# CODING GAIN AND TUNING FOR PARAMETRIZED VISUAL QUALITY METRICS

*S. de Waele and M. J. Verberne*

Philips Research Europe – Eindhoven
Eindhoven, The Netherlands.
[stijn.de.waele, michael.verberne]@philips.com

## ABSTRACT

A visual quality metric for video provides a mathematical expression for visual quality. We show how the coding gain of rate-distortion optimal (RDopt) encoding of the quantizer step size depends on the exact shape of this expression. For the Mean Square Error, we find a small coding gain (1% on average) compared to using a constant quantizer. However, we do find a significant gain for the Structural Similarity (SSIM) (12% on average). Finally, we show parameter tuning using the new Perceptually Weighted Error (PWE). The tuning results in a parameter setting of PWE that offers the best visual quality of RDopt encoding.

***Index Terms***— Video compression, rate-distortion optimization, objective visual quality metrics.

## 1. INTRODUCTION

The standard quality metric for evaluating the quality of compressed video is the Mean Square Error (MSE), or, equivalently, the Peak Signal to Noise Ratio (PSNR). Many quality metrics have been proposed that capture different aspects of the human visual system (HVS). Examples are the Picture Quality Rating (PQR) [1] and the Structural Similarly (SSIM) [2].

Besides using quality metrics for quality evaluation, they are also used in rate-distortion optimization for optimized encoding. The H.264 reference encoder contains rate-distortion optimization towards MSE for motion vector estimation and for mode decision [3]. Other authors report the usage of SSIM for the same coding decisions of motion vector estimation and mode decision [4;5].

In this paper, we examine the compression gain of rate-distortion optimization of the quantizer step size. As a constant quantizer step size is close to optimal for MSE, rate-distortion optimization for the quantizer is not needed in encoders that aim for the lowest MSE, or, conversely, maximum PSNR. However, it is known that adaptive quantization, where the quantization is modulated according to perception rules, yields a clearly better visual quality. Therefore, we expect a larger gain of rate-distortion optimization for visual quality metrics.

To examine the dependency of coding gain on the used quality metric, we introduce the parameterized Perceptually Weighted Error (PWE). Depending on parameter settings in this metric, PWE can be similar to both MSE and SSIM.

Besides measuring the influence on coding gain of parameter settings in PWE, we show how this metric tuned such that it gives the optimal visual quality for a given bitrate.

The outline of this paper is as follows. In section 2, we introduce PWE and describe its relation to existing quality metrics. Parameter tuning is discussed in section 3. Experimental results for the coding gain depending on PWE parameters are given in section 4. Finally, we give some concluding remarks in section 5.

## 2. QUALITY METRICS

In this section we first give the definition of the Mean Square Error (MSE) and Structural Similarity. Then we introduce the new Perceptually Weighted Error PWE. For brevity of notation, we will give the luminance contribution for the metrics discussed. The expressions for the chrominance contributions are very similar to that of luminance. They are included in the final distortion by weighted addition.

### 2.1 Mean Square Error / Peak Signal to Noise Ratio

The Mean Square Error (MSE) is a mean over all $N$ pixels in the sequence:

$$\mathrm{MSE} = \frac{1}{N}\sum_i (s_r - s)^2 , \qquad (1)$$

The Mean Square Error is equivalent to the Peak Signal to Noise Ratio, given by:

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}} \qquad (2)$$

### 2.2 Structural Similarity

The Structural Similarity (SSIM) is a sum of local contributions:

$$\mathrm{SSIM} = \sum_i d_{\mathrm{SSIM},i} , \qquad (3)$$

where the sum is taken over all $n$=16x16 macroblocks in the sequence. The local contribution is given by:

$$d_{\mathrm{SSIM},i} = \frac{2\mu\mu_r + C_1}{(\mu_r^2 + \mu^2 + C_1)} \frac{2\rho + C_2}{(\sigma_r^2 + \sigma^2 + C_2)}, \quad (4)$$

where ρ is the local cross-correlation between original and processed video **s** and $\mathbf{s_r}$. $\mu$ and $\mu_r$ are the local means; $\sigma^2$ and $\sigma_r^2$ are the local variances. The local contribution can be re-written as:

$$d_{\mathrm{SSIM},i} = \left(1 - \frac{(\mu_r - \mu)^2}{(\mu^2 + \mu_r^2 + C_1)}\right)\left(1 - \frac{\overline{(s'_r - s')^2}}{(\sigma^2 + \sigma_r^2 + C_2)}\right) \quad (5)$$

Where the *s'* are the variations around the local mean:

$$s' = s - \mu. \quad (6)$$

and $\overline{(s'_r - s')^2}$ is the mean square difference between *s* and *s'*:

$$\overline{(s'_r - s')^2} = \frac{1}{n}\sum_j (s'_{r,j} - s'_j)^2 \quad (7)$$

SSIM is a quality metric: a higher value means better quality. To obtain a distortion metric needed for rate-distortion optimization we use (1-SSIM). Following [2], we use $C_1 = (0.01 \cdot 255)^2 = 6.50$ and $C_2 = (0.03 \cdot 255)^2 = 58.52$.

