ANALYSIS OF IN-LOOP DENOISING IN LOSSY TRANSFORM CODING

Eugen Wige[1], *Gilbert Yammine*[1], *Peter Amon*[2], *Andreas Hutter*[2], *André Kaup*[1]

[1] Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg
Erlangen, Germany

[2] Siemens Corporate Technology, Information and Automation Technologies
Munich, Germany

ABSTRACT

When compressing noisy image sequences, the compression efficiency is limited by the noise amount within these image sequences as the noise part cannot be predicted. In this paper, we investigate the influence of noise within the reference frame on lossy video coding of noisy image sequences. We would like to estimate how much noise is left within a lossy coded reference frame. Therefore we analyze the transform and quantization step inside a hybrid video coder, specifically H.264/AVC. The noise power after transform, quantization, and inverse transform is calculated analytically. We use knowledge of the noise power within the reference frame in order to improve the inter frame prediction. For noise filtering of the reference frame, we implemented a simple denoising algorithm inside the H.264/AVC reference software JM15.1. We show that the bitrate can be decreased by up to 8.1% compared to the H.264/AVC standard for high resolution noisy image sequences.

Index Terms— High quality compression, predictor denoising, quantization effects

1. INTRODUCTION

Video and image signals which are generated through a physical acquisition process are known to be degraded by noise. The noise characteristics as well as the noise strength within these signals depend strongly on the acquisition process itself. For example, in digital photography the signal-to-noise ratio depends on the amount of incoming photons on the sensor. If the sensor gets larger for the same photo resolution, more photons can be collected in order to generate a pixel in the photo. In other words, the noise amount gets larger if the resolution increases using the same sensor size. Similar problems appear in digital video capturing.

Sometimes, such video signals have to be compressed with a very high quality, i.e., in professional video applications like monitoring systems for buildings and industrial manufacturing, and medical imaging. However, the state of the art video compression standard H.264/AVC, which exploits the temporal correlation between adjacent frames, is optimized for consumer quality. It has been shown that removing noise from the video before encoding leads to reasonable gains in compression efficiency for noisy image sequences [1] [2] [3]. The problem of using such schemes is that although the subjective quality is getting higher, the noise reduction algorithm could also damage the useful signal part significantly and thus it is not allowed in some professional video applications.

In [4] we showed that the coding efficiency of lossless compression of noisy video signals can be improved by a denoising operation applied on the reference frame, which improves the temporal prediction. The same idea can be extended to lossy video coding without influencing the quality adjustment of a hybrid video coder. Fig. 1 illustrates a video encoder with an in-loop denoising filter denoted by WF.

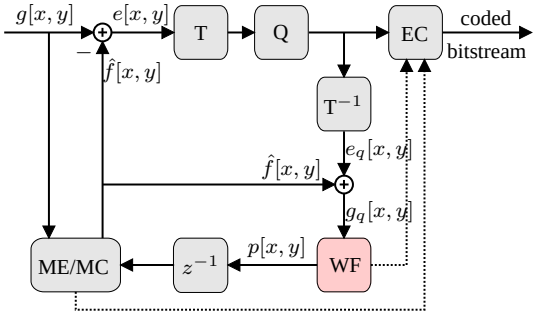

Fig. 1. Diagram of the lossy encoder.

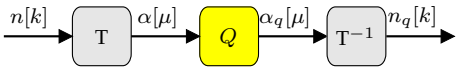

Fig. 2. Diagram of the transform, quantization, and inverse transform process within a lossy transform coder.

The question that arises for lossy coding is: How much noise is left within the reference frame after transform, quantization and inverse transform? In Section 2 we analyze the noise after transform coding with a uniform quantization and apply it to the H.264/AVC quantization scheme afterwards. In Section 3 we introduce the MMSE estimator used for denoising of the reference frame and give some simulation results. Section 4 concludes the paper.

