IN-LOOP DENOISING OF REFERENCE FRAMES FOR LOSSLESS CODING OF NOISY IMAGE SEQUENCES

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

[1] Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg
Erlangen, Germany
{wige, kaup}@lnt.de
[2] Siemens Corporate Technology, Real-time Vision & Industrial Imaging
Munich, Germany
{p.amon, andreas.hutter}@siemens.com

ABSTRACT

The major gain in video coding applications compared to single image coding is the use of temporal prediction, which exploits the correlation between adjacent frames. However, in high quality video coding, especially lossless video coding, the compression gain of P-frames over I-Frames becomes very small. The reason for that is that the reference frame for inter prediction is not good enough, and therefore the amount of inter-predicted blocks in a P-Frame becomes relatively small compared to the number of intra-predicted blocks. In order to generate a better predictor for inter-prediction, we propose to remove additive noise from the reference frame using an adaptive Wiener filter. This way, we could achieve a maximum compression gain of 4.6% and an average compression gain of 3.3% in contrast to the H.264/AVC standard for lossless coding of high quality image sequences without affecting the encoding time noticeably.

***Index Terms*—** Lossless Video Compression, Predictor Denoising, Motion Compensated Prediction

1. INTRODUCTION

Professional video aplications such as monitoring systems for buildings, industrial manufacturing, and medical applications need video signals with very high quality. These signals are likely to have very high resolutions in spatial as well as temporal direction, so the uncompressed data can become very large. For that, it is important to compress these signals as much as possible without visible information loss. For example, the DICOM standard which has been introduced for medical applications uses the JPEG-LS [1] and the JPEG-2000 [2] standards for high quality compression of medical images. In Digital Cinema, the JPEG-2000 standard is used for compression and archiving of digital video material. However, these standards do not exploit the temporal correlation in image sequences.

Other recent video compression systems such as VC-1 [3] and H.264/AVC [4] exploit the temporal correlation between images, but they are optimized for consumer quality applications. In [5], the authors compared the H.264/AVC standard with the JPEG-2000 standard when high resolution image sequences were coded with very high quality. They showed that at very high quality inter-frame coding of H.264/AVC becomes less efficient relative to intra-frame coding of H.264/AVC or JPEG-2000. It is expected that the reason for that is the (additive) noise within a video signal, which affects the motion compensated prediction inside a video encoder/decoder. In order to improve the coding efficiency, different algorithms (i.e., [6], [7], [8]) have been proposed, which should be applied to the noisy image sequences prior to encoding. In the latter it is shown that the noise prefiltering can be done inside the video encoder. Thus, the bitrate for coding a noisy image sequence is reduced while its subjective quality is improved. However, the prefiltering process introduces errors, which is not allowed in lossless compression of video signals.

Another approach for bitrate reduction is in-loop filtering of lossy encoded video data. In [9], [10] and [11] it is shown that in lossy coding a deblocking filter or quantization noise removal filter (the algorithms are developed for high quantization inside a video encoder) can improve the coding efficiency as well as the visual quality of the signal. Also in [12] the authors show that in-loop denoising filter for reduction of impulse noise leads to bitrate reduction and subjective quality improvement. Because the mentioned in-loop filters modify the visual output signal, they can not be used for lossless coding of image sequences.

In this paper, we want to show how the additive noise can be minimized, whereas keeping the coding process lossless. In contrast to the mentioned algorithms, we do not modify the visual output signal. Thus, it is possible to achieve a higher compression rate for lossless compression of video signals. We use an adaptive Wiener filter, which is implemented in the spatial domain for denoising of the reference frame only. This way we generate a better predictor for inter-frame coding by a low commplexity algorithm.

2. DENOISING FOR MOTION COMPENSATION

In this section, the motivation for the denoising of the reference frame will be reviewed and the denoising block in the encoder and the decoder will be introduced. Figure 1 shows a block diagram of a lossless video encoder. For simplification, the block diagram is observed at one specific time point, where only one inter frame $g[x, y]$ has to be coded and another motion compensated reference frame $\hat{f}[x, y]$ is available. Lossless coding implies the absense of quantization as it can be seen in this block diagram. The T-block stands for any further decorrelation of the error signal (i.e., lossless reversible transformation). ME/MC describes the motion estimation and motion compensation within a video encoder. The z^{-1}-block illustrates the reference picture buffer. The WF-Block is a denoising filter, which will be introduced in this section. The EC block describes the entropy encoder whithin a video encoder.

