

MODEL-BASED TRANSRATING OF H.264 INTRA-CODED FRAMES

Naama Hait and David Malah

Technion IIT, Haifa 32000, Israel
 Department of Electrical Engineering
 naama.hait@gmail.com, malah@ee.technion.ac.il

ABSTRACT

This paper presents a transrating (bit-rate reduction) algorithm for H.264 intra-coded frames via requantization. Previous works focused on adapting the input prediction modes to the lower bit rate and hence performed requantization using a one-pass algorithm. We propose a model-based algorithm for uniform requantization of the transform coefficients in intra-coded frames. The spatial prediction in such frames introduces block dependencies. We suggest a novel statistical-based closed-loop model for estimating the relation between the rate and the requantization step that overcomes the dependency problem. The performance of an overall transrating system for H.264 coded video, incorporating this work for intra-coded frames and our previous work for inter-coded frames is also examined.

Index Terms— Requantization, H.264, video coding

1. INTRODUCTION

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 [1], at the expense of higher complexity. As the choices of quantization step-size and coding modes are dependent, the rate control becomes computationally expensive.

Transrating of coded video is the process of reducing the bit rate of a high-quality pre-encoded video to match user-specific bit rate requirements. Requantization of the transform coefficients is a common approach for video transrating. However, previous works on transrating in H.264 [2, 3, 4] focus on changing the input coding decisions (intra prediction modes and motion) rather on the rate control, and the requantization is addressed by a simple one-pass algorithm [2]. In our previous work [5], we suggested a model-based optimal requantization for H.264 *inter-coded* frames (P-frames).

In this paper, we focus on the requantization of H.264 *intra-coded* frames (I-frames). As the spatial prediction in I-frames introduces dependencies between neighboring residual blocks, full decoding is required to avoid a drift error. Moreover, due to these dependencies, the residual coefficients

to be requantized are not available in advance, when the requantization step-size should be selected. We propose to select a uniform requantization step, using ρ domain models, where ρ is the fraction of zero quantized coefficients [6]. Section 2 reviews the open-loop approach for requantization step-size determination. Section 3 proposes a closed-loop statistical estimator for the relation between ρ and the requantization step-size. It overcomes the block dependency problem by modeling the correction signal of the requantized residual.

2. OPEN-LOOP APPROACH FOR REQUANTIZATION STEP-SIZE SELECTION

Different models in the literature suggest different relations for rate vs. quantization step size. In [6], a robust linear *rate*– ρ model is suggested: $R(\rho) = \theta \cdot (1 - \rho)$, where ρ is the fraction of zero coefficients among the quantized transformed coefficients in a frame. This model is used to set a uniform requantization step-size for an I-frame. The model parameter θ is estimated using the input rate– ρ point, (ρ_{in}, R_{in}) and an anchor point at $(1, 0)$, see Fig. 1(a). Given the target rate for that frame, R_{target} , we extract the expected fraction of zeros by $\rho_{target} = 1 - R_{target}/\theta$. 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 3. Then, the target step is found by $Q_{2,target} = f^{-1}(\rho_{target})$.

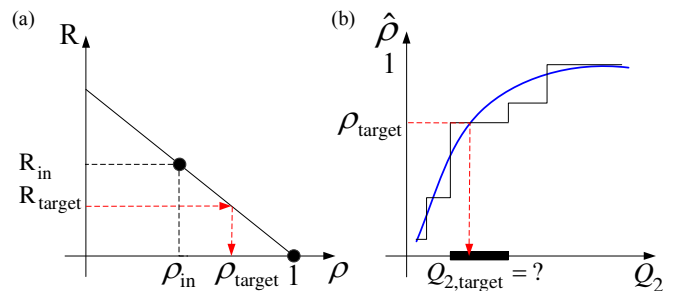


Fig. 1. Uniform requantization using a *rate*– ρ model. (a): *rate*– ρ relation. (b): ρ – Q_2 relation, where black staircase: open-loop estimator; blue: closed-loop estimator.

Due to spatial prediction, requantization of the prediction residual at one block changes the residual in neighboring ca-

sual blocks. To avoid a drift error, I-frames are fully decoded to 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. Since the requantization is performed in the transform domain, we seek for a transform domain $\rho(Q_2)$ estimator.

