Long-term Global Motion Compensation for Advanced Video Coding

Aljoscha Smolić, Yuriy Vatis, Heiko Schwarz, and Thomas Wiegand
Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute (FHG-HHI), 10587 Berlin, Germany

Abstract

We present a new approach to video coding that utilizes video analysis based on global motion features. For each frame to be encoded a global motion parameter set is estimated with respect to a number of previously transmitted frames. Using this global motion parameter set, a new motion compensation technique inspired by super-resolution mosaicing is applied. Rate-distortion optimization is used to identify macroblocks that are only affected by global motion. For such macroblocks no prediction error is transmitted. They are purely reconstructed using long-term global motion-compensated prediction. Our results indicate significant bit rate savings compared to a state of the art H.264/AVC codec at the same visual quality.

1 Introduction

Global motion compensation (GMC) is an important tool for a variety of video processing applications including for instance segmentation and coding. The basic idea is that a part of the visible 2-D motion within video sequences is caused by camera operation (translation, rotation, zoom). A common approach is to model this *global motion* by a parametric 2-D model. In the past, efficient algorithms have been developed to estimate the global motion parameters between consecutive frames or over a longer period [4]. Finally, the estimated parameters are used to compensate the global motion.

In a video coding application the goal is to predict those parts of the images that move consistently with the global motion, in order to increase the coding efficiency. Such systems have been studied in numerous publications and it has been shown that the coding efficiency can be increased for sequences that contain significant global motion. As a result GMC has been adopted for the MPEG-4 video coding standard [1]. The subjective quality at a given bit-rate can be improved even more with a related technology called sprite coding [1], [4], which employs video mosaics for coding. However, the usage of sprite coding is quite restricted due to inherent constraints of these approaches.

In [5] we have presented an algorithm that exploits the spatial alias to produce mosaics and video with a higher spatial resolution than the original video. It is a combination of video mosaicing and super-resolution techniques. The resulting mosaics and video are much sharper and contain a lot more details compared to the original. Further we have shown in [6] how these super-resolution mosaics and video can be used for long-term global motion compensation (LT-GMC). It combines elements from both, GMC and sprite prediction, resulting in prediction results that are significantly better compared to standard GMC.

In this paper we present the integration of LT-GMC into the H.264/AVC video codec [2]. LT-GMC is applied on a macroblock basis to detect such macroblocks that are only affected by global motion. The decision is taken based on rate-distortion criteria, i.e. the operational control of the reference H.264/AVC encoder is used as detection module for global motion regions. The corresponding macroblocks are encoded in a special way without transmission of the prediction error. This results in significant bit rate savings without affecting the visual quality of the decoded images. This approach is very similar in spirit to the algorithm presented in [3] that applies texture analysis and synthesis to detect image regions with specific texture properties and to encode them in a special way.

In the next Section, we describe the proposed algorithm. Section 3 presents experimental results. Finally, Section 4 concludes the paper and gives an outlook to future work.

2 Algorithm Description

In this section we give a detailed description of the algorithms that consists of the parts global motion estimation, macroblock-based LT-GMC, mode decision, and encoding.

2.1 Global Motion Estimation

The multi-frame prediction tool [7] of H.264/AVC typically provides 5 decoded reference images to be used for motion compensation at the decoder. As first step the global motion of these 5 images is estimated with respect to the actual image to be encoded using a robust recursive algorithm [4]. The global motion is represented using the perspective motion model:

$$x_1 = \frac{a_1 + a_2 x_0 + a_3 y_0}{1 + c_1 x_0 + c_2 y_0} \qquad y_1 = \frac{b_1 + b_2 x_0 + b_3 y_0}{1 + c_1 x_0 + c_2 y_0}. \qquad (1)$$

These well known equations describe the transformation of a position in the reference image (x_0, y_0) to the position in the actual image (x_1, y_1). The transformation is controlled by a set of warping parameters with elements a_i, b_i, and c_i , thus we have 8 parameters per reference image and 40 parameters in total per encoded image that define the transformation.

2.2 Macroblock-based LT-GMC

LT-GMC as described in [5], [6] includes the generation of a super-resolution mosaic. All reference images are warped towards a mosaic of double resolution of the original images in both dimensions. The composition of the mosaic is done without interpolation. Fig. 1 shows on the left side details from 3 consecutive original frames of test sequence Stefan. The right side shows the corresponding details from the corresponding super-resolution mosaics. The images are scaled to same size using the word processor's utility. The details from the super-resolution video mosaics clearly show superior visual quality. They are much sharper, contain much more details, appear much less blocky, and most of all aliasing is highly reduced. Looking at original video in motion, annoying flicker artefacts can be noticed, that result from block structure artefacts, as shown on the left-hand side of Fig. 1, changing over time. In the sequence of super-resolution mosaics these block structure artefacts are drastically reduced, as shown on the right-hand side of Fig. 1, due to the spatial alias elimination capability of our algorithm. Therefore the resulting video sequence is flicker free.