The subtraction of the motion compensated predicted frame

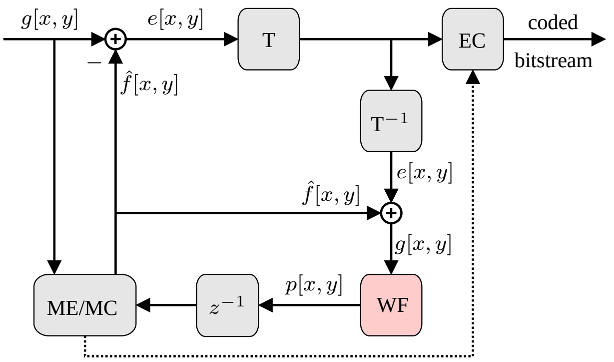

Fig. 1. Diagram of the lossless encoder.

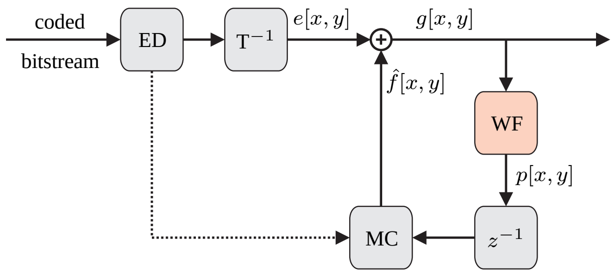

Fig. 2. Diagram of the lossless decoder.

from the actual frame in a video encoder is described by the following equation:

$$e[x,y] = g[x,y] - \hat{f}[x,y], \tag{1}$$

where x and y are the spatial coordinates, $g[x,y]$ is the signal which describes the actual frame which has to be compressed, $\hat{f}[x,y]$ is the predictor signal which describes the motion compensated reference frame and $e[x,y]$ is the error signal which has to be transmitted to the decoder after transformation and entropy encoding. Assuming that the image sequence is degraded by independent additive white noise, the signals can be written as:

$$g[x,y] = s_g[x,y] + n_g[x,y], \tag{2}$$
$$\hat{f}[x,y] = s_{\hat{f}}[x,y] + n_{\hat{f}}[x,y], \tag{3}$$

where s_ν is the useful part of the signal and n_ν is the noise part of the signal. From (1), (2) and (3) it follows that

$$e[x,y] = \underbrace{s_g[x,y] - s_{\hat{f}}[x,y]}_{s_e[x,y]} + \underbrace{n_g[x,y] - n_{\hat{f}}[x,y]}_{n_e[x,y]}, \tag{4}$$

From (4) it can be seen that the error signal consists of two major components. The signal component $s_e[x,y]$ can be minimized if the correlation between the actual noise-free frame and the noise-free reference frame increases (i.e., no motion occurs between these two images). However, the noise component $n_e[x,y]$ becomes higher because the noise of the actual frame is independent from the noise of the reference frame. In case of additive white Gaussian noise, the noise variance of the error image becomes twice as big as the noise variance of one of these two images.

In lossless coding applications the noise part of the actual picture has also to be encoded. In order to increase the compression efficiency, the noise in the reference frame should be minimized without introducing high degradations in the useful part of the signal. This would allow to minimize the noise amount in the error signal whereas keeping the coding process lossless.

The WF block, which is introduced in Fig. 1 is a denoising filter for additive white Gaussian noise, is located on the same place in the diagram as the usually used deblocking filter in lossy coding applications. Before saving the actual image to the reference buffer it should be processed by a denoising algorithm. The denoising algorithm should be adapted to the noise nature present in the video. In order to keep the coding process lossless, the same procedure should be applied in the decoder. In Fig. 2, the diagram of the lossless decoder is shown. In this figure, the ED block stands for an entropy decoder. Also in the decoder, the reference picture is first denoised before it is stored in the reference picture buffer. The major difference to a quantization noise removal filter or deblocking filter is that the image should be displayed before the noise filtering is applied, because otherwise the encoding/decoding process would not be lossless.

3. DESCRIPTION OF THE DENOISING ALGORITHM

In this section, the denoising algorithm within the denoising block in Fig. 1 and 2 is described. We suggest an adaptive Wiener filter algorithm to be used for reduction of the additive white noise in the reference frame. The filter minimizes the squared error between the noisy and the noise-free image sequence, which is the nature of a Wiener filter for Gaussian distributed processes. Furthermore, the filter prevents smoothing of contours by using of local statistics of the image signal, which is important for accurate motion compensation. Moreover, this filter algorithm does not need any transformation to a frequency domain, so it can be applied directly on the spatial signal.

