A Rate-Distortion Metric Targeting Perceptual Video Coding

Bruno George de Moraes, Ismael Seidel, José Luís A. Güntzel
{brunogm,ismaelseidel,guntzel}@inf.ufsc.br

Embedded Computing Lab. (LCE)
Federal University of Santa Catarina, Florianópolis, Brazil

Abstract
This paper presents a new metric, called Quantization Distortion Energy (QDE), based on perceptual video coding, proposed for block matching considering the losses in quantization, and its evaluation in an H.264/AVC encoder. Three tests were elaborated to compare the metric using rate–distortion optimization (RDO) with Sum of absolute transformed differences (SATD) metric. The first test shows the curve of structural dissimilarity (DSSIM) and peak signal to noise ratio (PSNR) at various bit-rates. The second test shows the size reduction with the quality normalized by SSIM, resulting in a variable constant ratefactor (CRF) for the proposed encoder and a fixed CRF for the original encoder. In the third test an algorithm was used to perform the second test automatically spanning CRF range. The results show that the proposed metric, at the same quality in SSIM, provides gains in bit-rate ranging from 2% to 60%. Furthermore, in most cases considering the same bit-rate, gains were obtained in both quality metrics: SSIM and PSNR.

1. Introduction
The H.264/AVC standard provides at least 50% of bit rate improvement over MPEG2, which has guaranteed its widespread adoption for transmission and storage of high definition video in applications such as some digital television standards (e.g., SBTVD, DVB-S2), recent movie formats (e.g., Blu-ray and AVCHD), etc. Despite being the state-of-art for its efficiency, the bandwidth and costs associated with, per example, a 1080p IPTV (Internet Protocol television) application has limits in user-base by definition of broadband penetration on the market. Therefore, improving the analysis step of an H.264/AVC encoder contributes to expand its use to other applications.

Motion pictures or videos have lots of data. To compress them, each frame is divided into macroblocks and further into smaller blocks that are submitted to the Inter and Intra Frame Prediction techniques. Those techniques explore temporal and spatial redundancy (i.e., nearby video blocks exhibiting high correlation) by coding only the differences between similar blocks. A search algorithm will look for candidate blocks, and the best candidate block is selected in a simple view with a specific distortion metric [1]. Thus, by exploring the encoder metrics used to find matching blocks with minimum residues and distortion, it is possible to achieve better compression rates, while keeping the video quality or, conversely, increasing quality for the same bit-rate.

Distortion metrics are divided into objective and subjective. Objective metrics are purely mathematical models derived from signal theory and do not consider human visual system (HVS) characteristics. However, for performance reasons, they are the most used ones. Subjective metrics exploit the HVS models. The few existing ones suffer from poor performance and explore only a small set of features. For that reason, they are neither widely used nor well accepted [2]. Examples of the common used metrics are the SAD (Sum of Absolute Differences) [3], the SSD (Sum of Squared Differences) [4] and the SATD (Sum of Absolute Transformed Differences) [3]. This paper presents, as contribution, the development and evaluation of a subjective metric, named Quantization Distortion Energy (QDE) which is explained in Section 2.

The rest of this paper is organized as follows. Section 3 shows the evaluation process for the metric, whereas Section 4 comments the evaluation results. Conclusions and perspective work are outlined in Section 5.

2. Proposed Metric
Most of the metrics used for matching block does not directly consider the quantization step, which occurs after the discrete cosine transform (DCT) [5], for the current block. Typically, only the reconstructed data is considered containing losses in quantization, inverse quantization and inverse transform. The DCT coefficients represent frequency bands, for which it is known that the HVS has different sensitivity, as defined by the contrast sensitivity function (CSF) [6]. The quantization eliminates some of those frequency bands in order to achieve high compression rates. However, as drawback, the quantization causes a great amount of image distortion, either by rounding errors or by frequency band elimination [4].

