1}^{N_B}$. Since we would like to reduce the bit rate, we constrain the requantized step-sizes not be finer than the original step-sizes. Thus, states that correspond to QP smaller than the original are assigned an infinite cost and discarded from the search procedure. The dynamic programming algorithm is performed at each Lagrangian iteration. The Lagrangian iterations convergence criterion is that the resulting rate deviates from the target rate by no more than 4%. In addition, in case the bisection algorithm is stuck, there is also a tolerance of 0.1% on the minimal amount of change in λ between consecutive Lagrangian iterations. The number of Lagrangian iterations required until convergence is 6 to 8, on average.

3) Coefficient elimination: After applying the transform and quantization, the quantized indices blocks are typically sparse. At the encoder, or the transcoder for that matter, it is possible to modify the obtained indices levels to achieve a lower cost, in terms of rate-distortion. In [32], [33], [34], indices modification was examined by evaluating the modified costs exhaustively, that is, evaluate a few optional rates directly from the entropy coding tables *without using models*. A simpler case of indices modification is coefficient elimination, or thresholding [35], [36], [37]. Specifically, [37] considers the coefficient elimination rule used in the H.264 recommended encoder. It zeroes sparse blocks that are almost zeroed except for a few high-frequency trailing-ones (± 1 at the end of the block) corresponding to transform coefficients at high frequencies.

We examined incorporating selective coefficient elimination into the proposed rate-distortion optimization algorithm. To reduce the computational load regarding which coefficient to eliminate, we follow the simple elimination rule used in the recommended encoder software.

We optimally decide for each quantized MB whether to encode it as is or to perform coefficient elimination first, as follows: Two rate-distortion pairs are evaluated for each combination of quantization parameter QP and macroblock index i: $\{d_i^0(QP), r_i^0(QP)\}$ and $\{d_i^1(QP), r_i^1(QP)\}$, for the case of no elimination and the case of elimination according to the reference software rule, respectively. As a result, a two layer array for the rate and the distortion is generated over the set of states $\{(QP, i, elim)\}$, where $elim \in \{0, 1\}$ is a binary flag that denotes whether or not elimination is performed, see Fig. 10. The optimization problem is then defined by:

$$\begin{aligned} &\min_{\{QP_i, elim_i\}} D \\ &subject \quad to: \\ &R \leq R_{target} \\ &|\Delta QP| \leq 3 \end{aligned} \tag{31}$$

where $D = \sum_{i=1}^{N_B} d_i^{elim_i}(QP_i)$ and $R = \sum_{i=1}^{N_B} r_i^{elim_i}(QP_i)$.

To solve the optimization problem of (31), we first follow the Lagrangian iterations extended by a dynamic programming algorithm, as explained earlier. The dynamic programming algorithm is then extended from a single 2D layer to two layers. When the algorithm reaches the last stage ($i = N_B$), the best path $\{QP_i^*, elim_i^*\}_{i=1}^{N_B}$ is the optimal path over the entire frame:

$$\begin{gathered} (BestPathEnd, BestElimEnd) = \\ \operatorname*{argmin}_{QP} \operatorname*{argmin}_{elim} V_{N_B}^{elim_{N_B}}(QP) \end{gathered} \tag{32}$$

We compared the performance of the selective coefficient elimination with that of no elimination, where in both cases the requantization step-sizes were optimally selected. The current implementation of the selective elimination algorithm shows a small gain in terms of PSNR vs. bit rate (about 0.07 [dB]). This gain is small as only a small part of the frame blocks are selected for elimination. Full elimination (without selection) is not recommended and the PSNR loss at high bit rates can get to 0.4 [dB]. Even though, we believe that this algorithm can potentially achieve a higher gain, by using more sophisticated elimination rules.

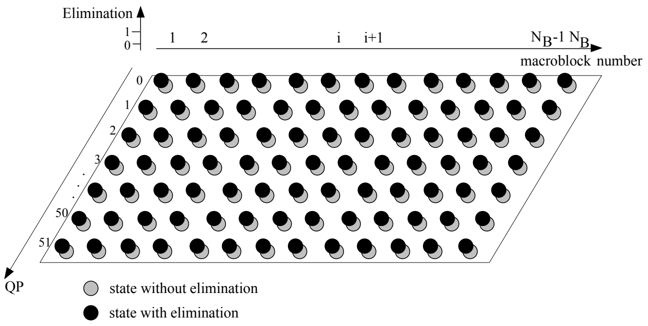

Fig. 10. Two layer array illustration. Horizontal axis: macroblock number, vertical axis: the quantization parameter QP. Each disc denotes a state, where the black and gray colors correspond to states with and without elimination, respectively.

B. Rate-Distortion Modeling

The optimization algorithm described above requires the evaluation of the rate and distortion obtained by requantizing each macroblock at multiple step-sizes. If no prior knowledge is used, such rate assessment involves the simulation of the actual requantization followed by entropy coding. As this procedure must be repeated multiple times, the optimization becomes computationally expensive. The computational complexity can be greatly reduced by using an analytic model for the relation between rate and quantization step-size, *for each macroblock*. In order to incorporate the ρ-domain models into the optimization, we suggest [10] modified models for H.264 at the macroblock level. The proposed rate-ρ model is especially adapted for requantization in the H.264 standard. Therefore, we briefly outline the H.264 entropy coding first and then describe the proposed model.

1) H.264 Context Adaptive Entropy Coding: The H.264 context adaptive entropy coding with VLC tables (CAVLC), is designed to take advantage of the sparse (compact energy) characteristics of the quantized transform coefficients [4]. To this end, it uses a set of syntax elements, that includes both the customary run-level representation and additional overhead counts that mainly describe the zero valued coefficients distribution. On top of that, it switches between several VLC tables for each syntax element, in a context adaptive manner.