The simplest $\rho - Q_2$ estimator is the *open-loop estimator*, evaluated from the output of the scheme depicted in Fig. 2. 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 dz , the output indices are derived by $Z_{out} = \text{sign}(Y) \cdot \lfloor \frac{|Y|}{Q_2} + dz \rfloor$. Therefore, all transform coefficients that fall in the interval $[-t(Q_2), t(Q_2)]$ are requantized to zero, where $t(Q_2) = (1 - dz)Q_2 = \frac{2}{3}Q_2$. This process is repeated for each Q_2 step-size, to derive the $\rho - Q_2$ relation.

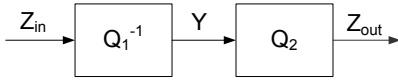


Fig. 2. Open-loop requantization scheme.

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. Given a target ρ value, the estimator may encounter an uncertainty as to which requantization step-size to choose, see Fig. 1(b).

3. 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 and requires special attention.

3.1. Closed-loop residual modeling architecture

We propose to estimate $\rho(Q_2)$ using a closed-loop residual modeling architecture in the transform domain, as depicted in Fig. 3. The closed-loop estimator models the required correction of the requantized residual coefficients, thereby overcoming the dependency problem. Instead of evaluating $\rho(Q_2)$ based on Y , it estimates how many corrected transform coefficients W fall in the deadzone interval. The corrected residual is defined as $W \triangleq Y - C$, where C is the correction signal. This signal is formed by feeding the transrating error ε , into the transform-domain spatial-predictor. 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 input and output images, where the output image is decoded using the requantized indices $Z_{out} = Q_2(W)$.

This scheme is merely used in order to model the distribution of W , from which ρ is estimated. During actual transrating, we do not follow this scheme that calculates exactly the output Z_{out} for each step Q_2 .

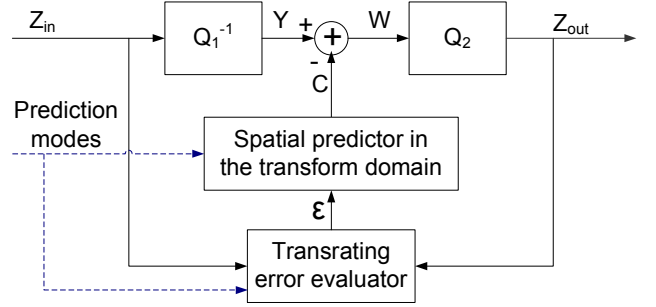


Fig. 3. A closed-loop modeling scheme for estimating $\rho(Q_2)$. The transrating error ε is fed into the predictor to yield the correction signal C . Then, $\rho(Q_2)$ is estimated based on $W \triangleq Y - C$.

In order to evaluate $\rho(Q_2)$ from W , we first characterize the 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 quantization step-size Q_1 , their distribution is discrete, and given as:

$$p_Y(y) = \sum_{m=-M}^M p_m \cdot \delta(y - mQ_1) \quad (1)$$

where $\delta(y)$ is the impulse function and $\{p_m\}_{m=-M}^M$ 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 (section 3.2) and its statistical modeling (section 3.3). 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. 4. As we cannot assume that C is independent of Y , we use the joint probability of $(Y, -C)$: $p_{Y,-C}(y, c) = p_{-C|Y}(c|y) \cdot p_Y(y)$ to calculate the cumulative distribution of W :

$$\begin{aligned} Pr.(W \leq w_0) &= \int_{-\infty}^{\infty} \int_{-\infty}^{w_0-y} p_{Y,-C}(y, c) dc dy = \\ &= \sum_{m=-M}^M p_m \cdot \int_{-\infty}^{w_0-mQ_1} p_{-C|Y}(c|Y = mQ_1) dc \end{aligned} \quad (2)$$

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

$$\rho(Q_2) = Pr.(|W| \leq t(Q_2)) = \sum_{m=-M}^M p_m \cdot \phi(m|Y) \quad (3)$$

where

$$\phi(m|Y) = \int_{-t(Q_2)-mQ_1}^{t(Q_2)-mQ_1} p_{-C|Y}(c|Y = mQ_1) dc \quad (4)$$

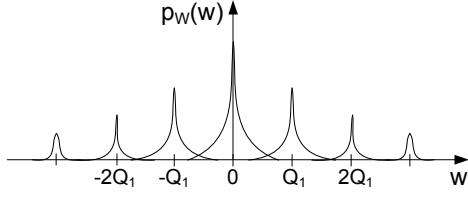


Fig. 4. Schematic illustration of the probability distribution of W .