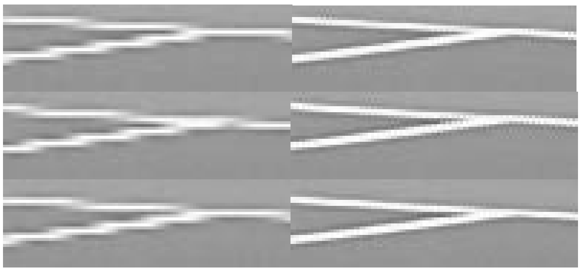

Fig. 1 Left: details from consecutive original frames of sequence Stefan, right: corresponding details from consecutive super-resolution mosaics

In [6] we have demonstrated that such super-resolution mosaics can significantly improve the prediction performance in terms of the residual error compared to standard GMC.

For integration into the H.264/AVC codec we developed a macroblock-based LT-GMC algorithm. We also modified the composition of the super-resolution prediction. Each macroblock to be predicted is scanned in lines and rows. Each sample is warped towards all reference frames using the respective global motion parameters as illustrated in Fig. 2. In general the transformed positions do not fall onto integer positions in the sample raster of the reference frames, but somewhere in between 4 neighbouring samples. A distance can be calculated to each of the neighbouring samples. The minimum distance is recorded along with the corresponding intensity or colour value. Such a minimum distance d_i is calculated for all 5 reference frames and finally the minimum of d_i is determined. The intensity or colour value of the corresponding sample is used for prediction of the current macroblock sample. Note, that no interpolation is performed to preserve the original sharpness.

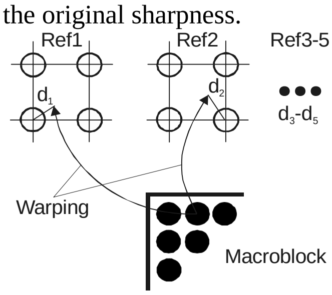

Fig. 2 Generation of prediction for a macroblock

2.3 Mode Decision and Encoding

The H.264/AVC codec supports a variety of different macroblock coding modes. In principle each encoder implementation has the freedom to decide about the encoding modes using any kind of proprietary operational control. The official reference encoder implementation applies a Lagrangian encoder control as described in [8] that optimizes encoding decisions with respect to their rate-distortion efficiency.

Given the quantization parameter QP for a macroblock k as well as the motion vectors and reference indices for all motion-compensated macroblock modes of this macroblock, the mode decision for the macroblock k proceeds by minimizing the Lagrangian functional

$$J_k(p \mid QP,\lambda) = D_k(p \mid QP) + \lambda \cdot R_k(p \mid QP), \qquad (2)$$

where the macroblock mode p is varied over the set of possible macroblock modes given by the slice type. The distortion $D_k(p \mid QP)$ is measured as the sum of squared differences (SSD) between the original samples of the macroblock k and their reconstructions that are obtained by using the coding mode p (with associated motion parameters) and the quantization parameter QP. The rate term $R_k(p \mid QP)$ specifies the number of bits associated with choosing the coding mode p and the quantization parameter QP including the bits for encoding the macroblock header, the motion parameters, and all transform coefficient blocks. The Lagrangian multiplier λ is determined according to

$$\lambda = 0.85 \cdot 2\text{^}(QP / 3 - 4). \qquad (3)$$

We have introduced a new macroblock mode, i.e. LT-GMC as described in the previous section. This new mode is included in the mode decision process, where the distortion and rate terms are calculated as if the LT-GMC mode comprises the transmission of a residual signal. However, for the encoding of LT-GMC macroblocks we introduce a heuristic assumption that is justified by intuition and the results. If LT-GMC results in the minimum of the cost functions, i.e. it is the best coding mode, then it is very likely that the corresponding macroblock is only affected by global motion. Then it can be reconstructed without transmission of the prediction error and of course without transmission of local motion vectors. This results in a reduction of the bit rate without affecting the visual quality although the PSNR performance drops. This algorithm is in the spirit of sprite coding schemes that also perform best if the content is purely reconstructed from previously transmitted content without transmission of texture updates [4]. In a sense the operational control of H.264/AVC is used as recognition tool for global motion regions and performs an automatic segmentation of the content.

Since the algorithm is fully integrated into the recursive encoding coding loop, errors cannot accumulate. If there are inaccuracies, LT-GMC will most likely not be the best coding mode and the errors will be corrected by transmission of prediction errors using other modes. Therefore our approach incorporates the advantages of standard GMC over sprite coding, i.e. being fully automatic, integrated and suitable for any type of content without relying on a priori knowledge (e.g. segmentation) of the content.

Our algorithm saves bits for transform coefficients and motion vectors, but on the other hand it introduces overhead for transmission of the global motion parameters. Having 40 parameters per encoded frame results e.g. in 960 parameters per second for a 25Hz sequence. Fortunately these parameters are highly redundant, and we have developed an efficient compression scheme.