Though the run and level are encoded separately, their encoding is efficient due to the context based VLC tables switching. The additional overhead counts consist of two symbols. One describes the combination of the number of non-zero coefficients and the high-frequency trailing-ones (± 1 at the end of the block). We shall refer to it as (TotalCoefficients, TrailingOnes). The other symbol, called TotalZeros, denotes the number of zeroed coefficients from the DC coefficient to the highest frequency non-zero coefficient. Both of which use multiple VLC tables. Fig.11 shows an example for a 4x4 zig-zag scanned block, with 6 non-zero coefficients, 2 trailing-ones, and 2 TotalZeros (that are marked in gray).

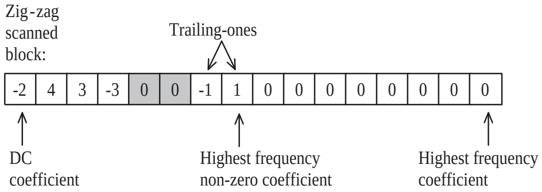

Fig. 11. An example of the additional overhead syntax elements in H.264.

2) Rate-ρ Model for H.264 Requantization: Examination of the rate-ρ relation at the macroblock level has shown that a linear relation is not a good descriptor of the empirical data. Therefore, and in light of H.264 new entropy coding features, we suggest a different rate-ρ model at the macroblock level. We decompose the rate into ”*data*” and ”*overhead*” components, where the ”*data*” stands for the bits spent on coding the run-level, and the ”*overhead*” designates the bits spent on coding the new syntax elements. For the model parameters estimation we use prior information, such as the original input quantized transform coefficients and their encoded rate.

”Data” Component

The ”data” texture bits component is composed of coding the run-level syntax elements that form the majority of the texture bits at moderate to high bitrates. This component rate-ρ relation is a monotonically decreasing convex function.

Therefore, for the ”data” component rate-ρ relation, we suggest the following closed-form model:

$$r^{data}(\rho) = \theta \cdot ln(1 + (1 - \rho)^{\eta}) \tag{33}$$

where $\theta \geq 0, \eta \geq 1$. The parameter θ controls the scale of the curve, whereas the parameter η changes its shape. Now, given this component's original input encoded rate of a macroblock, $r_{in}^{data}(\rho_{in})$, we can fit one of the parameters. Since this model requires fitting two parameters, we apply a two-dimensional search to fit its shape parameter η and an average scale parameter $\bar{\theta}$ using the input ensemble $\{r_{in,i}^{data}(\rho_{in,i})\}_{i=1}^{N_B}$ of all the frame macroblocks. The estimated shape parameter η is used for all the frame macroblocks. The scale parameter θ_i is then matched to each macroblock separately by:

$$\theta_i = \frac{r_{in,i}^{data}}{ln(1 + (1 - \rho_{in,i})^{\eta})} \tag{34}$$

The luminance and the chrominance components are modeled separately.

Since the frame macroblocks share the same parameter η, but each has a different parameter θ_i, we cannot depict their model-based fittings on a single graph. However, we can scale all macroblock level relations using the average frame level parameter $\bar{\theta}$, by drawing $r_i^{data}(\rho_i) \cdot \frac{\bar{\theta}}{\theta_i}$ and then draw their common fit $r^{data}(\rho_i) = \bar{\theta} \cdot ln(1 + (1 - \rho_i)^{\eta})$. Fig. 12 depicts for each macroblock its scaled rate-ρ relation by blue dots and the common fit by a red line.

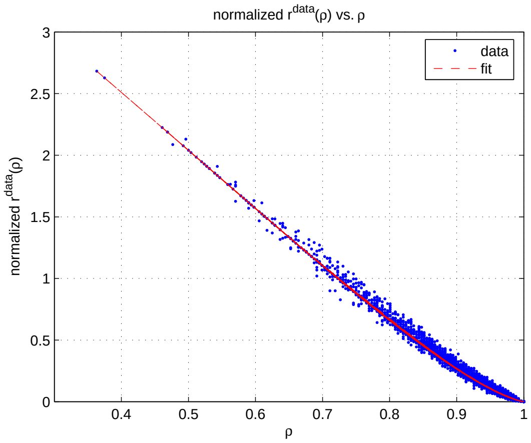

Fig. 12. Blue dots: normalized $r^{data}(\rho)$ relation of one frame's macroblocks; red solid line: the fit with the common shape parameter η. Here, $\eta = 1.36$ and $\bar{\theta} = 6.2$.

"Overhead" Component

The "overhead" component rate-ρ relation is very noisy due to two reasons. One is that the overhead syntax elements values (e.g. (TotalCoefficients, TrailingOnes)=(6,2) and TotalZeros=2 in the example of Fig. 11) are not uniquely defined by the local block's ρ. The other is the use of multiple VLC tables for each syntax element, which means that the number of bits spent on coding the same syntax element value changes with the context. As a result, fitting a closed-form model for it becomes practically impossible. However, due to the partial dependency on the local ρ, we chose to use a statistical model to characterize the average code length at the 4x4 block level, and then average over the 16 blocks in the macroblock.