We will give a short motivation for the second contribution of SSIM expression in equation (7), which is the most important contribution in the coding context. See reference [2] for a further motivation. The variance $\sigma$ of the original video is present to reflect masking: A given error in a high-activity area (high $\sigma$) is perceived less than the same error in a low-activity area. A completely proportional relation would result in extremely low SSIM values near $\sigma$=0, which is not in agreement with how errors are perceived. This explains the presence of the factor $C_2$.

### 2.3 Perceptually Weighted Error

As will be demonstrated by the experiments in section 3, SSIM has some disadvantages when used as a distortion metric in rate-distortion optimized encoding. We therefore introduce the new Perceptually Weighted Error. As SSIM, the Perceptually Weighted Error (PWE) is also a sum of local contributions:

$$\mathrm{PWE} = \sum_i d_{\mathrm{PWE},i}, \quad d_{\mathrm{PWE},i} = \frac{(\mu_r - \mu)^2}{\mu^{p_1} + k_1^{p_1}} + \frac{\overline{(s'_r - s')^2}}{\sigma^{p_2} + k_2^{p_2}} \quad (8)$$

With parameter settings $p_1$=$p_2$=0 we find that PWE is equivalent to MSE:

$$\mathrm{MSE} = 2\mathrm{PWE}[p_1 = p_2 = 0] \quad (9)$$

With $p_1$=$p_2$=0, the setting for $k_1$ and $k_2$ are irrelevant.

PWE and SSIM are equal in the second order approximation around $\mathbf{s}_r$ = **s** with the following parameter settings:

$$d_{\mathrm{SSIM}} = 1 - \frac{1}{2} d_{\mathrm{PWE}}[p_1 = p_2 = 2; k_i = \sqrt{C_i/2}] + \mathrm{O}(|\mathbf{s}_r - \mathbf{s}|^4) \quad (10)$$

Where $\mathrm{O}(|\mathbf{s}_r - \mathbf{s}|^4)$ denotes the higher order contributions. This result is derived by relating equation 5 for SSIM to the definition of PWE (equation 8). We conclude that with the right setting for $p_1$ and $p_2$, PWE can be more similar to MSE ($p$≈0) or to SSIM ($p$≈2).

The motivation for PWE is for the larger part the same as that of SSIM given at the end of the previous section. Again we focus on the second contribution. The constant $k_2^{p2}$ has the same role as $C_2$ in SSIM in adjusting the amount masking at low $\sigma$. With the introduction of the parameter $p$ as the exponent of $\sigma$, we add the possibility to adjust the amount of masking for medium and large $\sigma$ ($\sigma > k_2$).

## 3. PARAMETER TUNING

In this section we introduce an efficient method to tune the parameters $p$ in a distortion metric M($p$) for video coding using rate-distortion optimized encoding:

Suppose we want to compare two different metrics $\mathrm{M_A}$ and $\mathrm{M_B}$, where $\mathrm{M_A}$ and $\mathrm{M_B}$ are two forms of the same general metric M($p$), with different parameter settings $p_\mathrm{A}$ and $p_\mathrm{B}$. First, we perform optimized encoding to obtain sequences A and B. We have now generated 2 video sequences at the same bitrate, where the two distortion metrics $\mathrm{M_A}$ and $\mathrm{M_B}$ disagree on which has the best visual quality. According to metric A, sequence **A** has the better quality; according to metric B, B is to be preferred. By visual inspection, we can select which video actually has the best quality and thus which of the parameter settings $p_\mathrm{A}$ or $p_\mathrm{B}$ is to be preferred.

We briefly show what the benefits of this technique are compared to other techniques for parameter tuning. One alternative technique is to fit the parameters to a set of images or videos and the associated rating given by a group of observers, such as the test set used by the Video Quality Experts Group [6]. Unlike the proposed technique, this technique does not give insight in the influence of parameter variations on the details that are emphasized by the metric.