By considering the errors caused by the quantization step, it is possible to predict what block match candidate will cause less distortion after the inverses of quantization and transform steps, for frame re-assembly. The worst errors occur when some transformed coefficients are eliminated. Based on that fact, we propose to apply a greater weight for the coefficients that will cause the most distortions. This is done by
predicting the eliminated coefficients before the quantization. Thus, for all the 52 quantization parameters (QP), the thresholds of DCT coefficients that will be eliminated shall be calculated, for the multiplicative factors (MF) table of H.264/AVC [3].

$$0 = \text{round}(\text{Coeff} \times \text{MF}[\text{QP}\%6,n]/2^{\text{floor}(\text{QP}/6)}) \times \frac{1}{2^{15}} \quad [3] \quad (1)$$

In order to calculate the metric some threshold values are obtained from the quantization equation (1) which was solved for “Coeff”, obtaining the root value for all MF, as mentioned before. The thresholds were calculated for 4x4 and 8x8 DCT of H.264/AVC. Only the QP 51 thresholds were stored, whose values can be seen in Tab.1, for 4x4 blocks. The others thresholds can be obtained by doing a simple shift operation by \(8 - \text{floor}(\text{QP}/6)\]. By comparison of the DCT coefficients with the thresholds values, the coefficients can be classified in two groups, eliminated and encoded. The encoded will be handled as in SATD, by gathering absolute sums. The eliminated coefficients will be quadratically summed as in SSD. This ensures that the encoded group will be representative for the bit-rate and the eliminated group will estimate the distortion caused by quantization. Such proposed metric will be called Quantization Distortion Energy (QDE).

| Tab. 1 - The QP 51 thresholds coefficients for 4x4 block |
|-----------------|-------|-------|
| 640             | 1039  | 1599  |
| 703             | 1119  | 1800  |
| 832             | 1279  | 2000  |
| 896             | 1440  | 2300  |
| 1023            | 1599  | 2500  |
| 1151            | 1840  | 2899  |

3. Testing QDE

As test platform for QDE the H.264/AVC encoder of the x264 project [7] was used. In the x264 there are three modes of rate control: Constant Quantizer, Bit-rate and Constant Ratefactor, which is similar to constant quantizer, but acts on reducing/improving quality according to frame importance, maximizing perceptual quality of the encoded video. To compare to the original version of the x264, which uses SATD, appropriate modifications in the encoder were accomplished, in order to use QDE thresholds to perform coefficient classification. Residues were summed in absolute and squared parts respectively, constituting the modified x264 version. It is important to notice that in x264, after quantization there is one step called decimation which eliminates some coefficients. This step is present in both encoders, and hence does not influence the comparison itself. Those two versions of x264 encoder were compared against each other, to show the possible gains or losses of the new metric. The three performed tests are referred to as “A”, “B” and “C”.

In the experiment “A”, the encoder is parameterized to compress the video by restricting bit-rates at the values of 400, 700, 1000, 2000, 4000 and 8000 Kbit/s, using the multi pass controller, which guarantees the specified rates. A script iterated all bit-rate values for both encoders, and then two charts were generated from the statistics file, one chart showing the curves for the bit-rates against PSNR [4], and another one with bit-rates against DSSIM [4], which is a distance metric derived from SSIM [9].

The SSIM measures the structural similarity between two images: an original (distortion-free) to which the distorted image will be compared. This kind of approach is known as full-reference (FR). The other two approaches are no-reference (NR), in which the reference image is not available, and reduced-reference (RR), in which the reference image is only partially available. The most traditional image/video quality evaluation metric is the PSNR. However, the PSNR values do not correspond to the perceived visual quality due to the non-linear behavior of the HVS. In principle, the SSIM assumes that the HVS perception is highly adapted to extract structural information from a scene. Hence, SSIM is one of the most effective perceptual quality evaluation metrics and that justifies its adoption in this work to measure quality.

For experiment “B”, the quality measured in SSIM was restricted by encoding the video with CRF [7] equals to 20.0 (23.0 by default at x264 encoder, used 20.0 to have high quality samples). For each test sample, the same quality was sought by doing manual changes in CRF on the modified version of the encoder until the quality was reached. By doing so, it was possible to measure the differences in video size for the two encoders.