2. QUANTIZATION OF NOISE IN TRANSFORM CODING

If considering the prediction to be perfect, only the noise signal $n[x,y]$ is present in the error signal $e[x,y]$ in Fig. 1. For simplification, we analyze the transform coding of 1-D noise signals $n[k]$ and we assume the noise $n[k]$ to be white and Gaussian distributed with zero mean and standard deviation σ_n. Fig. 2 illustrates the information reduction step inside a transform coder.

The spatial signal $n[k]$ is transformed into the frequency domain, where the frequency coefficients $\alpha[\mu]$ are quantized. The quantized coefficients $\alpha_q[\mu]$ are sent to the decoder where they are inverse transformed to get the lossy spatial signal $n_q[k]$. In order to prevent drift between the encoder and the decoder, the spatial signal $n_q[k]$ is also reconstructed inside the encoder and this signal is used

for prediction.

2.1. Uniform Quantization in Frequency Domain

We consider unitary transforms like the discrete cosine transform (DCT) for the forward transform and the inverse discrete cosine transform (IDCT) for the inverse transform operation. According to Parseval's theorem, the variance of the noise signal $n[k]$ can be calculated in the spatial as well as the frequency domain:

$$\sigma_n^2 = \frac{1}{N}\sum_{k=0}^{N-1} n[k]^2 = \frac{1}{N}\sum_{\mu=0}^{N} \alpha[\mu]^2, \tag{1}$$

where N is the transform size. If the probability density function (PDF) of each DCT coefficient $\alpha[\mu]$ is known (which is the same for all frequency components in the case of white noise $n[k]$), it can also be used for calculating the variance:

$$\sigma_n^2 = \int_{-\infty}^{\infty} p_A(\alpha)\alpha^2 d\alpha = 2 \cdot \int_{0}^{\infty} p_A(\alpha)\alpha^2 d\alpha, \tag{2}$$

where $p_A(\alpha)$ is the PDF of the DCT coefficients. As the DCT is a unitary transform and the input signal is Gaussian distributed with standard deviation σ_n, the PDF of the DCT coefficients $\alpha[\mu]$ is given as

$$p_A(\alpha) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{\alpha^2}{2\sigma_n^2}}. \tag{3}$$

Even if the signal $n[k]$ is not Gaussian distributed (i.e., uniformly distributed), the PDF of the DCT coefficients tends to a Gaussian distribution due to the central limit theorem. Thus the derivation is also appropriate for transform, quantization, and inverse transform of non Gaussian distributed signals if the transform size is large enough. In order to calculate the variance of the quantized signal $n_q[k]$ we can also use Parseval's theorem and calculate the variance from the quantized DCT coefficients $\alpha_q[\mu]$:

$$\sigma_{n_q}^2 = \int_{-\infty}^{\infty} p_{A_q}(\alpha_q)\alpha_q^2 d\alpha_q, \tag{4}$$

where $p_{A_q}(\alpha_q)$ is the PDF of the quantized DCT coefficients $\alpha_q[\mu]$. The uniform quantization of an input signal can be considered as a sampling of its PDF [5]. Thus the PDF of the quantized DCT coefficients $\alpha_q[\mu]$ is given as

$$p_{A_q}(\alpha_q) = \sum_{\lambda=-\infty}^{\infty} w(\lambda)\delta(\alpha - \lambda\Delta), \quad \lambda \in \mathbb{Z}, \tag{5}$$

where δ is the unit impulse function, $w(\lambda) = \int_{\lambda\Delta-\frac{\Delta}{2}}^{\lambda\Delta+\frac{\Delta}{2}} f_A(\alpha)d\alpha$ are the weighting factors of the unit impulse functions, and Δ is the quantization interval. Sampling of the Gaussian distributed PDF is illustrated in Fig. 3.

The figure shows the input PDF $p_A(\alpha)$ of the DCT coefficients $\alpha[\mu]$ and the PDF $p_{A_q}(\alpha_q)$ of the quantized DCT coefficients $\alpha_q[\mu]$. From the figure it is clear that the DCT coefficients which are rounded up contribute to an increase of the noise variance $\sigma_{n_q}^2$ and the DCT coefficients which are rounded down contribute to a decrease of the noise variance $\sigma_{n_q}^2$.