Lacking a known model for the correlation between Y and C , we are left with the unfeasible task of modeling $\phi(m|Y)$, for every possible value of Y (corresponding to $-M \leq m \leq M$). 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 $\phi(m|Y \neq 0)$ separately. In that case, the model in (5) for $\rho(Q_2)$ is simpler, as there are two possible input dependencies instead of $2M + 1$. To complete the evaluation of $\rho(Q_2)$, the following two subsections address the evaluation of $\phi(0|Y=0)$ and $\phi(m|Y \neq 0)$, by characterizing the correction signal C and modeling its distribution.

$$\rho(Q_2) = p_0 \cdot \phi(0|Y=0) + \sum_{m=-M, m \neq 0}^M p_m \cdot \phi(m|Y \neq 0) \quad (5)$$

3.2. 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 pre-defined for each prediction mode by an "affected coefficients mask" whose shape is characterized by the prediction mode type, see Fig. 5. 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.

The third partition distinguishes between the corrections applied to zero/non-zero input coefficients, following the approximation we made in section 3.1. In the next subsection, a probability distribution is fitted to each data group allowing evaluation of its $\rho - Q_2$ relation according to (5).

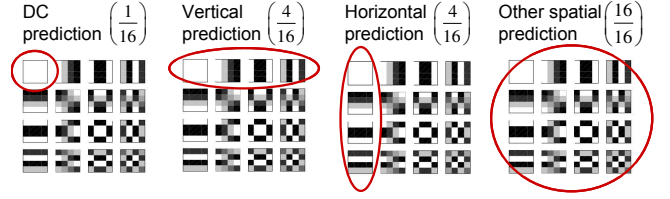


Fig. 5. Affected/unaffected transform coefficients map denoted over the ICT basis images. The classification is done according to the prediction modes. Affected coefficients are encircled in red, and their fraction is denoted in parenthesis.

3.3. Correction signal modeling using a Γ distribution

To evaluate (5) for each data group, a statistical description of $\phi(0|Y=0)$ and $\phi(m|Y \neq 0)$ is required. To study the issue, we evaluated the correction signal C offline, according to the scheme of Fig. 3, and performed the partitioning described in subsection 3.2. 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 [7]:

$$p_X(x; \beta) = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{\beta}{|x|}} \cdot \exp\{-\beta|x|\} \quad (6)$$

where $\beta > 0$ is the scale parameter, whose decrease results in a wider distribution. The Γ cumulative distribution function is defined by (7), where $\Gamma(a, 0.5) \triangleq \int_0^a t^{-0.5} \exp(-t) dt$.

$$Pr.(X \leq x; \beta) = \frac{1}{2} + \text{sgn}(x) \frac{1}{2\sqrt{\pi}} \Gamma(\beta|x|, 0.5) \quad (7)$$

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 (7) and these estimated parameters, the functions $\phi(0|Y=0)$ and $\phi(m|Y \neq 0)$ take the form of (8), and $\rho(Q_2)$ can be evaluated for each data-group by substituting (8) into (5). Then, all data-groups $\rho - Q_2$ relations are linearly weighted to obtain the frame level relation.

$$\begin{aligned} \phi(0|Y=0) &= Pr.(|C| \leq t(Q_2); \beta^{C|Y=0}) \\ \phi(m|Y \neq 0) &= Pr.(|C + mQ_1| \leq t(Q_2); \beta^{C|Y \neq 0}) \end{aligned} \quad (8)$$

In a real-time scenario, the scheme of Fig. 3 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. 6. 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 $\beta \rightarrow \infty$. The $\|\varepsilon\|_1$ vs. Q_2 relation was 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 $\|Y\|_2$.

