Lossy Parametric Motion Model Compression for Global Motion Temporal Filtering

Michael Tok, Andreas Krutz, Alexander Glantz, and Thomas Sikora
Technische Universität Berlin, Communication Systems Group

Abstract—It has been shown that techniques using higher-order motion parameters outperform common translational motion compensated prediction for hybrid video coders. A critical issue is the transmission of accurate higher-order motion parameters with as little additional bits as possible to maximize the compression gain of the whole system. For that, we propose an estimation method for perspective motion models using transformation before quantization and temporal redundancy reduction and integrate this scheme into a video coding environment using adaptive global motion temporal filtering. Experimental results show that using the proposed compression scheme for the perspective motion models, the BD-rate can be improved up to 8.8% in average in the higher bit rate range and up to 7.7% in average in the lower bit rate range compared to the latest version of the HEVC test model HM 4.0.

I. INTRODUCTION

For temporal redundancy reduction in common video coding standards like MPEG-2 [1], MPEG-4 Visual [2] or H.264/AVC [3], block-based motion estimation followed by motion compensation is utilized. However, this kind of motion compensation can only describe translational motion and delivers suboptimal results for higher-order motion such as rotation, zoom, or perspective deformation. Additionally, for each translationally compensated block, the motion information has to be encoded and transmitted in the form of motion vectors as well.

In contrast to block-based translational motion models, higher-order motion models have the ability to describe the camera transformation between two frames more precisely. Thus, by using such models, the generation of better prediction signals is possible leading to less prediction error information to be encoded. Additionally, for all regions that are described by such a higher-order motion model, no additional motion vector information has to be transmitted.

In [4], Glantz et al. introduced a method for H.264/AVC inter coding by using an 8 parameter perspective motion model for prediction signal generation and temporal filtering. Thereby, they save bits usually used for motion vectors and improve the prediction signal at the same time. Another way of increasing coding efficiency is to improve the quality of the decoded frames by filtering. For this purpose, the work in [5] describes how to use global, or parametric motion models (PMM) in more general terms, for temporal frame filtering to reduce block artifacts. The authors call their technique global motion temporal filtering (GMTF). Since highly precise parametric motion model estimation is not always possible at the decoder side, the required model parameters have to be transmitted in addition. This increases the bit rate. To reduce the amount of overhead caused by the additional parameters, motion model compression can be utilized.

For lossy compression of polynomial motion models with up to 12 parameters, Karczewicz et al. proposed to orthonormalize the coefficients of the models to obtain higher robustness to quantization [6]. Steinbach et al. [7] employ this technique to compress affine six parameter models. That way they extend the H.263 [8] inter prediction with a set of affine transformed reference frames. However, these polynomial models can only cover motion describable with linear combinations of basis functions while perspective motion models with 8 parameters can even follow perspective deformations which are also often introduced by camera motion.

We propose a method for compression of such perspective models to improve the performance of video coding methods that use PMMs. This compression scheme transforms a given model to global motion vectors at the corners of each frame. Subsequently, these vectors are quantized. As the motion between frames changes slowly in time, the temporal redundancy of the corner vectors is reduced by temporal difference coding. Then exponential Golomb coding is applied on these differences. With this proposed method, the amount of bits needed to transmit the PMMs can be reduced by a factor of up to 4 while the compensation quality decrease stays negligible. To demonstrate the benefit of this method in application to coding approaches based on PMMs, a block-adaptive version of the GMTF is incorporated into the latest HEVC test model HM 4.0 and extended by the motion model compression scheme.