Experiment “C” can be considered as the most important one because it allows the comparison of gains (in bit-rate) for the same SSIM on both encoded videos. As long as in x264 there was not a way to choose final SSIM, an algorithm was created to iterate over several test samples spanning CRF from 10 to 30, looking for convergence between the original SSIM and the modified one by adjusting its CRF.
4. Results

The following results present the QDE performance in quality against the SATD as defined in the RDO of the x264, both metrics act as D metrics (part of RDO) in such case. Charts showed in Fig. 1 indicates, for the sequences "Riverbed" and "Sunflower, the behavior of QDE and SATD in quality. DSSIM (upside charts) and PSNR (downside charts) versus the bit-rate of the final encoded sample, as defined in experiment "A". The charts of "Riverbed" show that, for all samples, QDE had better DSSIM and PSNR over SATD. "Sunflower" is a video sample that achieved gains up to 60% in compression at experiment "C" (see Fig. 3), for the same quality in original encoder. This sample in lower bit-rates has losses in PSNR and DSSIM. However, as its bit-rate increases, the QDE shows better results than SATD. Thus, experiment “A” shows an apparent tendency of QDE superiority over SATD, as the bit-rate increases. On the other hand, experiment “C” shows that there are cases, for the same quality, where higher bit-rates result in a worse compression with QDE.

![Fig. 1 - Graphics generated for PSNR and DSSIM for sequences “Sunflower” and “Riverbed”, as part of the experiment “A”](image)

![Fig. 2 - Experiment “B”: results at size of QDE relative to the original](image)

The simplest experiment, “B” (Fig. 2), is a short overview of the metric at CRF 20 for all nine tested video samples. In Fig. 3 it is possible to see the results for experiment “C”. In this figure each marked point at
the curves represents, from right to left, a CRF value of the original coded video, ranging from 10 to 30. The horizontal axis informs the quality in SSIM achieved for both encoders at different CRFs. The vertical axis shows the gain in size (kb) relative to the original encoded sample. In the case of “Ducks take off” there is a steady improvement at CRF range from 10 to 19 of 0.5% to 2.5%. Above that, from CRF 20 to 27, there are higher gains until a peak of ~5%. From CRF 28 to 30 diminishing gains occur due to the limits imposed by lower qualities. Yet for the experiment “C”, the results for “Sunflower” samples show a minor improvement range until CRF 15 (in practice, lossless). Above that, a substantial gain, up to a peak of ~60% in CRF 20. Above such CRF the gains diminish down to some penalty at CRF 28 (with a bottom of ~ -5%). These effects in CRFs greater than 28 are felt even at CRF 51 in which QDE presents better SSIM quality than attainable, because there is no higher CRF. Thus, the problem relies on achieving higher quality for the entire quantizer range allowed by H.264/AVC. For CRF after ~28, the encoder tries to match the available space in the encoding buffer, for which a low bit-rate is needed.

![Graph showing the curves for different CRFs](image.png)

Fig. 3 - Experiment “C” shows the result of DSSIM vs. size gain over the original using the proposed metric for samples “Ducks Take Off” and “Sunflower”, respectively

5. Conclusion and perspective work

The evaluation of the Quantization Distortion Energy (QDE) metric, based on the prediction of eliminated coefficients caused by quantization, showed that there is much information available during the encoding process which could be used for better block-matching, and this metric successfully utilizes some of this information. The generated graphics show gains in videos coded with QDE. This efficacy has been presented with the three conducted experiments for most of the tested video samples. Regarding experiment “B”, only one among the nine tested videos had a loss in bit-rate, for the same quality. Experiment “A” shows that, switching the bit-rate, the quality can be increased not only by increasing the bit-rate itself, but also by using new block match metrics, such as QDE. Experiment “C” shows that there are certain CRF ranges where, at the same quality, compression rates are better for QDE than for SATD.

Ongoing work includes the implementation and comparison of the QDE within the JM reference software of H.264/AVC [10] to also investigate how this proposed metric behaves. In addition, a hardware implementation of QDE is currently under development.

6. References