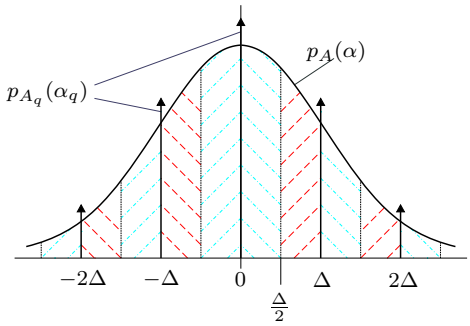

Fig. 3. Uniform sampling of the Gaussian PDF with sampling interval Δ. According to the absolute value, the dashed lined area illustrates the DCT coefficients to be rounded up and the dot-dashed lined area illustrates the DCT coefficients to be rounded down.

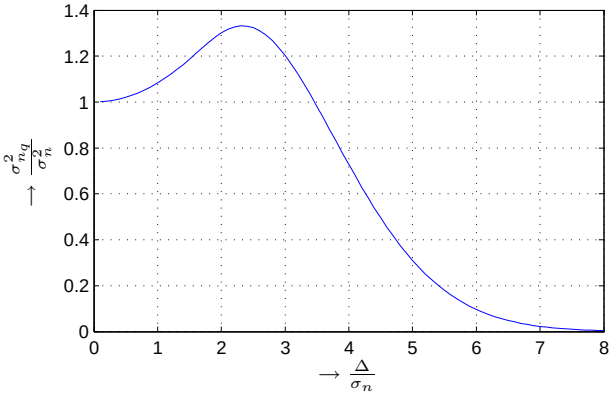

Fig. 4. Relationship between quantization interval Δ and output variance $\sigma_{n_q}^2$ normalized to input noise.

From 4 and 5 and because of the symmetry of the PDF the variance of the quantized noise is given as

$$\begin{aligned}\sigma_{n_q}^2(\Delta) &= \int_{-\infty}^{\infty}\sum_{\lambda=-\infty}^{\infty} w(\lambda)\delta(\alpha-\lambda\Delta)\alpha_q^2 d\alpha_q \\ &= \Delta^2\sum_{\lambda=-\infty}^{\infty}\lambda^2 w(\lambda) = 2\cdot\Delta^2\sum_{\lambda=1}^{\infty}\lambda^2 w(\lambda).\end{aligned} \tag{6}$$

For a Gaussian PDF, the weights $w(\lambda)$ in (6) can be calculated as

$$w(\lambda) = \mathrm{erf}\left(\frac{\lambda\Delta+\frac{\Delta}{2}}{\sigma_n\sqrt{2}}\right) - \mathrm{erf}\left(\frac{\lambda\Delta-\frac{\Delta}{2}}{\sigma_n\sqrt{2}}\right), \tag{7}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{u^2}du$ is the Gauss error function. As it can be seen from (6) and (7), the variance $\sigma_{n_q}^2$ of the quantized noise depends on the input variance σ_n^2 and on the quantization interval Δ. These equations and thus the behavior of the noise variance of the quantized signal is illustrated in Fig. 4.

The figure shows the output variance $\sigma_{n_q}^2$ of the quantized noise signal in dependency of the quantization interval Δ. Generally, it is clear that the signal power becomes less if the quantization interval gets bigger for a decaying PDF $p_A(\alpha)$. However, for uniform quantization it is interesting to notice that the noise becomes higher for

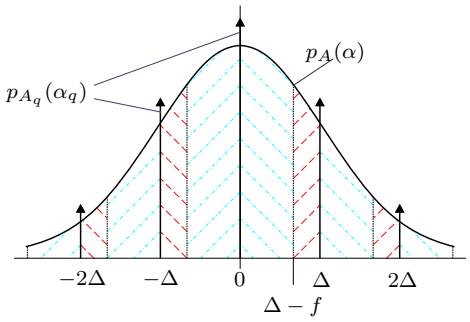

Fig. 5. Uniform sampling of the Gaussian PDF with $f \neq \Delta/2$ and sampling interval Δ.

small quantization intervals first before it becomes lower for larger quantization intervals. The variance of the quantized signal has its maximum when the quantization interval Δ equals $2.33 \cdot \sigma_n$.