In a local region where the signal and the additive noise are uncorrelated and are considered to be stationary, the Wiener filter is described by

$$H(e^{j\Omega_x}, e^{j\Omega_y}) = \frac{\Phi_{ff}(e^{j\Omega_x}, e^{j\Omega_y})}{\Phi_{ff}(e^{j\Omega_x}, e^{j\Omega_y}) + \Phi_{nn}(e^{j\Omega_x}, e^{j\Omega_y})}, \tag{5}$$

where $\Phi_{ff}(e^{j\Omega_x}, e^{j\Omega_y})$ and $\Phi_{nn}(e^{j\Omega_x}, e^{j\Omega_y})$ are the power spectral densities of the signal $f[x,y]$ and the noise $n[x,y]$ respectively. In the case of white noise the power spectral density is reduced to

$$\Phi_{nn}(e^{j\Omega_x}, e^{j\Omega_y}) = \sigma_n^2, \tag{6}$$

where σ_n^2 is the space invariant variance of the noise signal $n[x,y]$. The signal $f[x,y]$ can be modelled by a sum of a space variant local mean $\mu_f[x,y]$ and a space variant local variance $\sigma_f^2[x,y]$ (the reader is refered to [13] for more details). If the signal $f[x,y]$ is zero mean the power spectral density within the local region is reduced to

$$\Phi_{ff}(e^{j\Omega_x}, e^{j\Omega_y}) = \sigma_f^2[x,y], \tag{7}$$

where $\sigma_f^2[x,y]$ is the local variance of the image signal $f[x,y]$. Thus the Wiener filter within the local region is described by

$$H(e^{j\Omega_x}, e^{j\Omega_y}) = \frac{\sigma_f^2[x,y]}{\sigma_f^2[x,y] + \sigma_n^2}. \tag{8}$$

Generally, the signal $f[x,y]$ is not zero mean and thus the mean $\mu_f[x,y]$ has to be subtracted from $f[x,y]$ before filtering and has to be added again after filtering. The filtering process is described by the following equation:

$$p[x,y] = \mu_f[x,y] + \frac{\sigma_f^2[x,y]}{\sigma_f^2[x,y]+\sigma_n^2} \cdot (g[x,y]-\mu_f[x,y])\,. \quad (9)$$

The noise variance σ_n^2 is assumed to be known (i.e., from the acquisition process).

As the noise is considered to be zero mean, $\mu_f[x,y]$ should be equal to $\mu_g[x,y]$ and the estimation of the local mean is reduced to:

$$\hat{\mu}_f[x,y] = \frac{1}{(2M+1)^2} \sum_{k=x-M}^{x+M} \sum_{l=y-M}^{y+M} g[k,l], \quad (10)$$

where M describes 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:

$$\hat{\sigma}_f^2[x,y] = \begin{cases} \hat{\sigma}_g^2[x,y] - \sigma_n^2, & \text{if } \hat{\sigma}_g^2[x,y] > \sigma_n^2 \\ 0, & \text{else} \end{cases}, \quad (11)$$

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

$$\hat{\sigma}_g^2[x,y] = \frac{1}{(2M+1)^2} \sum_{k=x-M}^{x+M} \sum_{l=y-M}^{y+M} (g[k,l] - \hat{\mu}_f[x,y])^2, \quad (12)$$

From these equations it is clear that noise reduction is obtained by averaging the pixel values in a rectangular window in dependancy on the local signal and noise variances. Choosing a larger window by variation of M leads to higher noise reduction, but it also introduces more blurring of the useful signal part. Therefore, M can be considered as an optimization parameter for the lossless compression. Also the variation of σ_n^2 can be considered as an optimization parameter. Increasing σ_n^2 leads to higher noise reduction whereas the useful signal part is more blurred.

The local filters in (9) - (12) have a low computational complexity and therefore do not affect the encoding time noticeably.

4. SIMULATION RESULTS

Different simulations have been made in order to evaluate the efficiency of the introduced denoising block. The proposed algorithm in Section 3 was implemented in the H.264/AVC reference software JM15.1 [14]. All of the presented test results above were generated with the 4:4:4 predictive profile, which supports lossless encoding, and the standard parameter sets for lossless encoding. In this section, two test results will be presented.

