GLOBAL MOTION TEMPORAL FILTERING FOR IN-LOOP DEBLOCKING

Alexander Glantz, Andreas Krutz, and Thomas Sikora

Communication Systems Group
Technische Universität Berlin
Berlin, Germany

ABSTRACT

One of the most severe problems in hybrid video coding is its block-based approach, which leads to distortions called blocking artifacts. These artifacts affect not only the subjective perception at the receiver but also the motion compensated prediction (MCP) that generates a prediction signal from previously decoded pictures. It is therefore directly connected to the amount of data that has to be transmitted. In this paper, we propose a technique called global motion temporal filtering for blocking artifact reduction. Other than common deblocking techniques, this approach does not reduce the blocking artifacts spatially. Filtering is performed temporally using a set of neighboring pictures from the picture buffer. This approach is incorporated into an H.264/AVC reference software. Experimental evaluation shows that the proposed technique significantly improves the quality in terms of rate-distortion performance.

***Index Terms*—** H.264/AVC, video coding, deblocking, temporal filtering, motion compensated prediction

1. INTRODUCTION

Research on video compression techniques has evolved over the last decades, which lead to the latest state-of-the-art codec H.264/AVC [1]. This video coding standard aims at compressing high-quality video contents at low bit rates. Nevertheless, the importance of ongoing research in that area has been outlined before in [2] and is today more urgent than ever, since for emerging high-definition video content more powerful compression algorithms need to be found.

One of the still existing problems in video coding today are distortions, i.e. blocking artifacts that strongly affect the perceived video quality at the receiver and decrease the motion estimation quality at the encoder leading to non-optimal prediction signal results. There are mainly two reasons for these artifacts. One of them is the quantization of transform coefficients. Depending on the coarseness of quantization this can cause visually disturbing edges between block boundaries. The second source is the motion compensated prediction. Here, blocks are predicted from temporally neighboring pictures that already have been locally decoded and therefore contain discontinuities at block boundaries. These are often copied into the interior of the prediction signal.

Research on deblocking filters has been vast. The H.264/AVC standard itself defines a deblocking filter that is based on the work by List et al. [3]. This algorithm first tries to distiguish between different kinds of discontinuities using boundary analysis. Here, it is assumed that depending on the kind of neighboring blocks, i.e. intra or inter coded, boundaries are more or less severe. The second step is spatial filtering of horizontal and vertical edges depending on the previously analyzed boundary strength. Although subjective quality and the prediction signal could be improved significantly, block artifacts are still visible in the decoded video especially at low bit rates.

Besides spatial deblocking approaches, research on using the temporal correlation between successive pictures in a video sequence has been done. Bi-predictive coding techniques or hierarchical B-pictures use information from two locally decoded pictures to build a prediction signal by taking a weighted average of two signals. However, due to the low amount of filtered signals and due to the low order of the motion model, i.e. a 2-parameter translational model, the quality of the prediction is not really high. This is especially true if complex motion like zoom or rotation is present. Other techniques use a long-term memory to use temporally further distant pictures for prediction [4] or a combination of translationally compensated blocks from many neighboring pictures [5]. In [6], the authors use an affine motion model on subareas to generate several reference pictures for MCP.

In [7], we have proposed an approach that combines global motion compensation using higher-order motion models and a large set of temporally neighboring pictures to create a single filtered representation of the current picture. Thereby, blocking artifacts could be reduced, which both increased the subjective and the objective quality. However, this approach has only been performed as a post-processing filter. Having access to the original video sequence, the encoder merely estimated the amount of pictures the receiver should use to

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2011] under grant agreement no. 216444.

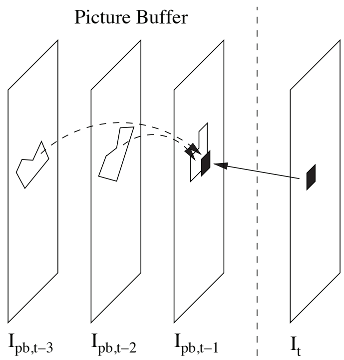


Fig. 1. Example for the generation of a filtered reference picture $I_{\text{pb},t-1,\text{filtered}}$ from previously decoded pictures for block-based MCP. Step 1: The pictures inside the picture buffer can be past and/or future pictures of the sequence depending on the GOP structure set in the encoder and are transformed into the coordinate system of the reference picture.

optimally reconstruct the sequence. Since not only the subjective perception is affected by blocking artifacts but also the results of motion estimation at the encoder, the next step is to incorporate the global motion temporal filtering (GMTF) technique into the encoder loop generating an enhanced representation of the reference picture in MCP. This is done by performing GMTF besides common spatial deblocking [3]. Two deblocked representations of the reference picture are generated, i.e. a spatially deblocked and a temporally filtered representation that uses the pictures available in the picture buffer. A block decision then merges both pictures on a macroblock level depending on a mean squared error-based (MSE) threshold. The resulting picture is then taken as reference for MCP.