Each 4x4 block has a local percentage of zeroed coefficients, ρ_b, which is related to the local total non-zero coefficients count TC_b, by $\rho_b = 1 - \frac{TC_b}{16}$. The macroblock's level ρ is simply the average of these local ρ_b's: $\rho = \frac{1}{16}\sum_{b=1}^{16} \rho_b$. Using the statistical model that follows, we calculate once the average code lengths $\bar{c}_{(TC,Tr)}(\rho_b | context\, prior)$ and $\bar{c}_{TZ}(\rho_b | input\, prior)$ of the (TotalCoefficients, TrailingOnes) and TotalZeros syntax elements, respectively. These average lengths are kept in look-up tables and the rate "overhead" component is obtained by averaging over all the blocks in the macroblock:

$$r^{overhead}(\rho) = \frac{1}{16}\sum_{b=1}^{16} \bar{c}_{(TC,Tr)}(\rho_b | context\, prior) + \frac{1}{16}\sum_{b=1}^{16} \bar{c}_{TZ}(\rho_b | input\, prior) \tag{35}$$

We assume that the quantized transform coefficients are not correlated and follow a Laplacian distribution. Another assumption is that all ± 1 quantized coefficients appearances occur at the highest nonzero frequencies, and are thus considered as high-frequency trailing-ones. Using the Laplacian distribution, the probability that the magnitude of a quantized transform coefficient, l, will take the value k is:

$$Pr.(|l| = k) = \begin{cases} \rho & k = 0 \\ \frac{(1-\rho)^{2k}\rho(2-\rho)}{1-\rho} & k > 0 \end{cases} \tag{36}$$

and therefore the probability of a trailing-one coefficient, given that it is non-zero is: $Pr.(TR) = Pr.(|l| = 1||l| > 0) = \rho(2-\rho)$.

We define a binomial random variable that denotes the number of trailing-ones appearances given ρ_b and sum over the joint (TotalCoefficients, TrailingOnes) code length tables (there are 4 different tables) to obtain the average VLC tables $\overline{c}_{(TC,Tr)}(\rho_b|context\,prior)$. We switch between these four average VLC tables by predicting the number of non-zero coefficients from the neighboring blocks, in accordance with the standard's context-based encoding.

Since the quantized blocks are typically sparse and most of the energy is concentrated at low frequencies, there is usually a tail of zeros at the end of the scanned block (see example in Fig.13). So, instead of counting the TotalZeros syntax element, TZ, as the number of zeroed coefficients from the DC coefficient to the highest frequency non-zero coefficient, we can count its complement, the tail, since $TC + TZ + Ztail = 16$. As we increase the requantization step, the number of non-zero coefficients, TC, decreases, and the tail length monotonically increases. Therefore, $TC + TZ$ monotonically decreases.

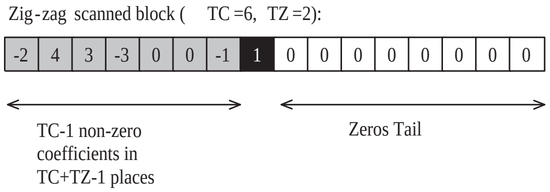
Zig-zag scanned block (TC =6, TZ =2):
-2
4
3
-3
0
0
-1
1
0
0
0
0
0
0
0
0
TC-1 non-zero coefficients in TC+TZ-1 places
Zeros Tail

Fig. 13. The example of Fig.11 with TC, TZ and the zeros tail. There are TC=6 non-zero coefficients and TZ=2 zeros counted from the DC coefficient to the highest frequency non-zero coefficient (which is denoted in black).

Given the input prior information (TC_{in}, TZ_{in}), we find the probability of having TZ TotalZeros given ρ_b. The average code length for each of the 15 (TC_{in}, TZ_{in}) input priors is evaluated by summing over the joint (TotalCoefficients,TotalZeros) code length tables.

Finally, the total rate-ρ relation is evaluated by:

$$r(\rho) = r^{data}(\rho) + r^{overhead}(\rho) \tag{37}$$

where $r^{data}(\rho)$ and $r^{overhead}(\rho)$ are evaluated from (33) and (35), respectively.

3) Distortion$-\rho$ model: The PSNR is a widely used objective quality metric that is related to the MSE distortion. That is why we examined the validity of the exponential $distortion - \rho$ model suggested in [3] in describing the MSE. According to this model, $ln(\overline{d}(\rho))$ should be linearly proportional to $1-\rho$, where $\overline{d}(\rho) = \frac{d(\rho)}{\sigma^2}$ is the normalized distortion. Examining this relation at the macroblock level, we found that a linear model does not describe it with sufficient accuracy. We therefore suggest to extend the model to an exponential-quadratic relation:

$$d(\rho) = \sigma^2 \cdot e^{\alpha_1 \cdot (1-\rho)^2 + \alpha_2 \cdot (1-\rho)} \tag{38}$$

that better matches the empirical data, see Fig. 14 and a quantitative accuracy comparison in Table II.

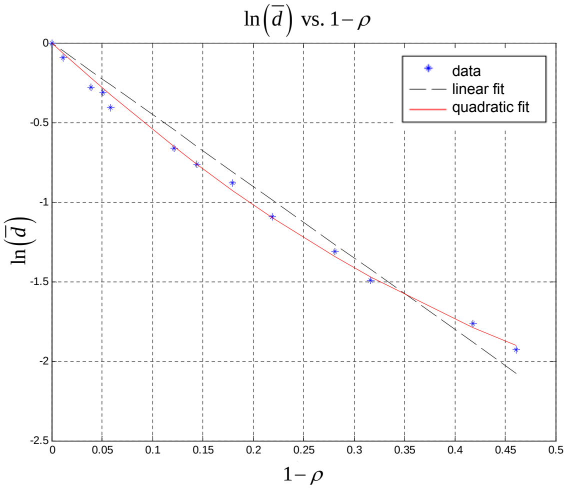

Fig. 14. Distortion$-\rho$ model. Blue points: $ln(\overline{d}(\rho))$; black dashed line: linear fit; red solid line: quadratic fit.