In order to investigate the influence of the noise on the compression efficiency, we degraded computer generated image (CGI) sequences from [15] with different amount of additive white Gaussian noise, and coded them afterwards. Here, we present the simulation results for the *Big Bucks Bunny* sequence beginning with the frame 2000. This sequence has a resolution of 1920x1080 pixels, a color bit depth of 8 bits per channel, and a framerate of 24 frames per second. 100 frames of the sequence were coded. The first frame was coded as I-frame and the adjacent 99 frames were coded as P-frames. For coding of the P-frames, 5 reference frames were used.

First, the sequence was coded with the reference software JM15.1. Afterwards, the same sequence was coded with the proposed coding scheme in Section 2. For denoising of the reference frame we have used the algorithm which is described in Section 3. A uniform window of 3x3 pixels was chosen for the calculation of the local statistics ($M = 1$). The knowledge of σ_n^2 of the noise within the degraded image was used for denoising of the reference frame. Because in this case the noise-free image sequence is available too, it is possible to show a theoretical bound according to the argumentation in Section 2. Therefore, the noise-free image sequence was used for motion compensated prediction of the noisy image sequence. This way, it is possible to minimize the noise in the error signal. The coded sequence has very low motion activity, so the bitrate for P-frames should be much lower than the bitrate for I-frames. In the case of the noise-free image sequence, the bitrate for the I-frame is 2.12 bpp (Y-channel) and the average bitrate for the P-frames is 1.11 bpp (Y-channel). The coding results for this sequence degraded by additive white Gaussian noise can be observed in Figure 3. The figure shows the bitrate in dependency of the noise

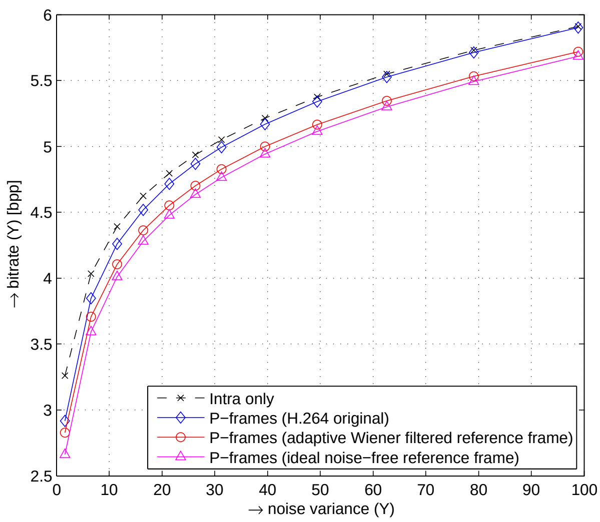

Fig. 3. Influence of the additive white Gaussian noise within an image sequence on the compression efficiency.

variance within the image sequence that has to be coded. The lower a curve is located in this figure, the higher is the compression ratio. The bitrates of H.264 standard, H.264 with adaptive Wiener filtering of the reference frame, and H.264 with the use of the ideal noise-free reference frame are compared in Fig. 3.

It is not surprising that the compression efficiency decreases with a growing noise variance in general. Also, it is clear that the efficiency of P-frames compared to I-frames becomes less, if more uncorrelated noise between adjacent frames is present in the video signal. However, the curves determine that the compression gain of the P-frames over the I-frame decreases rapidly when the H.264 standard is used. If the reference frame is filtered by the adaptive Wiener filter before motion compensated prediction, the average bitrate for the P-frames becomes noticeably lower than the bitrate for the I-frame. For example, if the Y-signal contains additive white Gaussian noise with a variance of 21.4, an average H.264 standard conform coded P-frame saves about 1.8% of bitrate in comparison to the I-frame. When the proposed algorithm is used for denoising of the reference frame, an average P-frame saves about 5.2% of bitrate, and if the ideal denoised reference frame is used the bitrate savings for an average P-frame is about 6.7%.

As mentioned in the previous section, the test results can be improved by variation of the parameter σ_n^2 of the adaptive Wiener filter. Coding of the above mentioned sequences using $\sigma^2_{n,par} = 1.6 \cdot \sigma^2_{n,seq}$ leads to slightly better compression results. For example, if the luminance channel of the *Big Bucks Bunny* sequence is degradded by additive white gaussian noise with $\sigma^2_{n,seq} = 1.65$ adjusting the parameter to $\sigma^2_{n,par} = 2.64$ results in additionally 0.5% bitrate savings.

