Adaptive Global Motion Temporal Filtering

Andreas Krutz, Alexander Glantz, Michael Tok, and Thomas Sikora
Communication Systems Group
Technische Universitaet Berlin
Einsteinufer 17, D-10587 Berlin, Germany
Email: {krutz,glantz,tok,sikora}@nue.tu-berlin.de

***Abstract*—The emerging standardization on high efficiency video coding (HEVC) has brought a huge improvement in terms of coding performance comparing to existing standards and approaches. One tool that provides a significant portion of the coding gain so far is loop filtering. In H.264/AVC, a deblocking filter has been used to reduce blocking artifacts. HEVC added loop filter approaches to this deblocking method that further reduce noise in the decoded video frame. All these techniques work spatially. Besides that, work has been done to further improve the quality of decoded frames by applying a temporal filtering approach. In this work, we propose adaptive global motion temporal filtering (AGMTF) that reduces noise along temporal trajectories. Experimental results show that the coding performance of the current HEVC test model HM 4.0 can be improved by up to 8.8% and 3.7% in average over a large bit rate range using this technique.**

I. Introduction

In the near future, almost every multimedia device in the market will be able to display Full-HD video and beyond. For that, new highly efficient video codecs have to be developed to deliver this large amount of data. Current video coding standards are not designed for very high resolution video. Therefore, a new standardization activity is underway to develop a video coding standard that is designed to efficiently compress high and very high resolution video data. This joint activity from ISO/IEC MPEG and ITU, the Joint Collaborative Team on Video Coding (JCT-VC), is developing a High Efficient Video Coding (HEVC) Test Model HM that is still emerging. Several tools have made very large steps towards a much better coding gain in comparison to the state of the art. One of these tools is adaptive loop filtering for reduction of noise in decoded video frames. In the current HEVC Test Model HM 4.0 [1] , three loop filters are integrated. First, the deblocking filter [2] that was adopted from the latest video coding standard H.264/AVC. The second method is Sample Adaptive Offset (SAO) which was first introduced by [3] and has been improved by [4]. The third method is called adaptive loop filtering (ALF) and is based on a Wiener filtering approach. This filter design has significantly improved the coding efficiency and therefore, a large number of implementations including the one in [1] has been proposed. It has been shown that all these approaches bring a significant amount of gain compared to the state of the art. However, these filter approaches work spatially. On the other side it has been shown in the literature that additional quality improvements are possible using temporal filtering. Motion compensated temporal filtering has been widely used and published in several methods over the last two decades. A significant quality improvement of decoded video frames has been reported in [5]. Here, motion compensation, which is a key issue for the performance of temporal filtering, was performed based on higher-order motion parameters applied on the whole frame, i.e. estimating the camera motion. Using a very accurate motion estimation [6], it is possible to filter along a large amount of motion compensated frames. It is widely known from theory that this results in a much better noise reduction and quality improvement. One drawback of using a camera motion estimation for temporal filtering is that only pixels that correspond to the higher-order motion model, in this case background pixels, can be filtered. Pixels that do not correspond to the motion model remain and have to be filtered otherwise. In this work, we apply this concept to the latest HEVC test model as an encoder-assisted post-processing scheme. While using only camera motion estimation, we are able to filter background pixels over a large amount of motion compensated frames. Pixels that are different from the background motion will be filtered with the standard spatial filters integrated in the test model. We use an automatic block-wise decision that chooses either a temporally or a spatially filtered block for quality enhancement of decoded video frames. We call this method adaptive global motion temporal filtering (AGMTF).

The paper is organized as follows. In the next section, we overview the concept of global motion temporal filtering. To use this technique in a video coding environment, we have to transmit additional side information, i.e. higher-order motion parameters to build the motion compensated frame stack for temporal filtering. To be more efficient, we use a compression scheme for these motion parameters. Section III introduces this method shortly. In Section IV, we describe the proposed video coding scheme. The experimental set up and results are reported in Section V and the last section summarizes the paper.