The modified $disortion - \rho$ model has three parameters that need to be estimated: α_1, α_2, and σ^2. Since we can only measure the requantization distortion, and not the total degradation from the reference, as we do not have the signal at the input of the first encoder. The scale parameter σ^2 is calculated once as the sum of squares of the input transform coefficients, as this would be the MSE if the block is zeroed. Given the scale parameter, we evaluate the normalized distortion $\overline{d}(\rho)$, that has two parameters to be estimated: α_1, α_2. To this end, we should get two different (ρ, d) points. Our suggestion is to first evaluate the $\rho - Q_2$ relation for each macroblock, see subsection IV-B4. Then, estimate the distortion at the finest requantization step-size (that is coarser or equal to the original step) that corresponds to a fraction ρ_1 of zeroed coefficients. Based on ρ_1, we would like to find a second point, ρ_2, far enough from both ρ_1 and 1. We arbitrarily choose ρ_2 such that $1 - \rho_2 \simeq \frac{1}{2} \cdot (1 - \rho_1)$. Since we can only find ρ_2 at the resolution of the available quantization step-sizes, we choose the closest available ρ_2 (using the $\rho - Q_2$ table we already have at hand). Based on these two points, $(1 - \rho_1, ln(\overline{d_1}))$ and $(1 - \rho_2, ln(\overline{d_2}))$, we can estimate the quadratic fit for $ln(\overline{d})$ vs. $1 - \rho$ curve (see illustration in Fig. 15) and extract the α_1, α_2 parameters. The luminance and chrominance components are modeled separately.

4) $\rho-Q_2$ relation: Contrary to intra-coded frames, the estimation of ρ for inter-coded frames is fairly simple and has a low computational complexity. Since the inter-coded blocks are predicted using previously decoded frames,

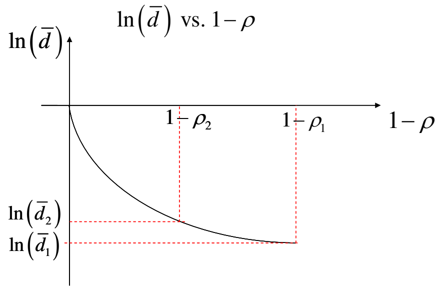

Fig. 15. Parameters estimation for the $distortion - \rho$ model.

their closed loop correction signal is available and the models evaluation is performed based on the corrected transform coefficients to be requantized.

Therefore, we count the number of coefficients that fall in the second quantizer deadzone, $[-t(Q_2), t(Q_2)]$, where $t(Q_2) = (1 - \Delta z)Q_2$ and Δz is the deadzone. The $\rho - Q_2$ relation is evaluated using this histogram count by normalizing the expected number of zeros at the quantizer output to the data size (either 256 coefficients or 128 coefficients for the luminance and chrominance MB components, respectively). It is evaluated for each macroblock for all the step-sizes that are coarser than the input step-size, prior to the rate and the distortion evaluation.

In case the selective elimination algorithm is applied, ρ is evaluated by applying the same histogram count on the quantized coefficients after elimination.

V. Results

In this section we summarize and report the main simulation results of the developed algorithm. The original video sequences were first encoded at 2[Mbps] using H.264 baseline profile and then transrated at four transrating ratios. The standard video sequences used for the analysis are 'flower garden', 'football', 'mobile & calendar' at SIF format (352x240 resolution) and 'foreman' at a CIF format (352x288 resolution). We also examined the 'pedestrian' sequence at SDTV format (720x576 resolution) originally encoded at 8[Mbps].

A. System architecture

The chosen system architecture is FD-GE for intra frames and PD-PE for inter frames (see subsection I-B). The PD-PE architecture reduces the run-time of inter frames transrating by about 15% as compared to a FD-GE architecture, at negligible quality loss. If the FD-GE architecture is used for inter frames too, one could also modify the input motion vectors (MVs). Our attempt to modify the input MVs by locally merging them has shown that a further MV refinement search is required to avoid quality degradation. Such a refinement further increases the computational complexity, therefore we chose to reuse the input motion decisions. Another extension of our work using the FD-GE architecture for inter frames is discussed in subsection V-E.

B. Intra Frames Transrating

1) Model-based Uniform Requantization: In subsection III-A2, we proposed a closed-loop statistical model for estimating the $\rho(Q_2)$ relation for an intra frame. Fig. 16 depicts an example for this $\rho(Q_2)$ estimator at the frame level, as compared to other estimators. The open-loop estimator is biased as compared to the true data relation and as noted earlier has a staircase characteristic. The proposed estimators follow closely the data and their average relative error is less than 1.7%. We examined the average rate deviation from the target, where the uniform requantization step-size was selected using different $\rho(Q_2)$ estimators, as listed in Table I. The true data $\rho(Q_2)$ relation was used as a yardstick for the performance, as it cannot be evaluated in a real-time scenario. It shows some small rate estimation error (2.5%), mainly because of the rate-ρ model's inaccuracy. Due to the inherent bias of the open-loop estimator, it tends to choose finer steps than required, resulting in increased rate. That is, it has a large rate estimation error. The proposed $\rho(Q_2)$ estimator outperforms the open-loop estimator, providing a smaller rate estimation error, close to the rate estimation error from the true data.