2.2. Quantization in H.264/AVC

H.264/AVC uses a 4×4 transform in the baseline profile, which is an approximated version of the 4x4 DCT. Similarly, an 8×8 transform is used within the high profile. However in H.264/AVC a rounding control parameter for quantization of the transform coefficients has been introduced. The mapping of the transform coefficients $\alpha[\mu]$ to the quantized coefficients $\alpha_q[\mu]$ is described by the following equation:

$$\alpha_q[\mu] = \left\lfloor \left(\frac{|\alpha[\mu]| + f}{\Delta} \right) \right\rfloor \cdot \text{sign}(\alpha[\mu]) \cdot \Delta, \quad 0 < f < \Delta \quad (8)$$

where f is the rounding control parameter, $\lfloor \alpha \rfloor$ means rounding α towards minus infinity and sign(α) is the algebraic sign of α.

For example, if f equals to $\Delta/2$ the quantization corresponds to the uniform quantization discussed in the previous section. Usually two different rounding control parameters are used [6]:

- $f = \Delta/3$ for intra
- $f = \Delta/6$ for inter

Similarly to Fig. 3, Fig. 5 illustrates the sampling of the Gaussian PDF with $f \neq \Delta/2$.

It can be seen from Fig. 5 that the thresholds for rounding up and rounding down the transform coefficients within a quantization interval are shifted by the parameter f. For generation of this plot, f was chosen to be $\Delta/3$, which corresponds to the quantization of intra coded blocks. In this case more coefficients are rounded down and less coefficients are rounded up compared with the uniform quantization where $f = \Delta/2$ from the previous section. Therefore the curve for $f = \Delta/3$ should be below the curve for $f = \Delta/2$ in Fig. 4.

For calculation of $\sigma^2_{n_q}$, we can use (6) where the weighting coefficients $w_f(\lambda)$ additionally depend on the rounding control parameter f. It can be calculated for Gaussian noise as

$$w_f(\lambda) = \text{erf}\left(\frac{(\lambda+1)\Delta - f}{\sigma_n \sqrt{2}} \right) - \text{erf}\left(\frac{\lambda\Delta - f}{\sigma_n \sqrt{2}} \right). \quad (9)$$

The noise variance after transform, quantization, and inverse transform in H.264 is illustrated in Fig. 6.

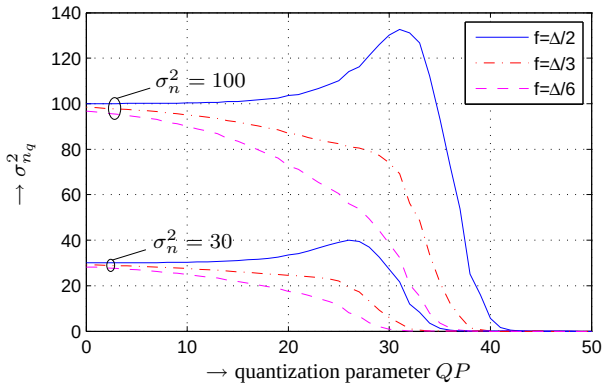

Fig. 6. Relationship between quantization parameter QP and output variance $\sigma^2_{n_q}$ for H.264/AVC.

The figure shows two different noise strength resulting from quantization of noise with the H.264/AVC quantization scheme. Additionaly, uniform quantization without deadzone is illustrated in the plot. It can be seen that the introduced rounding offset parameter f has noise reduction capabilities. However, the noise is still present for a wide range of quantization parameters. Thus it has to be considered during the inter prediction process where it should be removed from the reference frame as much as possible.