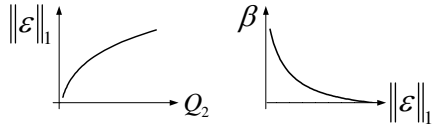


Fig. 6. Decomposition of the β vs. Q_2 relation, using $\|\varepsilon\|_1$.

4. RESULTS

Fig. 7 depicts an example for a $\rho - Q_2$ relation at the frame level. The open-loop estimator is biased as compared to the true data relation and as noted earlier has a staircase characteristic. The proposed estimators accurately follow the same trend as 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 1. 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, 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, at the cost of an increased rate. Therefore, 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 estimation from the true data.

We incorporated the proposed uniform requantization for I-frames with the non-uniform optimal requantization suggested in [5] for P-frames. As compared to re-encoding (cascaded decoder-encoder), the proposed system reduces the computational complexity by a factor of about 4, at a maximal cost of 1.4[dB] in PSNR. In comparison with a simple one-pass requantization, the proposed algorithm achieves better performance (PSNR gain of up to 1.6[dB]), at the cost of twice the complexity.

5. CONCLUSION

This paper suggests to solve the problem of H.264 I-frames requantization. The spatial prediction in I-frames introduces block dependencies, so that the residual coefficients to be requantized are not available in advance, when the requantization step-size should be selected. A uniform requantization step-size is chosen using the *rate* - ρ model. To this end, a novel closed-loop statistical estimator for the $\rho - Q_2$ relation was developed. Its average rate deviation from the target is 3%, as compared to 10.8% average deviation, obtained using an open-loop $\rho - Q_2$ estimator.

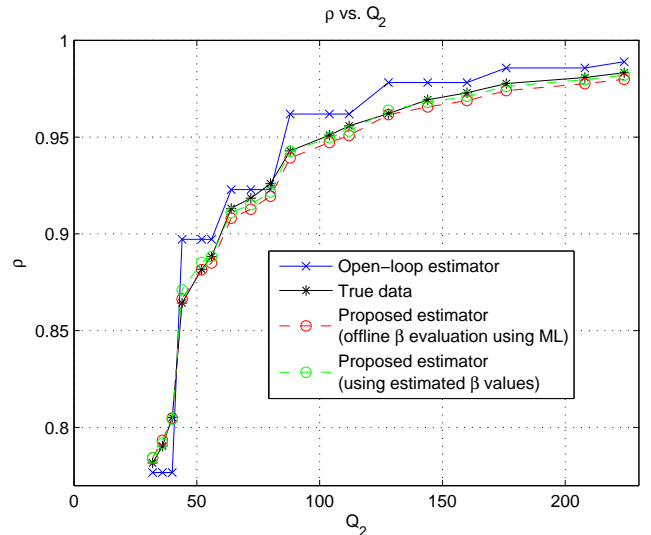


Fig. 7. Frame level $\rho - Q_2$ relation (from the 'flower garden' sequence). Blue x: open-loop estimator. Black asterisk: data. Red circles: proposed estimator (offline β evaluation using ML). Green circles: proposed estimator (using estimated β values).

Table 1. Mean relative rate deviation from the target, measured for the 'flower garden', 'football' and 'mobile & calendar' 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

6. REFERENCES

- [1] I.E.G. Richardson, *H.264 and MPEG-4 Video Compression*, John Wiley, 2003.
- [2] P. Zhang, Q.M. Huang, and W. Gao, "Key techniques of bit rate reduction for H.264 streams," in *Lecture Notes in Computer Science - PCM 2004*, pp. 985–992. Springer, Oct. 2004.
- [3] H.M. Nam et al., "Low complexity h.264 transcoder for bi-trate reduction," in *International Symposium on Communications and Information Technologies, ISCIT*, Bangkok, Thailand, Oct. 2006, pp. 679–682.
- [4] 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.
- [5] N. Hait and D. Malah, "Towards model-based transrating of H.264 coded video," in *IEEE 24th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, Nov. 2006, pp. 133–137.
- [6] Z. He and S.K. 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.
- [7] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw-Hill, 2nd edition, 1986.