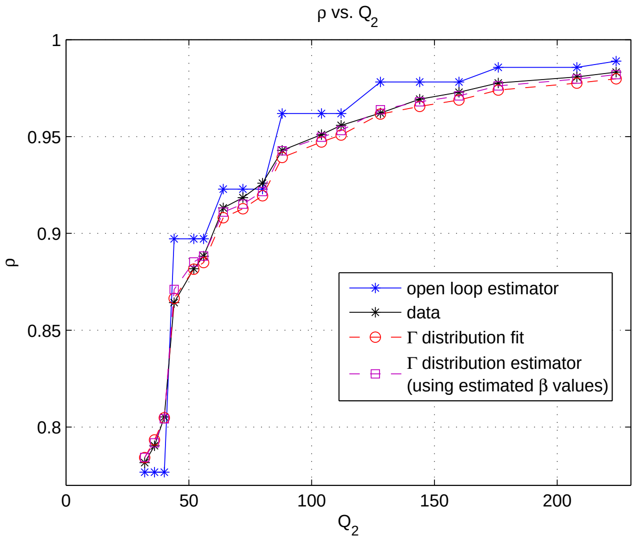

Fig. 16. Frame level $\rho(Q_2)$ relation (from the 'flower garden' sequence). Blue x: open-loop estimator. Black asterisk: data. Red circles: proposed estimator (offline β evaluation using ML). Magenta squares: proposed estimator (using estimated β values).

TABLE I

MEAN RELATIVE RATE DEVIATION FROM THE TARGET, MEASURED FOR THE 4 EXAMINED SEQUENCES INITIALLY ENCODED AT 2[MBPS], AT INTRA TRANSRATING FACTORS OF 1.5 TO 3.

$\rho(Q_2)$ estimator	Mean relative rate deviation [%]
True data	2.5
Open-loop	10.8
Proposed (closed-loop)	3.0

2) Modification of Prediction Modes: In subsection III-B2, we considered weighting the distortion of {texture regions, smooth regions, and edges} differently. Since HVS-based considerations are not in the main focus of our work, the weighting factors were set empirically. We found that $f_{texture} = 1.2$, $f_{smooth} = 0.8$, and $f_{edge} = 1$ are suitable. The visual effect of the prediction modes modification is more noticeable at smooth regions, e.g. the sky in the 'flower garden' sequence. Reusing the input prediction modes reduces the run-time of intra frame transrating by a factor of about 4.5, on average, as compared to re-encoding, at a PSNR loss of up to 1[dB] (for a transrating factor of 3). The proposed selective modes modification scheme, suggested in subsection III-B, has practically the same performance as the intra frame re-encoding scheme in terms of PSNR vs. bit rate, at about 37.5% less computations. Comparing the two methods, reuse of input modes is faster and more suitable for small transrating factors, since the transrated frame prediction modes are expected to be similar to the input modes.

C. Inter Frames Transrating

The motivation of using the rate-distortion models proposed in subsection IV-B is to provide an accurate and low computational rate-distortion evaluation. We now discuss the performance of the proposed MB-level models in terms of accuracy and computational complexity. Fig. 17 depicts the mean rate-model error, for both the proposed and the linear rate models, measured as the deviation of the model-based rate estimation from the actual encoded number of bits. The proposed rate-ρ model errors are smaller than the linear rate-ρ model errors. As the bit rate is reduced, the "overhead" component in the rate model gets more dominant and more accurate. As a result, the overall accuracy of the proposed model is improved for higher transrating ratios. Table II compares the accuracy of the exponential-linear distortion-ρ model suggested in the literature with the proposed exponential-quadratic model. It shows an average error of only 2% for the proposed exponential-quadratic model vs. 11% for the exponential-linear model. To evaluate the computational complexity of the inter frames transrating its two phases shoud be considered: model-based rate-distortion evaluation, and the optimization procedure. We compared the run time of inter frame transrating, when an optimal requantization is performed (see subsection IV-A), once using the proposed rate-distortion models and once using an exhaustive rate-distortion evaluation (i.e., without models). By evaluating the proposed rate-distortion models, the run time is reduced by a factor of about 2.3, on average, as compared to the exhaustive evaluation. As for the optimization procedure complexity, it takes about 6 to 8 Lagrangian iterations until convergence. Each such Lagrangian iteration requires $MB_{num}QP_{num}$ basic operations of finding the best previous value (minimum of a 7-length array), where MB_{num} is the number of macroblocks in the frame and $QP_{num} = 52$.

D. Overall System Performance

We summarize and compare by simulations the following transrating algorithms:

- Re-encoding.
- Proposed algorithm.

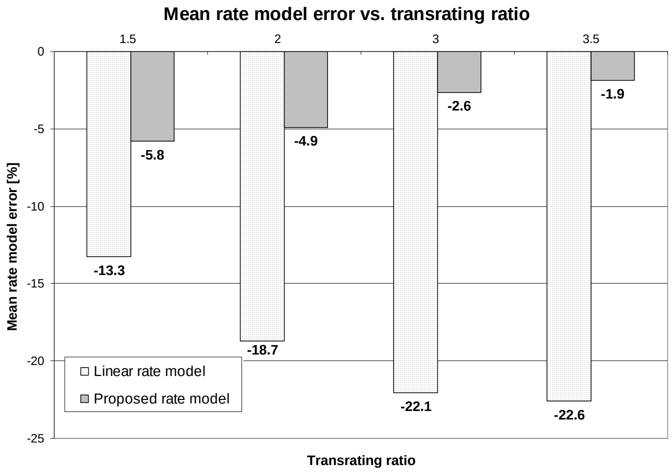

Fig. 17. Mean rate-model error vs. transrating ratio. White: linear rate model. Gray: proposed rate model.