Usually, image sequences that are acquired by a physical process can be considered to be degraded by noise caused through the acquisition process (i.e., sensor noise, amplifier noise, quantization noise, ...). Therefore, we applied the proposed coding scheme also for lossless coding of natural test sequences without degrading them artificially. The following test results are based on high resolution test image sequences from [16]. The sequences have a spatial resolution of 3840x2160 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. For generation of the test results, we have used the same settings as mentioned before. It is worth to mention here that changing of different settings did not affect the gain accomplished by our algorithm noticeably. The compression results can be observed in Table 1.

Table 1. Compression results for high definition sequences. ($\hat{\sigma}^2_n$ is the noise variance to be used for filtering, I is the bitrate for I-frames, P orig. is the average bitrate for P-frames achieved by the standard, P mod. is the average bitrate for P-frames achieved by the modified algorithm and G is the gain of P mod. over P orig. in $\%$)

Sequence	$\hat{\sigma}^2_n$	Y-bitrate in bits per pixel			G [%]
		I	P orig.	P mod.	
ParkJoy	10	4.53	4.33	4.23	2.03
CrowdRun	30	4.62	4.44	4.30	3.15
InToTree	50	4.51	4.40	4.25	3.41
OldTownCross	60	5.00	4.96	4.73	4.64
DucksTakeOff	80	5.26	5.22	5.05	3.26

Generally it can be observed that our algorithm provides better compression results than the H.264/AVC standard. Under the assumptions of additive and white noise, the proposed algorithm achieves bitrate savings of up to 4.6% (for *OldTownCross*) in comparisson to H.264/AVC for P-frames and up to 6.9% (for *CrowdRun*) in comparison to Intra-Only coded sequence.

5. CONCLUSIONS

In this paper, we have shown how the bitrate of a lossless encoded video can be decreased by reducing noise in the reference frame. We have used the adaptive Wiener filter for denoising of the reference frame. The test results have shown that significant bitrate savings can be achieved. In the future, we are going to work on better adaptive filter algorithms for denoising of the reference frame.

6. 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 [17].

7. REFERENCES

[1] Marcelo J. Weinberger, Gadiel Seroussi, and Guillermo Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Transactions on Image Processing*, vol. 9, pp. 1309–1324, 2000.

[2] Michael W. Marcellin, Ali Bilgin, Michael J. Gormish, and Martin P. Boliek, "An overview of JPEG-2000," *Data Compression Conference*, vol. 0, pp. 523, 2000.

[3] Hari Kalva and Jae-Beom Lee, "The VC-1 video coding standard," *IEEE MultiMedia*, vol. 14, no. 4, pp. 88–91, 2007.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "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, August 2003.

[5] Michael Smith and John Villasenor, "Intra-frame JPEG-2000 vs. inter-frame compression comparison: The benefits and trade-offs for very high quality, high resolutin sequences," *SMPTE Technical Conference and Exhibition*, 2004.

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

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

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

[9] O.G. Guleryuz, "A nonlinear loop filter for quantization noise removal in hybrid video compression," in *Proceedings of ICIP 2005. IEEE International Conference on Image Processing*, Sept. 2005, vol. 2, pp. II–333–6.

[10] C.C. Dorea, O.D. Escoda, Peng Yin, and C. Gomila, "A direction-adaptive in-loop deartifacting filter for video coding," in *Proceedings of ICIP 2008. 15th IEEE International Conference on Image Processing*, Oct. 2008, pp. 1624–1627.

[11] Takashi Watanabe, Naofumi Wada, Goki Yasuda, Akiyuki Tanizawa, Takeshi Chujoh, and Tomoo Yamakage, "In-loop filter using block-based filter control for video coding," in *Proceedings of ICIP 2009. 16th IEEE International Conference on Image Processing*, Nov. 2009.

[12] S. Roy and S. Sengupta, "An improved video encoder with in-the-loop de-noising filter for impulse noise reduction," in *Proceedings of ICIP 2006. IEEE International Conference on Image Processing*, Oct. 2006, pp. 2605–2608.

[13] Jae S. Lim, *Two-dimensional signal and image processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1990.

[14] Heinrich-Hertz-Institut, "H.264/AVC software coordination," `www.iphome.hhi.de/suehring/tml/`.

[15] Blender Foundation, "Big buck bunny," `www.bigbuckbunny.org`.

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

[17] EU Seventh Framework Programme for Research and Development, "OPTIMIX," `www.ict-optimix.eu`.