This paper is organized as follows. Section 2 shows how global motion compensation and temporal filtering are performed. Section 3 explains the experimental setup by incorporating the GMTF approach into an existing H.264/AVC reference software. Section 4 describes the experimental results and the last section summarizes the paper.

2. GLOBAL MOTION COMPENSATION AND TEMPORAL FILTERING

Since many video sequences have been recorded with a moving camera, the background region of these sequences is not fixed. Theoretically this means that the picture buffer does not contain a set of equal signals only differing by the superimposed noise, which could easily be removed by averaging, but of a set of displaced image signals containing noise. This problem can be solved using higher-order motion models that account for the displacement a camera performs. This motion can then be compensated so that the signals are spatially aligned. The aligned representations can then be considered as equal signals differing only by quantization noise and blocking artifacts, respectively.

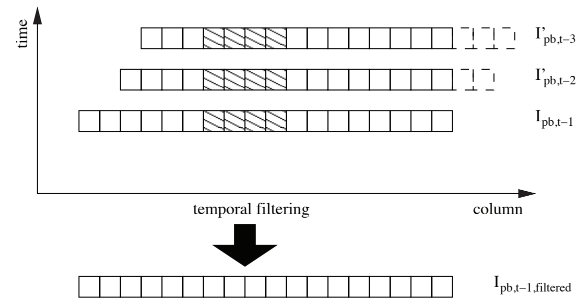


Fig. 2. Step 2: GMTF process for the generation of a filtered reference picture $I_{\text{pb},t-1,\text{filtered}}$ in block-based MCP. Depicted is one line y_k from a set of spatially aligned pictures.

Fig. 1 shows the exemplary transformation process of pictures from the picture buffer into the coordinate system of the reference for MCP. The current picture I_t is to be encoded using inter prediction. Therefore, its reference $I_{\text{pb},t-1}$ shall be used for finding block correspondences using common translational motion estimation, which is indicated by the black macroblock pointing to its spatially displaced representation. Assuming the reference contains blocking artifacts, these can be reduced using the GMTF approach. The regions that are subject to filtering from the temporal neighbors $I_{\text{pb},t-2}$ and $I_{\text{pb},t-3}$ are therefore transformed into the reference's coordinate system. For that, the long-term motion between the pictures involved in the process of global motion compensation has to be known. Thus, global motion estimation (GME) is incorporated into the approach following the algorithm previously presented in [8].

The transformation process creates an image stack of spatially aligned pictures. Part of this image stack can be seen in Fig. 2. One line of spatially aligned pictures is depicted. The pixel values inside the regions that are to be filtered are shaded. The picture $I_{\text{pb},t-1}$, i.e. the reference for MCP, and the transformed regions from its neighbors $I'_{\text{pb},t-2}$ and $I'_{\text{pb},t-3}$ are blended together using a median filter on the shaded regions. Thereby, a filtered representation $I_{\text{pb},t-1,\text{filtered}}$ is generated that can be used in MCP.

3. INCORPORATION INTO H.264/AVC

Fig. 3 shows the GMTF approach incorporated as a deblocking filter into a common hybrid video coding environment. Prediction modes other than inter prediction have been omitted from the block diagram. At the encoder, GMTF is performed besides common spatial deblocking [3] having access to the previously decoded pictures in the picture buffer. To generate a filtered reference for MCP, the GMTF algorithm iteratively transforms temporally neighboring pictures from the picture buffer into the coordinate system of the reference.