II. Global Motion Temporal Filtering

Fig. 1 illustrates the concept of global motion temporal filtering. Adjacent frames of a reference to be filtered are motion compensated with respect to it using a higher-order motion estimation method. In this work, we use an algorithm based on feature correspondences and a robust Helmholtz esitmater along with the well-known 8-parameter perspective

978-1-4577-2049-9/12/$26.00

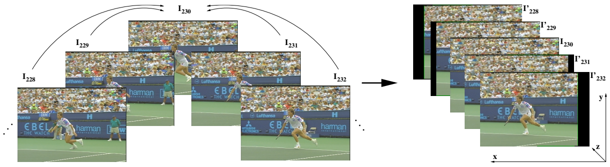


Fig. 1. Concept of Global Motion Temporal Filtering

motion model [6]. The estimation method is first applied on a frame-by-frame basis. Then, the resulted motion parameters are accumulated to build long-term parameters that every considered frame is compensated with respect to the reference. This results in an frame stack, on which the temporal filtering can be performed along every pixel trajectory using a simple mean operation. This compares to the known noise reduction method of overlapping noisy versions of the same signal and applying a mean function. To get a closer view, we assume a two-dimensional Gaussian distributed memoryless signal $x = x_n$ with a known D-R-function and consider two error components of the temporal noise reduction. In [7], it was found that the temporally overlapped quantization error represented by its variance $\sigma^2_{e_q}$ and the prediction error variance due to the motion estimation $\sigma^2_{e_m}$ is:

$$\begin{aligned} \sigma^2_{e_q} &= 2^{-2R}\frac{\sigma^2_x}{N} \\ \sigma^2_{e_m} &= N \cdot 2\sigma^2_\Delta \sigma^2_x (2-\alpha_1-\alpha_2), \end{aligned} \tag{1}$$

where σ^2_x is the variance of the signal, σ^2_Δ is the motion estimation error variance, α_1 and α_2 are the correlation factors in $x-$ and $y-$direction of the two-dimensional signal, and N is the number of noisy versions of the same signal. We assume that the final error variance is built by the sum of the two components shown above. Thus, the D-R-function of our model for the temporal noise reduction with (1) is:

$$\sigma^2_{e_{tf}} = 2^{-2R}\frac{\sigma^2_x}{N} + N \cdot 2\sigma^2_\Delta \sigma^2_x (2-\alpha_1-\alpha_2). \tag{2}$$

Now, it is of interest how possible bit rate savings are carried out from this theoretical D-R-function. For that, the distortion values of the D-R-function without filtering $\sigma^2_{e_{x_q}} = 2^{-2R} \cdot \sigma^2_x$ and (2) are set equal. The bit rate of the general quantization error shall be R_1 and the bit rate using temporal noise reduction shall be R_2. An equation of the bit rate R_2 can now be derived:

$$\begin{aligned} \sigma^2_{e_{x_q}} &\stackrel{!}{=} \sigma^2_{e_{tf}} \\ 2^{-2R_1}\sigma^2_x &= 2^{-2R_2}\frac{\sigma^2_x}{N} + N \cdot 2\sigma^2_\Delta \sigma^2_x (2-\alpha_1-\alpha_2) \\ 2^{-2R_2}\frac{1}{N} &= 2^{-2R_1} - N \cdot 2\sigma^2_\Delta (2-\alpha_1-\alpha_2) \\ R_2 &= -\frac{1}{2}\Big\{ ld\{2^{-2R_1} - \\ &\quad N2\sigma^2_\Delta(2-\alpha_1-\alpha_2)\} + ld(N)\Big\}. \end{aligned} \tag{3}$$

This means that there exists an optimal number of frames for filtering to achieve the highest possible bit rate reduction. For a further information on theoretical considerations see [7].

Thus, applying this concept in a video coding environment, an optimal number of frames has to be found for filtering. Therefore, we introduce an encoder-assisted post-processing method in Section IV. The second important issue is the accuracy of the higher-order motion parameters. We developed a compression scheme that reduces the bits needed for transmitting the parameters while keeping the accuracy as high as possible. This scheme is introduced next.