The remainder of this paper is organized as follows. Section II shortly describes a robust estimation method based on the Helmholtz principle that is used for getting highly precise perspective 8 parameter motion models. In Section III, the steps of the presented method for compressing these models are described, the whole compression method is explained and the dependency of the quantization step size to the model quality is pointed out. The adaptive GMTF, which is extended by the model compression scheme is described in Section IV. Section V presents and discusses the results in terms of bit rate savings for adaptive GMTF solely and with the novel motion model compression. Finally, Section VI gives a summary of this paper.
II. MOTION MODEL ESTIMATION

To get a PMM that describes the complex transformations induced by camera motion, the parametric motion estimation method presented in [9] is used. Thus, for each frame 400 features are selected and tracked. Then, a robust estimator based on the Helmholtz principle is applied on the set of feature correspondences to reject outliers resulting from foreground motion and mistracking and to derive a precise PMM. This estimator takes \( m \) randomly selected subsets of two correspondences to generate one simplified 4 parameter motion model \( \mathbf{H}_k \) per subset

\[
\mathbf{H}_k = \begin{pmatrix}
\bar{m}_{0,k} & \bar{m}_{1,k} & \bar{m}_{2,k} \\
\bar{m}_{1,k} & \bar{m}_{0,k} & \bar{m}_{3,k} \\
0 & 0 & 1
\end{pmatrix}.
\]

This model is then used to define whether a feature correspondence of the whole set is an inlier or an outlier regarding to \( \mathbf{H}_k \). With the number of inliers \( N_k \) and the estimated error variance of these inliers \( \sigma_k \), a rating per subset is defined by

\[
\Phi_k = \frac{N_k}{\sigma_k}.
\]

Only for the inlier features \( \mathbf{x}_k \) and their tracked correspondences \( \mathbf{h}_k \) with the largest \( \Phi_k \), a final perspective PMM is calculated by Least Squares as

\[
\mathbf{h} = \left( \mathbf{A}_k^T \mathbf{A}_k \right)^{-1} \mathbf{A}_k^T \mathbf{x}_k,
\]

where \( \mathbf{A}_k \) is the perspective design matrix for the feature correspondences of the \( k \)th consensus set and \( \mathbf{h} = (m_0, \ldots, m_7)^T \) contains the final motion parameters.

III. MOTION MODEL COMPRESSION

As mentioned before, these perspective motion models \( \mathbf{h} \) have the ability to describe each displacement \( \mathbf{d} \) of a pixel at a position \( \mathbf{p} = (x,y)^T \) to a new position \( \mathbf{q} = (x',y')^T \) that follows a combination of translation, zoom, rotation, shearing, and perspective deformation:

\[
\begin{pmatrix}
x' \\
y'
\end{pmatrix} = \begin{pmatrix}
m_0 & m_1 & m_2 \\
m_3 & m_4 & m_5 \\
m_6 & m_7 & 1
\end{pmatrix} \cdot \begin{pmatrix}
x \\
y \\
1
\end{pmatrix}.
\]

The transmission of such models with single precision floating point accuracy (32 bit) needs \( 8 \times 32 \) bit = 256 bit. This e.g. means additional 6.4 kbit/s for a 25 Hz sequence or 15.4 kbit/s for a 60 Hz sequence. It would be possible to use lower model precision (24 or even 16 bit) or parameter quantization, to reduce this amount of data. But on the other hand such precision reduction would lead to extreme quality losses. Furthermore, for getting a set of motion parameters that are more robust to quantization, Karczewicz et al. propose to use orthonormalized versions of polynomial motion models. E.g., warping with affine models is done by the linear combination

\[
x' = m_0 \cdot x + m_1 \cdot y + m_2,
\]

\[
y' = m_3 \cdot x + m_4 \cdot y + m_5.
\]
HEVC utilizes two kinds of deblocking filters, both only working in spatial domain. In [5], a deblocking filter for the temporal domain, called global motion temporal filtering has been presented. This new type of filter uses parametric motion models to compensate a set of already decoded frames for building a so called image stack and fuses them to reduce blocking artifacts appearing in the H.264/AVC compression process. The optimal amount \( n_{\text{opt}} \) of images used for the filtering is determined at encoder side. One drawback of that filtering method is, that arbitrarily moving foreground objects are vanishing when the image stack is fused. To overcome this issue, Glantz et al. generate foreground segmentation masks at encoder-side and transmit them in addition to the motion models, \( m_{\text{opt}} \) and the original H.264/AVC bit stream. We use the GMTF approach to improve the image quality of encoded HEVC streams. Instead of segmentation, an adaptive block-based decision between temporal filtering and spatial filtering is made at encoder-side. Therefore, blocks with a size of 128 × 128 are taken and for each block a flag for using this adaptive GMTF or not is sent. Figure 4 overviews the whole encoder and decoder setting consisting of HM 4.0 and adaptive GMTF with model compression.