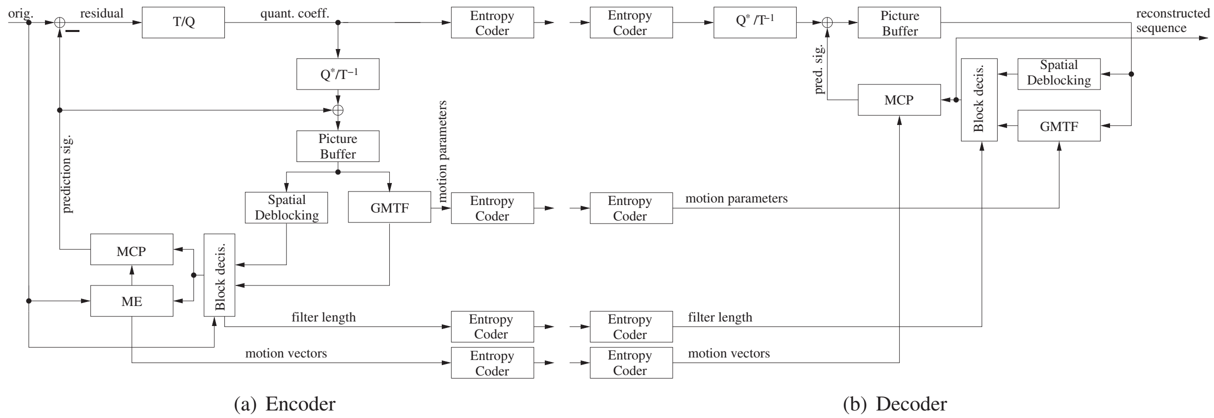


Fig. 3. GMTF for in-loop MCP refinement within a hybrid video coding environment: GMTF is performed using GME on the pictures inside the picture buffer besides common spatial deblocking. The encoder decides whether to use spatial deblocking or GMTF by minimization of the prediction error.

Sequence	Resolution	Frames	FPS [Hz]
Basketball	1024×576	300	25
BBC-Pan-13	720×576	110	25
Desert	720×400	240	25
Entertainment	720×576	250	25

Table 1. Test sequences used in the experimental evaluation.

A block decision module then decides on a macroblock-level whether to use the spatially deblocked representation or one of the temporally filtered blocks with an arbitrary filter length, i.e. an arbitrary amount of neighboring pictures from the picture buffer. This decision is based on the minimum MSE between original and filtered blocks. The filter length that produces the minimum MSE on a picture-level is chosen for all temporally filtered blocks in the picture. Therefore, only one filter length per picture has to be transmitted as side information besides the global motion parameters.

4. EXPERIMENTAL EVALUATION

For experimental evaluation, four test sequences have been considered that are listed in Table 1. The GMTF approach has been integrated into the H.264/AVC reference software KTA 2.6r1. The coding conditions are listed in Table 2.

Fig. 4 shows the experimental results in terms of rate-distortion performance. Bit rate savings (BRS) [9] up to 6.2% (*BBC-Pan-13*) could be reached. For all test sequences, the gains are high especially in upper bit rate ranges. In the lower regions, the proposed approach is similar or worse than common H.264/AVC deblocking, which is due to the fact that

Prediction structure	IPPP
Entropy coding	CAVLC
RD Optimization	enabled
QPPSlice $\in$	$\{28,33,38,43,48\}$
QPISlice	QPPSlice-1
MV search range	32 pels
Transform size	4x4

Table 2. Testing conditions used in the experimental evaluation for the H.264/AVC Baseline Profile.