III. Motion Model Compression

A single higher-order motion parameter set, or parametric motion model (PMM), consists of eight parameters each represented by a 32 bit single precision floating point value. So, for performing AGMTF at the decoder for each frame additional 256 bit have to be transmitted. This would mean 6.4kBit/s more for a 25Hz sequence or 15.36kBit/s for a 60Hz sequence. Thus, the used PMMs have to be compressed in an efficient way. Since the parameters $m_{0,\ldots,}m_7$ are highly correlated and have different ranges of values, and as the two perspective parameters m_6 and m_7 are very sensitive to quantization, each PMM is transformed to a set of four corner motion vectors at the positions $(\pm x_{\text{res}}/2, \pm y_{\text{res}}/2)^T$, just following:

$$\begin{pmatrix} x' \cdot w' \\ y' \cdot w' \\ w' \end{pmatrix} = \vec{H} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{4}$$

$$\vec{V_p} = \vec{p'} - \vec{p} \tag{5}$$

These vectors are more robust to quantization and can easily be transformed back to a perspective model at decoder side. Additionally, each vector is highly correlated with its temporal predecessor so that differential coding in combination with exponential Golomb coding is used for redundancy reduction. The whole coding process for the PMMs is illustrated in Figure 2.

As quantization step size for the corner motion vectors, $\frac{1}{32}$ was found to be a good trade-off between bit rate and model quality.

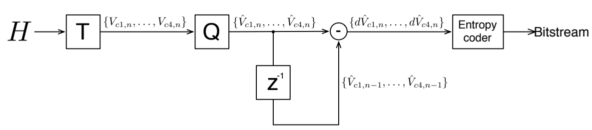


Fig. 2. Encoding process of PMM's

IV. Adaptive Global Motion Temporal Filtering using the HEVC Test Model HM 4.0

One critical issue of global motion temporal filtering is the exact amount of frames to be used for an optimal filtering. In Section II, it has been shown that for each sequence and noise