TABLE II

MEAN ABSOLUTE DISTORTION RELATIVE ERROR [%], MEASURED FOR THE 4 EXAMINED SEQUENCES INITIALLY ENCODED AT 2[MBPS], AT TRANSRATING FACTORS OF 1.5 TO 3.5

	Mean absolute distortion relative error [%]	
Transrating ratio	Exponential-linear model	Exponential-quadratic model
1.5	19.04	1.04
2	10.66	2.50
3	7.56	2.52
3.5	6.65	2.57

- One-pass requantization [38] (processes one macroblock at a time and sets its requantization step-size according to the output buffer fullness. For fair comparison, it also uses the optimal GOP level bit allocation suggested in section II).

The original video sequences (’flower garden’ SIF, ’football’ SIF, ’mobile and calendar’ SIF and ’foreman’ CIF) were first encoded at 2[Mbps] using H.264 baseline profile, with a GOP structure of an I-frame followed by 14 P-frames and no frame skipping allowed. The encoding was done using ”Nokia” H.264 baseline encoder. These were then transrated at four transrating ratios. The PSNR vs. bit rate graph for the ’football’ sequence is depicted in Fig. 18. As expected, it rates the performance of the transrating algorithms at the following order, from the best to the worst quality: re-encoding; the proposed algorithm and one-pass requantization. It also shows that the selective intra modes modification (denoted by red circles) has better performance than reusing the input intra modes (denoted by black x). It should be noted that for fair comparison, the model-based optimal GOP level bit allocation was applied for the one-pass method too, but it is more likely that such a simple requantization would use a simpler GOP allocation as well, which is expected to further decrease its performance. The same ranking of the algorithms in terms of PSNR vs. bit rate turned out at other cases, e.g. transrating ’football’ SIF initially encoded at 1[Mbps] and ’foreman’ QCIF initially encoded at 250[Kbps]. We also compared the tested algorithm at SDTV resolution by transrating the SD ’pedestrian’ sequence originally encoded at 8[Mbps], as depicted in Fig. 19. The results show the same algorithms ranking concluded from our previous experiments at lower spatial resolutions, with larger PSNR gaps.

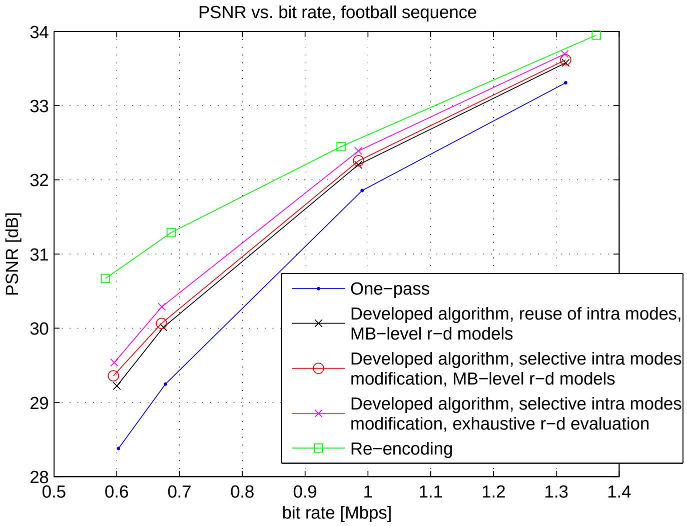

Fig. 18. PSNR vs. bit rate, for transrating a SIF ’football’ sequence, initially encoded at 2[Mbps]. Blue dots: one-pass requantization. Black x: developed algorithm, reuse of intra modes, MB-level r-d models. Red circles: developed algorithm, selective intra modes modification, MB-level r-d models. Magenta x: developed algorithm, selective intra modes modification, exhaustive r-d evaluation. Green squares: re-encoding.

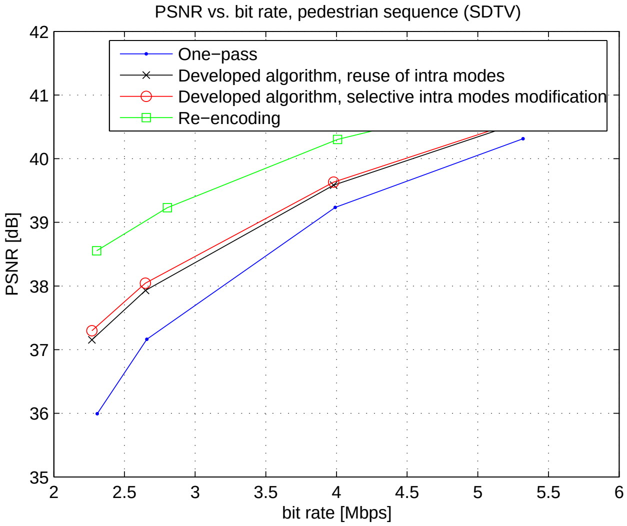

Fig. 19. PSNR vs. bit rate, for transrating a SD ’pedestrian’ sequence, initially encoded at 8[Mbps]. Blue dots: one-pass requantization. Black x: developed algorithm, reuse of intra modes, MB-level r-d models. Red circles: developed algorithm, selective intra modes modification, MB-level r-d models. Green squares: re-encoding.