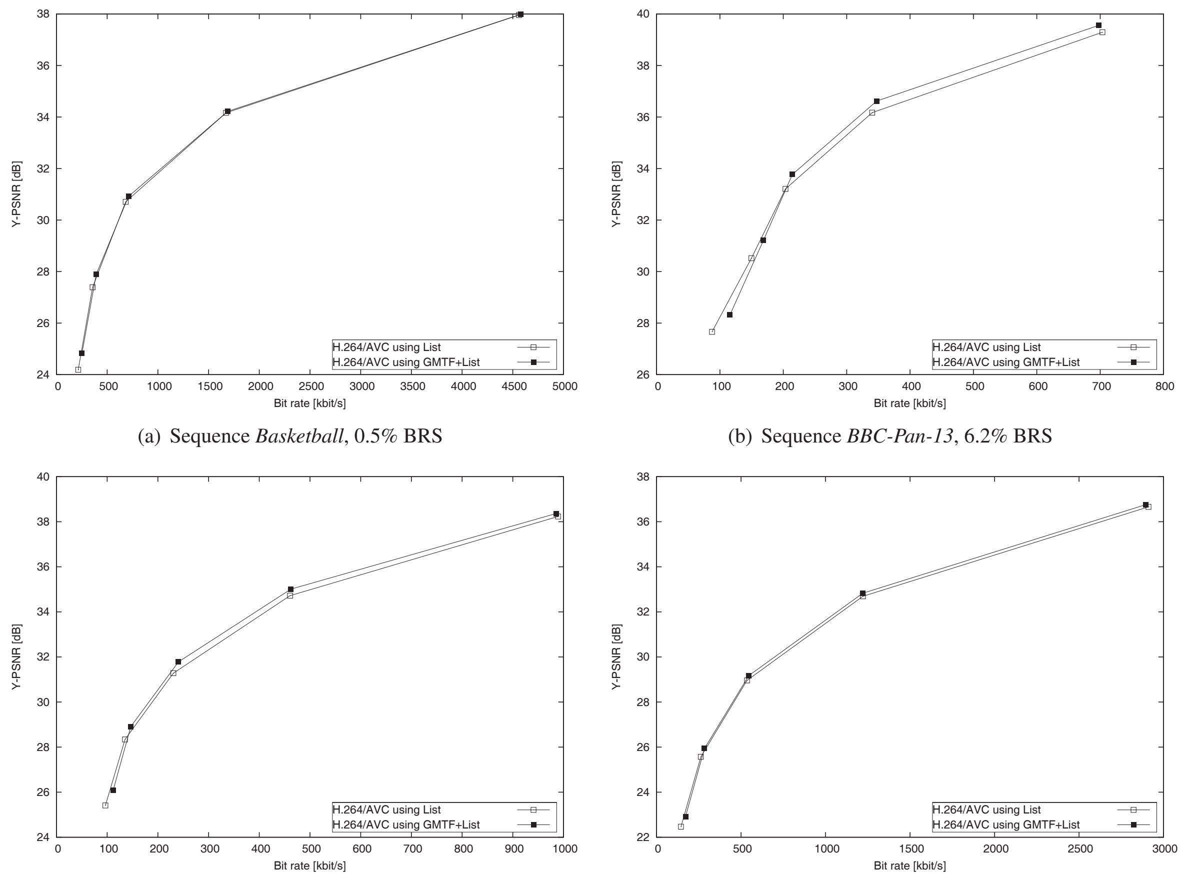
here, the side information is too expensive. It has to be explicitly remarked that no rate-distortion optimization is performed, which would solve this problem. Subjectively, the proposed approach outperforms H.264/AVC as well.

5. SUMMARY

We presented a new deblocking technique included inside the encoder loop of an H.264/AVC coding environment. For deblocking, temporal filtering of a set of spatially aligned pictures that have been previously encoded is performed besides spatial deblocking. Spatially and temporally filtered pictures are merged on a macroblock-level depending on minimum MSE to generate a refined reference picture for MCP. Experimental results show that the new approach significanly outperforms common spatial deblocking.

6. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,”



(a) Sequence *Basketball*, 0.5% BRS

(b) Sequence *BBC-Pan-13*, 6.2% BRS

(c) Sequence *Desert*, 4.3% BRS

(d) Sequence *Entertainment*, 2.6% BRS

Fig. 4. Rate-distortion results and bit rate savings (BRS) for the four test sequences considered. Rate-distortion curves for H.264/AVC coding using the spatial deblocking filter (List) only and for coding using a combination of spatial deblocking and temporal filtering (GMTF) are depicted.

Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, no. 7, pp. 560–576, Jul 2003.

[2] T. Sikora, "Trends and perspectives in image and video coding," in *Proceedings of the IEEE*, Jan 2005, vol. 93, pp. 6–17.

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

[4] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, 1999.

[5] G. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-93*, 27–30 April 1993, vol. 5, pp. 437–440.

[6] T. Wiegand, E. Steinbach, and B. Girod, "Affine multipicture motion-compensated prediction," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 2, pp. 197–209, Feb. 2005.

[7] A. Glantz, A. Krutz, M. Haller, and T. Sikora, "Video Coding using Global Motion Temporal Filtering," in *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP 2009)*, Cairo, Egypt, Nov 2009.

[8] A. Krutz, M. Frater, and T. Sikora, "Improved Image Registration using the Up-Sampled Domain," in *Proceedings of the 8th IEEE Workshop on Multimedia Signal Processing (MMSP 2006)*, Victoria, British Columbia, Canada, Oct 2006, pp. 447–450.

[9] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T SG16/Q.6 VCEG document VCEG-M33*, Mar 2001.