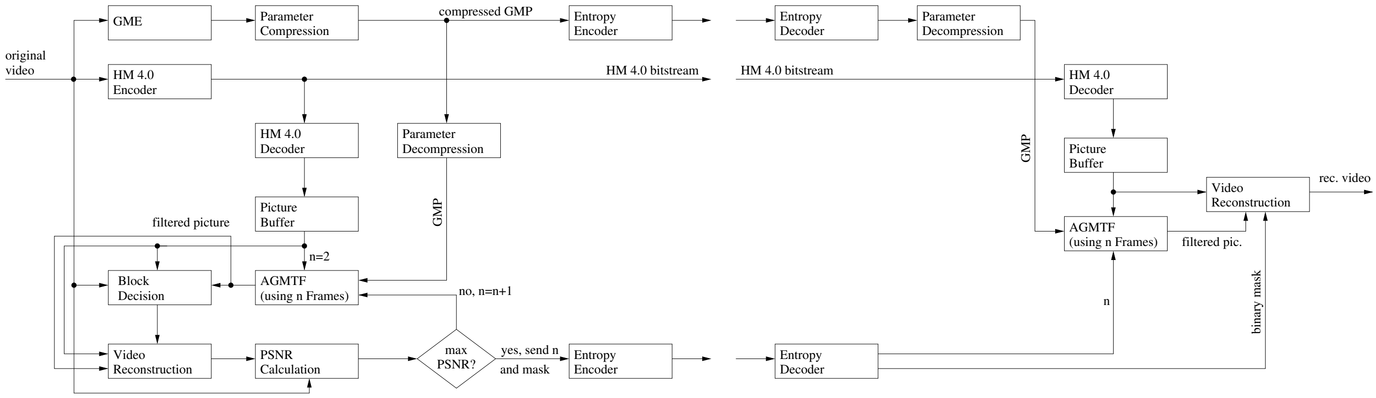


Fig. 3. HM 4.0 with AGMTF, encoder and decoder

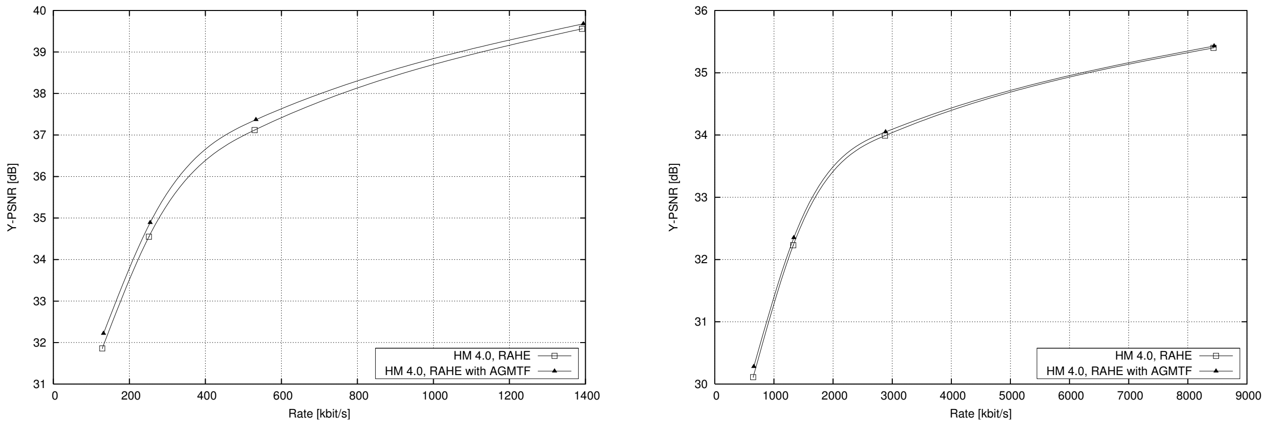
level an optimal number of aligned frames for filtering exists. To bring this to practice, we designed an encoder-assisted post-filtering scheme. Here, the higher-order motion estimation is performed along with a quality optimization method to find the optimal number of frames for filtering. This optimal number is determined for each frame of the video sequence. Fig. 3 shows the encoder and decoder with the proposed adaptive global motion temporal filtering using the HEVC Test Model HM 4.0. It can be seen that this scheme is independent from the codec itself. The scheme operates as follows. At the encoder, global motion estimation (GME) and encoding of the video sequence is performed in parallel (here using HM 4.0). Then, the resulting higher-order motion parameters are compressed with the method described in the previous section. To simulate the same conditions of the decoder, the motion parameters and the encoded video sequence are locally decoded and stored in a buffer. Afterwards, global motion temporal filtering is performed using the decoded frames and decompressed motion parameters starting with two frames (n=2). Having the reference frame temporally filtered, an MSE-based (MSE - mean-squared-error) block-wise decision is determined to evaluate if the spatially filtered block or the block filtered with GMTF results in a lower MSE which is then used to reconstruct the final frame. Here, a fixed block size is used. For Standard Definition (SD) resolution and lower, the block size of 64x64 pixel is used. For High Definition (HD) videos, the block size is set to 128x128. Having the final spatially/GMTF-filtered reconstucted frame, it is compared with its original using the PSNR. This procedure is performed until a predefined maximum number of frames for filtering is reached. In this scheme, we set the maximum number to 40. The number of frames for GMTF that results in the highest PSNR value is transmitted to the decoder along with a flag bit for each block to indicate whether the block is filtered using GMTF or remains spatially filtered. The numbers of frames for filtering are encoded using a simple Exp-Golomb code and the flag bit for each block is transmitted without any further processing. At the decoder, the side information is decompressed and the video data is decoded and stored in a frame buffer. Adaptive global motion temporal filtering is performed as a post-processing step using the number of frames for each decoded frame, flag bits and motion parameters to enhance the quality of the decoded frame.

V. Experimental Results

To evaluate our proposed scheme, we conducted experiments that show the performance of the adaptive global motion temporal filtering approach. To be as close to the current state of the art as possible, we used the HEVC Test Model HM version 4.0. For the encoding settings, the random access high efficiency (RAHE) case was considered as defined in [8]. We used two sets of QP values for encoding, i.e. $QP_{high} = \{22, 27, 32, 37\}$ and $QP_{low} = \{27, 32, 37, 42\}$ to evaluate a large bit rate range. For measuring the performance we used the widely known BD-rate [9]. The test sequences considered are listed in Table I. Two sequences are taken from the official HEVC test set and all remaining test videos are well-known in the field. It is emphasized that only sequences containing complex camera motion are considered in this study to see if the proposed concept still brings a gain on top of high efficiency video coding. It is obvious that global motion estimation has no effect on sequences where no camera motion exist. In these cases, a flag can be provided to switch off the GME or other motion estimation methods may be appled where temporal filtering can be performed. These are key issues for further work. In this study, we can see that AGMTF brings a significant improvement for HM 4.0 with the test sequences considered. The results are summarized in Table I. For two sequences, RD-curves are shown examplary in Fig. 4. A bit rate reduction of up to 9% roughly and almost 4% in average are achieved. It is also interesting to note that the average bit rate reduction in a higher bit rate range is the same as in the lower bit rate range. This means that despite the highly efficient spatial in-loop filters of HEVC there is an amount of noise that cannot be reduced by these filters. Adding the temporal filtering on top of the spatial filters results in a significant enhancement of the decoded frame quality. It is stated that a number of parameters in the proposed scheme including a fixed block size can be improved in future work.