The overall system performance is measured in terms of computational complexity (by run-time) and quality (by the PSNR difference). The quality vs. computational complexity, for the different algorithms, as compared to the proposed algorithm, is depicted by the black solid curve in Fig. 20. The graph shows average results over four video sequences encoded at 2[Mbps] and transrated to 1[Mbps]. As compared to re-encoding, the proposed algorithm saves the run-time by a factor of about 4, on average, with small PSNR loss at high to medium bit rates. In comparison with a simple one-pass requantization, the proposed algorithm achieves better performance, at the cost of twice the complexity. In [6], the authors compare their algorithm with re-encoding and report on saving a factor of about 2 in the run-time at a PSNR loss of about 0.5 [dB], which is worse than our proposed system performance. By examining the graph slopes in Fig. 20, we conclude that the proposed system's gain, as compared to the one-pass requantization, is higher than the re-encoding gain as compared to the proposed system.

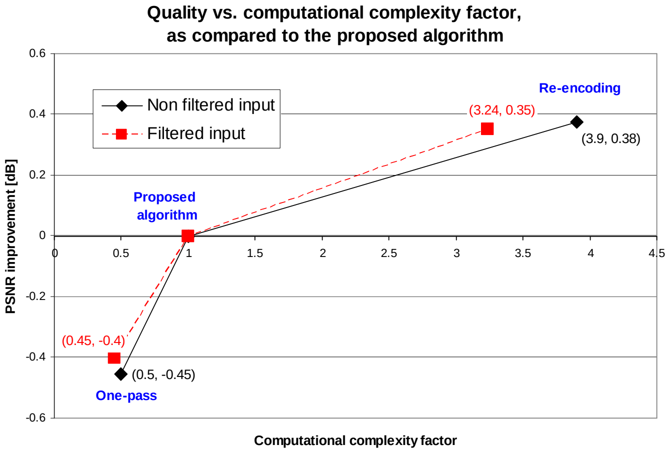

Fig. 20. Quality vs. computational complexity of re-encoding and one-pass algorithm, as compared to the proposed algorithm, for SIF/CIF resolution sequences. The quality is measured by PSNR improvement, and the computational complexity is measured by the run-time factor. Black solid diamond: Input encoded without deblocking filter, Red dashed square: Input encoded with deblocking filter.

E. Support of Input Coded with Deblocking Filter

H.264 may apply an adaptive in-loop deblocking filter on the decoded pictures to reduce blocking artifacts [39]. However, it is not clear whether the computational cost of the filter is justified considering the improvement in subjective quality [40]. In this work, we assumed that the deblocking filter was disabled during encoding of the input video and its transrating. To support input video that was initially encoded using the deblocking filter, we propose to fully decode the input (including the in-loop filtering) and then encode according to our algorithm without applying the filter. To evaluate the performance, we ran again the tests described in subsection V-D for an input video initially encoded with the deblocking filter, see red dashed curve in Fig. 20. Here, the proposed system run-time increases due to the decoding with a deblocking filter and therefore the complexity saving factor as compared to re-encoding is somewhat reduced. Still, the proposed system provides a good trade-off between quality and computational complexity.

VI. Conclusion

A model-based transrating system for H.264 encoded video via requantization is proposed. To keep a smooth constant video quality, it applies an optimal GOP level bit allocation that equalizes the frame distortions. For intra-coded frames, a uniform requantization step-size is chosen using the linear rate-ρ model and a novel closed-loop statistical estimator for the $\rho - Q_2$ relation. This estimator overcomes the spatial-block dependency problem by modeling the correction signal of the requantized residual. For the examined sequences, its average rate deviation from the target is 3%, as compared to 10.8% average deviation obtained by using an open-loop $\rho - Q_2$ estimator. The guided intra frames transrating allows to either reuse the input prediction modes, or selectively modify them, reducing the computational complexity. For inter-coded frames, a new optimal non-uniform requantization algorithm is developed, where the changes in the requantization step-sizes throughout the frame are regulated, to improve the subjective quality. To reduce that optimization computational burden, we suggest new macroblock level rate-distortion models in the ρ-domain, adapted to H.264 requantization. The incorporation of these models reduces the run-time of inter frames transrating by a factor of about 4, on average, with only a small PSNR loss at high to medium bit rates, for SIF/CIF resolution sequences.

References

[1] P. Assuncao and M. Ghanbari, “A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams,” *IEEE transactions on Circuits and Systems for Video Technology*, vol. 8, no. 8, pp. 953–967, Dec. 1998.

[2] J. Lan, W. Zeng, and X. Zhuang, “Operational distortion-quantization curve-based bit allocation for smooth video quality,” *Signal Processing: Image Communications*, vol. 16, pp. 527–543, 2005.

[3] Z. He and S. Mitra, “Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling,” *IEEE transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–894, Oct. 2002.

[4] I. Richardson, *H.264 and MPEG-4 Video Compression*. John Wiley, 2003.

[5] P. Zhang, Q. Huang, and W. Gao, “Key techniques of bit rate reduction for H.264 streams,” in *Lecture Notes in Computer Science, Book Advances in Multimedia Information Processing - PCM 2004*. Springer, Oct. 2004, pp. 985–992.

[6] H. Nam *et al.*, “Low complexity H.264 transcoder for bitrate reduction,” in *International Symposium on Communications and Information Technologies, ISCIT*, Bangkok, Thailand, Oct. 2006, pp. 679–682.

[7] D. Lefol, D. Bull, and N. Canagarajah, “An efficient complexity-scalable video transcoder with mode refinement,” *Signal Processing: Image Communications*, vol. 22, pp. 421–433, Apr. 2007.

[8] Z. He and S. Mitra, “A linear source model and a unified rate control algorithm for DCT video coding,” *IEEE transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970–982, Nov. 2002.

[9] “H.264 reference software,” http://bs.hhi.de/∼suehring/tml/download/.

[10] N. Hait and D. Malah, “Towards model-based transrating of H.264 coded video,” in *The 2006 IEEE 24’th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, Nov. 2006.