3. SIMULATION

In order to evaluate the efficiency of denoising of reference frames for lossy coding of noisy image sequences, we implemented a simple denoising algorithm into the H.264/AVC reference software JM15.1, i.e., block WF in Fig. 1. For coding tests we used high resolution sequences (*ParkJoy*, *CrowdRun*, *InToTree*, *OldTownCross*, *DucksTakeOff*) from [7], which we already used for lossless coding tests in [4]. The sequences have a spatial resolution of 3840×2160 pixels, a color bit depth of 8 bit per channel (original bit depth is 16 bit per channel) and a frame rate of 50 frames per second.

3.1. Denoising Algorithm

A modified version of the noise filtering algorithm from [8], the simplest form of a minimum mean-square error (MMSE) estimator, has been used for removing noise from the reference frame. Consider $f[x, y]$ to be the ideal noise-free reference frame. Thus the filtering process is described by the following equation:

$$p[x, y] = \mu_f[x, y] + \frac{\sigma_f^2[x, y]}{\sigma_f^2[x, y] + \xi\sigma^2_{n_q}} \cdot (g_q[x, y] - \mu_f[x, y]), \quad (10)$$

where x and y are the spatial coordinates, $p[x, y]$ is the processed image, $g_q[x, y]$ is the noisy reference image, and $\mu_f[x, y]$ and $\sigma_f^2[x, y]$ are the local mean and the local variance of the ideal noise-free reference frame. The noise variance $\sigma^2_{n_q}$ is assumed to be known (i.e., from the acquisition process). ξ can be sean as an adjusting parameter which optimizes the efficiency of the denoising according to the compression efficiency. As the noise is considered to be zero mean, $\mu_f[x, y]$ should be equal to $\mu_{g_q}[x, y]$ and the estimation of the local mean is reduced to:

$$\hat{\mu}_f[x, y] = \frac{1}{(2M+1)^2} \sum_{k=x-M}^{x+M} \sum_{l=y-M}^{y+M} g_q[k, l], \quad (11)$$

where M determines the window size in which the image signal is considered to be stationary. The estimation of the local variance of the noise-free image is described by the following equation:

$$\hat{\sigma}_f^2[x,y] = \begin{cases} \hat{\sigma}_{g_q}^2[x,y] - \sigma_n^2, & \text{if } \hat{\sigma}_{g_q}^2[x,y] > \sigma_{n_q}^2 \\ 0, & \text{else} \end{cases}, \quad (12)$$

where $\hat{\sigma}_{g_q}^2[x,y]$ is the local variance of the degraded image $g_q[x,y]$. The estimation of $\hat{\sigma}_{g_q}^2[x,y]$ is described by the following equation:

$$\hat{\sigma}_{g_q}^2[x,y] = \frac{1}{(2M+1)^2} \sum_{k=x-M}^{x+M} \sum_{l=y-M}^{y+M} (g_q[k,l] - \hat{\mu}_f[k,l])^2, \quad (13)$$

3.2. Simulation Results

Fifty frames of each of the high resolution sequences were coded. The first frame was coded as I-frame and the adjacent 49 frames were coded as P-frames, using a reference buffer size of 5 frames. For the noise filtering algorithm, we used a 3×3 window, i.e., $M = 1$. $\sigma_{n_q}^2$ was calculated from the estimated σ_n^2 and the quantization interval Δ (corresponding to a specific quantization parameter) as described in the previous section. The parameter $\xi = 3.5$ has been determined to be very good for this resolution. The estimated noise $\hat{\sigma}_n^2$ of the luminance channel of the unquantized sequences is shown in Table 1.

Table 1. Estimated noise $\hat{\sigma}_n^2$ within the high resolution sequences and maximum bitrate savings using in-loop-denoising in comparisson to the H.264/AVC standard.

Sequence	$\hat{\sigma}_n^2$	maximum bitrate savings
ParkJoy	3.1	3.3%
CrowdRun	6.5	7.7%
InToTree	10.5	7.7%
OldTownCross	15.9	8.1%
DucksTakeOff	21.6	5.1%

As shown in the section before, the noise is still present within the reference frame for low quantization parameters. The sequences were coded using the quantization parameters $QP \in \{10, 13, 16, 19, 22, 25, 28\}$. The rate-distortion curves of two sequences are illustrated in Fig. 7 below.