TABLE I
TEST SEQUENCES

Sequence name	Size	FPS	Frames	BD-rate [%] QP_{high}	BD-PSNR [dB] QP_{high}	BD-rate [%] QP_{low}	BD-PSNR [dB] QP_{low}
BQSquare	416 × 240	60	600	−0.5	0.0	−0.4	0.0
BQTerrace	1920 × 1080	60	600	−2.3	0.0	−3.4	0.1
Blue sky	1920 × 1080	25	218	−2.1	0.1	−2.7	0.1
Jets1	1280 × 720	60	300	−6.3	0.1	−4.3	0.2
Station2	1920 × 1080	25	313	**−8.8**	0.2	**−7.7**	0.3
Sunflower	1920 × 1080	25	500	1.3	0.0	−1.1	0.0
Waterfall	704 × 480	25	260	−7.0	0.2	−6.3	0.3
mean				**−3.7**	**0.1**	**−3.7**	**0.1**



(a) Test sequence "Waterfall"

(b) Test sequence "BQTerrace"

Fig. 4. HM 4.0 vs. HM 4.0 with AGMTF

VI. SUMMARY

In this work an adaptive global motion temporal filtering scheme as an encoder-assisted post-processing method was proposed. For encoding and decoding the video data, the HEVC Test Model HM 4.0 was taken into account. The main goal of this study was to show that temporal filtering can tackle further noise reduction on top of the highly efficient in-loop filters of HEVC. Since the proposed method relies on background pixels due to the design based on global motion estimation, a block-wise adaptive method was developed to automatically decide at the encoder in which region the quality of the current frame is to be improved by temporal filtering. Additionally, an optimal number of frames for filtering in terms of the best frame quality was also found for each decoded frame. A significant bit rate reduction was achieved in comparison to the HM 4.0 without temporal filtering. Further steps in this work is to find an optimal adaptivity and combine the method with temporal filtering that relies on different motion estimation methods and models to temporally filter all pixels contained in a decoded frame.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Communitys FP7 under grant agreement number 261743 (NoE VideoSense).

REFERENCES

[1] K. M. et al., "HM4: HEVC test model 4 encoder description," in *Tech. Rep. JCTVC- F802, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG*, Torino, Italy, Jul. 2011.

[2] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, "Adaptive deblocking filter," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 614 –619, july 2003.

[3] K. M. et al., "Samsungs response to the call for proposals on video compression technology," in *Tech. Rep. JCTVC- A124, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG*, Dresden, Germany, Apr. 2010.

[4] C.-M. F. et al., "Ce13: Sample adaptive offset with lcu-independent decoding," in *Tech. Rep. JCTVC- E049, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG*, Geneva, Switzerland, Mar. 2011.

[5] A. Glantz, A. Krutz, M. Haller, and T. Sikora, "Video coding using global motion temporal filtering," in *IEEE International Conference on Image Processing (ICIP2009)*, Cairo, Egypt, Nov. 2009.

[6] M. Tok, A. Glantz, A. Krutz, and T. Sikora, "Feature-based global motion estimation using the helmholtz principle," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, Prague, Czech Republic, May 2011.

[7] A. Krutz, A. Glantz, and T. Sikora, "Theoretical consideration of global motion temporal filtering," in *IEEE International Conference on Image Processing (ICIP2011)*, Brussles, Belgium, Sep. 2011.

[8] F. Bossen, "Common test conditions and software reference configurations," in *Tech. Rep. JCTVC- F900, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG*, Torino, Italy, Jul. 2011.

[9] G. Bjøntegaard, "Calculation of average psnr differences between rd-curves," in *Tech. Rep. ITU-T SG16/Q.6 VCEG document VCEG-M33*, Mar. 2003.