IV. ADAPTIVE GLOBAL MOTION TEMPORAL FILTERING

HEVC utilizes two kinds of deblocking filters, both only working in spatial domain. In [5], a deblocking filter for the temporal domain, called global motion temporal filtering has been presented. This new type of filter uses parametric motion models to compensate a set of already decoded frames for building a so called image stack and fuses them to reduce blocking artifacts appearing in the H.264/AVC compression process. The optimal amount \( m_{\text{opt}} \) of images used for the filtering is determined at encoder side. One drawback of that filtering method is, that arbitrarily moving foreground objects are vanishing when the image stack is fused. To overcome this issue, Glantz et al. generate foreground segmentation masks at encoder-side and transmit them in addition to the motion models, \( m_{\text{opt}} \) and the original H.264/AVC bit stream. We use the GMTF approach to improve the image quality of encoded HEVC streams. Instead of segmentation, an adaptive block-based decision between temporal filtering and spatial filtering is made at encoder-side. Therefore, blocks with a size of 128 × 128 are taken and for each block a flag for using this adaptive GMTF or not is sent. Figure 4 overviews the whole encoder and decoder setting consisting of HM 4.0 and adaptive GMTF with model compression.

V. EXPERIMENTAL EVALUATION

Six test sequences with differing resolution, frame rate and camera motion have been selected for evaluation. All sequences are encoded with the HEVC test model HM 4.0 [10], and by HM 4.0 with adaptive GMTF to show, how adaptive GMTF performs without model compression. Eventually, the model compression is incorporated into HM 4.0 in addition to adaptive GMTF. Table I shortly overviews the settings of the used HEVC reference coder.

As evaluation criterion for comparing the coding efficiency of the reference coder and the novel framework, the Bjøntegaard metric (BD-rate and BD-PSNR) is chosen [11]. Table II shows properties as resolution, frame rate and frame amount of the used test sequences and the coding results for adaptive GMTF with and without model motion compression in terms of BD-rate and BD-PSNR. For BQSquare, adaptive GMTF has an increased average BD-rate at higher QP ranges (2.3%) as well as at lower QP ranges (5.2%).

By using model compression in addition, it turns out, that bit rate savings of 0.5% for the high QP range and 0.4% for the low QP range are achievable for the same sequence. This example demonstrates the influence of an uncompressed models bit rate on coding efficiency. For Sunflower, the motion estimation itself is difficult due to the video content. Faulty parameters result in reduced filtering performance as well as in reduced model compression efficiency. Thus, losses in higher quality ranges appear. A reduction in coding gain when using model compression is only recognized for the Blue Sky sequence. Nevertheless, for all other sequences, in which adaptive GMTF provides coding gain, additional motion model compression further increases the coding efficiency.
VI. SUMMARY

A new method for lossy perspective motion model compression has been presented and explained. It uses transform, difference, and signed exponential Golomb coding. To show the performance of this approach, an adaptive version of the GMTF post filter is incorporated into the HM 4.0 and enhanced by our motion model compression method. Bit rate savings of up to 8.8% in comparison to the HM 3.2 reference can be observed, when switching from lossless model transmission to model compression. Furthermore, cases where GMTF without model compression results in bit rate increases. Then additional model compression can still result in coding gain.

For further improvements of the motion model compression scheme, spatial redundancy reduction and more efficient prediction techniques as polynomial prediction will be considered in further work. Increased coding gain can be achieved by using arithmetic encoding after Golomb coding as well.

REFERENCES