[11] N. Hait and D. Malah, “Model-based transrating of H.264 intra-coded frames,” in *Picture Coding Symposium - PCS’2007*, Lisbon, Portugal, Nov. 2007.

[12] T. Wiegand *et al.*, “Overview of the H.264/AVC video coding standard,” *IEEE transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[13] Z. He, Y. Kim, and S. Mitra, “Low-delay rate control for DCT video coding via ρ-domain source modeling,” *IEEE transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[14] S. Milani, L. Celetto, and G. Mian, “A rate control algorithm for the H.264 encoder,” in *Sixth Baiona workshop on Signal Processing in Communications*, Spain, Sep. 2003.

[15] I. Shin, Y. Lee, and H. Park, "Rate control using linear rate-ρ model for H.264," *Signal Processing: Image Communications*, vol. 19, no. 4, pp. 341–352, Apr. 2004.

[16] H. Sun, X. Chen, and T. Chiang, *Digital video transcoding for transmission and storage*. CRC press, 2005.

[17] I. Ahmad *et al.*, "Video transcoding: An overview of various techniques and research issues," *IEEE transactions on multimedia*, vol. 7, no. 5, pp. 793–804, Oct. 2005.

[18] Z. Lei and N. Georganas, "Rate adaptation transcoding for precoded video streams," in *Proceedings of the tenth ACM international conference on Multimedia*, Juan-les-Pins, France, Dec. 2002, pp. 127–136.

[19] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: an overview," *IEEE signal processing magazine*, vol. 20, no. 2, pp. 18–29, Mar. 2003.

[20] A. Vetro, J. Cai, and C. Chen, "Rate-reduction transcoding design for wireless video streaming," *Wireless Communications and Mobile Computing*, vol. 2, no. 6, pp. 625–641, Oct. 2002.

[21] D. Lefol, D. Bull, and N. Canagarajah, "Performance evaluation of transcoding algorithms for H.264," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, pp. 215–222, Feb. 2006.

[22] H. Sun, W. Kwok, and J. Zdepski, "Architectures for MPEG compressed bistream scaling," *IEEE transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 191–199, Apr. 1996.

[23] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 23–50, Nov. 1998.

[24] M. Militzer, M. Suchomski, and K. Meyer-Wegener, "Improved ρ-domain rate control and perceived quality optimizations for MPEG-4 real-time video applications," in *International Conference on Multimedia*, 2003, pp. 402–411.

[25] Y. Altunbasak and N. Kamaci, "ρ domain rate-distortion optimal rate control for DCT-based video coders," in *International Conference on Acoustics, Speech, and Signal Processing*, May 2004.

[26] C. Chen, P. Wu, and H. Chen, "Transform-Domain Intra Prediction for H.264," in IEEE I*nternational* S*ymposium on* C*ircuits and* S*ystems (*ISCAS *2005)*, May 2005, pp. 1497–1500.

[27] A. Papoulis, *Probability, random variables, and stochastic processes*, 2nd ed. McGraw-Hill, 1986.

[28] D. Lefol, D. Bull, and N. Canagarajah, "Mode refinement algorithm for H.264 intra frame requantization," in *International Symposium on Circuits and Systems*, 2006, pp. 4459–4462.

[29] L. Torres and M. Kunt, *Video Coding: The second generation approach*. Kluwer Academic Publishers, 1996, ch. 6, Segmentation-based motion estimation for second generation video coding techniques.

[30] K. Minoo and T. Nguyen, "Perceptual video coding with H.264," in *IEEE conference on Signals, Systems and Computers*, 2005.

[31] A. Nguyen and J. Hwang, "A novel hybrid HVPC/mathematical model rate control for low bit-rate streaming video," *Signal Processing: Image Communication*, vol. 17, pp. 423–440, 2002.

[32] M. Lavrentiev, "Transrating of Coded Video Signals via Optimized Requantization," M.Sc. thesis, TECHNION, 2004.

[33] M. Lavrentiev and D. Malah, "Transrating of MPEG-2 coded video via requantization with optimal trellis-based dct coefficients modification," in *XII European Signal Processing Conference - Eusipco*, Sep. 2004, pp. 1963–1966.

[34] W. Wang, H. Cui, and K. Tang, "Rate distortion optimized quantization for H.264/AVC based on dynamic programming," *Visual Communications and Image Processing, Proceedings of the SPIE*, vol. 5960, pp. 2100–2111, Jul. 2005.

[35] R. Lagendjik, E. Frimout, and J. Biemond, "Low-complexity rate-distortion optimal transcoding of MPEG I-frames," *Signal Processing: Image Communication*, vol. 15, pp. 531–544, 2000.

[36] A. Eleftheriadis and D. Anastassioiu, "Constrained and general dynamic rate shaping of compressed digital video," in *International Conference on Image Processing*, 1995, pp. 396–399.

[37] P. Carlsson, F. Pan, and L. T. Chia, "Coefficient thresholding and optimized selection of the lagrangian multiplier for non-reference frames in H.264 video coding," in *International Conference on Image Processing*, 2004, pp. 773–776.

[38] N. Hait, "Model-Based Transrating of Coded Video," Master's thesis, TECHNION, 2007, Downloadable from http://sipl.technion.ac.il/siglib/FP/Hait.pdf.

[39] P. List *et al.*, "Adaptive deblocking filter," *IEEE transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 614–619, Jul. 2003.

[40] Y. Zhong *et al.*, “Perceptual quality of H.264/AVC deblocking filter,” in *IEE International Conference on Visual Information Engineering*, Apr. 2005, pp. 379–384.