A second alternative technique is to use a qualitative physical model for the human visual system. Today, these models still contain some imperfections. Therefore, parameter tuning is still needed to get the best results.

The procedure of parameter tuning will be demonstrated for the parameter $p$, which is used for both $p_1$ and $p_2$ ($p_1$=$p_2$=$p$). For the parameters $k_1$ and $k_2$, we use the values based on the relation with SSIM as given in equation 13. Rate-distortion optimization is done for the quantizer step size $q$.

Results are given for MPEG-2 encoding of the "tennis" sequence. A part of an uncompressed frame is given in figure 1. Rate-distortion optimization for the quantizer step size $q$ has been implemented in the MSSG TM-5 encoder [7]. With this encoder, two sequences A and B have been generated as described above, with $p_A$=0 (MSE setting of

PWE) and $p_B$=2 (SSIM setting of PWE). Comparing sequences A and B in figures 2a and 2b, respectively, we find that with $p$=0, many bits are spent in the text area. Fewer bits are spent on the wall area, resulting in texture flattening. Conversely, for $p$=2, we find that with this parameter setting, less bits are spent in this area, resulting in lower image quality. Instead, more bits are spent the more subtle texture are on the brown wall area. This can be understood since for $p$=2, PWE displays a strong masking of errors in areas with large signal variations (large σ) such as the text area with its black-white transitions.

We now can tune the parameter $p$ by generating more sequences at different values. It was found that a value of $p$=1.5 provides a good trade-off between larger and smaller details as compared to both $p$=0 and $p$=2 for this set of sequences (see figure 1c). This example shows demonstrates how the proposed technique is used to quickly tune metric parameters for a specific type of content.

## 4. PARAMETER DEPENDENT CODING GAIN

In this section we consider the dependence of coding gain of RDopt encoding of the quantizer step size on the PWE parameter $p=p_1=p_2$. The same MPEG-2 encoder is used as described in the previous section. The reported results are average results of a set of 6 representative test sequences ("Cheer", "Suzie", "Tennis", "Shields", "Cityscape" and "Airshow") at a resolution 720x480 with progressive scan 25 frames per seconds.

The bitrate reduction is measured with respect to reference techniques. As a reference constant quantizer (constant Q) encoding and TM-5 adaptive quantization [7] are used. We have encoded 13 frames with a GOP length of 12 frames and GOP structure I B B P B B P… . The default TM-5 quantization matrices for intra and inter coding were used.

The bitrate reduction is determined as follows. First, encoding is done with a reference technique, resulting in distortion $D=D_0$, as measured distortion metric M, and reference rate $R_{\text{ref}}(D_0)$. Then, RDopt encoding for M is done such that we achieve the same distortion $D_0$, resulting in rate $R_{\text{RDopt}}(D=D_0)$. We express the compression efficiency gain as the bitrate reduction as a fraction of the reference bitrate:

$$G = \frac{R_{\text{ref}}(D = D_0) - R_{\text{RDopt}}(D = D_0)}{R_{\text{ref}}(D = D_0)}. \qquad (11)$$

The average bitrate reduction for bitrates ranging from 1.5 to 8 Mbit/s has been calculated using the Bjontegaard method [8]. The results for different values of $p$, averaged over the test set, are given in table 1 below. The value $p$=1.5 is a result of parameter tuning as discussed in section 3. Note that this experiment is different from the results given in figure 1 in that we consider the bitrate reduction at a constant distortion, whereas in figure 1, encoding results at a constant bitrate are compared.

The gain $G$ depends on the distortion metric used as a result of the requirement that the bitrate reduction is measured for


*original uncompressed frame*


*a) encoded PWE[p=0] optimal*


*b) encoded PWE[p=2] optimal*


*c) encoded PWE[p=1.5] optimal*

*Figure 1: Parameter tuning for the parameter* p *in the perceptually weighted error (PWE). The images a-c are encoded versions at the same bitrate of the original video (top) using rate-distortion optimal encoding for* p*=0 (MSE setting of PWE) in (a);* p*=2 (SSIM setting) in (b) and* p*=1.5 (tuned* p*) in (c). This experiment shows which details in a frame are emphasized, depending on distortion metric parameters.*

a constant value of M. Also, the metric M is used in the RDopt encoding.