As can be seen in this figure, for the *ParkJoy* sequence the coding gain of the proposed scheme is low because of little noise within the image sequence. However, in the case of the *OldTownCross* sequence, the coding gain is much higher and is still present for $QP = 28$. This is due to the relatively high amount of noise compared to the *ParkJoy* sequence, which is still present for higher QPs, and less movement within the sequence which means that the P-modes are chosen very often. The maximum coding gains for the simulated quantization parameters are approximately 3.3%, 7.7%, 7.7%, 8.1%, and 5.1% for *ParkJoy*, *CrowdRun*, *InToTree*, *OldTownCross*, and *DucksTakeOff*. As there is a lot of movement in the *DucksTakeOff* sequence and thus P-modes are selected less often, the coding gain is not as high as in the case of the other three sequences.

4. CONCLUSION

In this paper, we have shown how much noise is present within a quantized signal of a DCT like transform coder. The compression efficiency of the transform video codec can be improved for noisy image sequences when the reference frame is noise-filtered. We implemented the MMSE estimator in order to remove noise from the reference frame. Doing this we achieved a maximum coding gain of 8.1% for one of the coded sequences. In the future we would like to work on further improving this algorithm for high quality video coding.

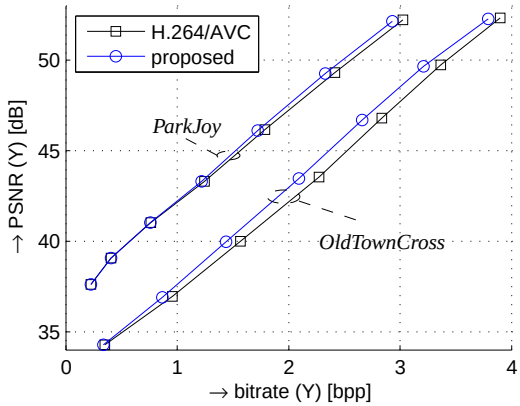

Fig. 7. Rate distortion curve for the *ParkJoy* and the *OldTownCross* sequence.

5. ACKNOWLEDGEMENT

This work was achieved with the help of the European Community's Seventh Framework Program through grant agreement ICT OPTIMIX n∘INFSO-ICT-214625.

6. REFERENCES

[1] H.-Y. Cheong, A.M. Tourapis, J. Llach, and J. Boyce, "Adaptive spatio-temporal filtering for video denoising," in *Proceedings of International Conference on Image Processing (ICIP 2004).*, Singapore, Oct. 2004, vol. 2, pp. 965–968.

[2] Byung Cheol Song and Kong Wook Chun, "Motion-compensated temporal prefiltering for noise reduction in a video encoder," in *Proceedings of International Conference on Image Processing (ICIP 2004).*, Singapore, Oct. 2004, vol. 2, pp. 1221–1224.

[3] Jim Deng, Alex Giladi, and Fernando Gómez Pancorbo, "Noise reduction prefiltering for video compression," *Stanford University*, 2008.

[4] E. Wige, P. Amon, A. Hutter, and A. Kaup, "In-loop denoising of reference frames for lossless coding of noisy image sequences," *accepted for International Conference on Image Processing (ICIP 2010).*, Hong Kong, Sept. 2010.

[5] B. Widrow, I. Kollar, and Ming-Chang Liu, "Statistical theory of quantization," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353–361, Apr. 1996.

[6] T. Wedi and S. Wittmann, "Quantization offsets for video coding," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2005).*, Kobe, Japan, May 2005, vol. 1, pp. 324–327.

[7] Sveriges Television, "The SVT high definition multi format test set," `ftp://vqeg.its.bldrdoc.gov/`.

[8] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transaction Pattern Analysis and Machine Intelligence (PAMI)*, vol. 2, no. 2, pp. 165–168, March 1980.