Table 1: Compression efficiency gain of optimized encoding for PWE with parameter setting $p$ with respect to constant quantization encoding (constant Q) and TM-5 adaptive quantization (adaptive Q). Results given as a percentage of the reference bitrate averaged over a range of bitrates (see figure 1). Also given is the standard deviation SD of the bitrate reduction across the test sequences.

| $G$ (%±SD) | Reference | |
|---|---|---|
| $p$ | constant Q | adaptive Q |
| 0 | 1±2 | 20±7 |
| 1.5 | 8±4 | 17±3 |
| 2 | 12±6 | 14±4 |

The results show that RDopt encoding for the quantizer $q$ yields little gain for $p$=0 (MSE setting) compared to the best reference: constant Q encoding. This result explains why RDopt for the quantizer step size is not used in encoders looking to minimize MSE (or maximize PSNR).

Conversely, the rate-distortion optimal quantizer yields significant gains for $p$=1.5 and $p$=2. For constant Q, this gain is explained from the fact that constant Q distributes the error equally over the video, not taking into account masking effects. For adaptive Q, masking effects are taken into account. However, here the trade-off with bitrate is not made as in the case of RDopt encoding.

In figure 2, bitrate reductions for $p$=2 with respect to constant Q are given as a function of bitrate. Note that the gain diminishes at low bitrates (< 1 Mbit/s). Amongst others, this is a result of the addition bit cost needed for signaling a change in quantization level.

## 5. CONCLUDING REMARKS

Experiments show that constant quantizer encoding is close to optimal for MSE. However, considerable gains can be achieved for perceptual metrics such as the tuned Perceptually Weighted Error (PWE) and Structural Similarity (SSIM).

Rate-distortion optimal encoding of the quantizer level provides an efficient way to tune parameters in a visual quality metric. It directly shows which type of details is emphasized as a result of a given parameter setting. For the tunable PWE, we find that a setting balancing MSE and SSIM behavior is optimal for coding purposes.

## REFERENCES


[1] Lubin, J. e. Sarnoff JND Vision Model, contribution to IEEE standards subcommittee. 1997. Princeton, USA.

[2] Zhou, W., Ligang, L., and Bovik, A. C. Video quality assessment based on structural distortion measurement. Signal.Processing.: Image Communication. 19[2], 121-132. 2004.

[3] Sullivan, G. J. and Wiegand, T. Rate-distortion optimization for video compression. IEEE signal processing magazine 15[6], 74-90. 1-11-1998.

[4] Zhi-Yi Mai Chun-Ling Yang Kai-Zhi Kuang Lai-Man Po. A Novel Motion Estimation Method Based on Structural Similarity for H.264 Inter Prediction. 2006. Proceedings of IEEE ICASSP conference.

[5] Zhi-Yi, M., Chun-Ling, Y., Lai-Man, P., Sheng-Li, X., Zhi-Yi, M., Chun-Ling, Y., and Sheng-Li, X. Improved best prediction mode(s) selection methods based on structural similarity in H.264 I-frame encoder. Adv.Concepts for Int.Vision Syst.7th.International.Conference., ACIVS.2005.Proceedings. 435-441. 2005.

[6] Video Quality Experts Group (VQEG). Final report from the video quality experts group on the validation of objective models of video quality assessment. 1-6-2000.

[7] MPEG-2 Test Model (TM) 5 software documentation, http://www.mpeg.org/MSSG/tm5/index.html. 1993.

[8] Bjontegaard, G. Calculations of Average PSNR Differences between RD curves. VCEG-M33. 1-4-2001. ITU-T SC16/Q6, 13th VCEG Meeting, Austin, USA.


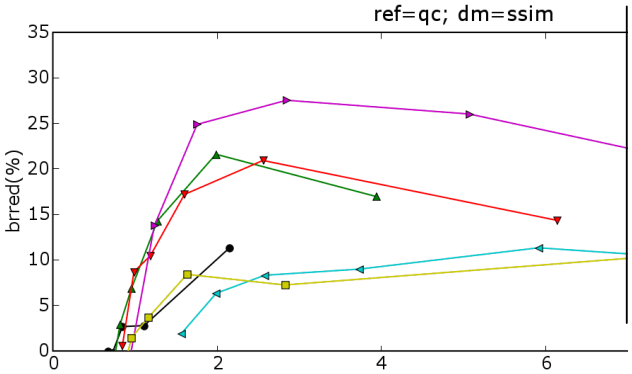


*Figure 2: Bitrate reduction for rate-distortion optimal encoding for PWE[p=2] (SSIM setting) as a function of the bitrate in Mbit/s. Bitrate reduction are calculated compared to constant quantizer encoding. Results are given for all 6 test sequences.*