We first normalize the parameters to the image dimensions. Knowing that the global motion cannot change discontinuously over time, we can predict the parameters from previously transmitted ones. The remaining residuals are γ-distributed and we apply scalar quantization and encode them using an exp-Golomb code.

The scheme has been integrated into the official H.264/AVC reference software, version JM2.1, so far only for P-frame encoding.

3 Experimental Results

We have tested the algorithm with a variety of test sequences, where we used P-frame only coding. Fig. 3 shows decoded images of the proposed and the reference codec for test sequence City (high definition 1280x720 samples, 30 frames). Visually no difference was observed in any of the decoded frames. The reference image was encoded with 445 kbit/s resulting in a PSNR of 40.3 dB. With LT-GMC we get 355 kbit and a PSNR of 39.3 dB. This means that we get a loss in PSNR of 1.0 dB which is not visible but expected since we do not perform waveform coding. On the other hand we get a bit rate saving of 20.2% at the same visual quality.

Fig. 3 Coding results for City, top: baseline H.264/AVC, bottom: with LT-GMC

The macroblock assignment mask is shown in Fig. 4. In total 1028 LT-GMC macroblocks have been assigned which corresponds to 28.5% of the total number. The sequence is captured by a camera in a helicopter flying over New York City. The skyscraper in the front is relatively close to the camera. Due to the translational motion of the camera it does not coincide with the global motion (motion parallax) and can therefore be regarded as a differently moving foreground object. Nevertheless the global motion is estimated very accurately since a lot of LT-GMC macroblocks are assigned in the background. But LT-GMC mode is not assigned to any of the foreground macroblocks. This example nicely illustrates the properties of our algorithm and validates the accuracy and efficiency.

Fig. 4 Assigned macroblock modes for the example in Fig. 3 bottom, grey: LT-GMC, black: other modes

Fig. 5 shows the relative bit rate savings. There is a significant gain for all bit rates. The largest gain of up to 26.5 % is achieved for lowest bit rates. Then the gain drops to 9.5 % for low bit rates. For medium and high bit rates the gain is relatively constant between

14 % and 16 %. 2 different effects influence the slope of the curve. At high and medium bit rates the total savings mainly come from savings on texture updates. Savings on motion vectors are less important at these bit rates. The relative savings on texture updates decrease to low bit rates. At very low bit rates the savings on motion vectors become important resulting in large total savings.

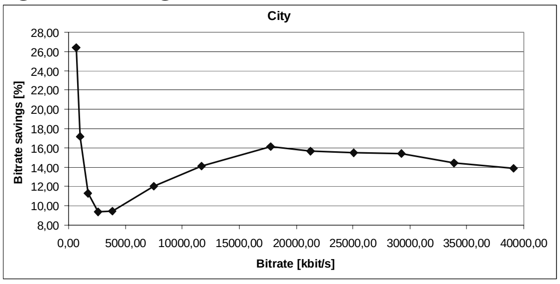

Fig. 5 Relative bit rate savings of LT-GMC codec compared H.264/AVC codec at the same visual quality over bit rate for City

These impressive gains were achieved for high resolution formats. The gain decreases with the resolution of the source video. Fig. 6 shows the bit rate savings for Mobile & Calendar (SIF 352x240 samples, 300 frames). The gain decreases from about 5% at highest bit rates to negative values at lowest bit rates. In this case the overhead for the global motion parameters becomes significant at low bit rates, we therefore don't get an increase as in Fig. 5. This overhead is quite constant over the bit rate and over the resolution of the source video. For high definition resolution it can be neglected, but for SIF resolution it becomes important at low overall bit rates.

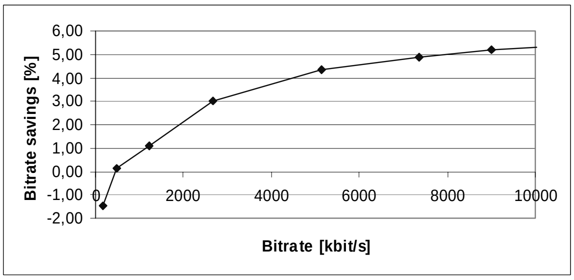

Fig. 6 Relative bit rate savings of LT-GMC codec compared H.264/AVC codec at the same visual quality over bit rate for Mobile & Calendar

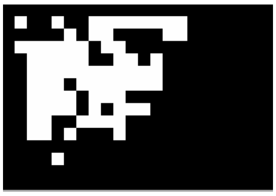

Fig. 7 Example macroblock modes for Mobile & Calendar, grey: LT-GMC, black: other modes

An example mode assignment mask for Mobile & Calendar is show in Fig. 7. In total 100 macroblocks are coded with LT-GMC. The average over the sequence is about 20-30 for Mobile & Calendar. Apparently the relative size of the macroblocks is larger compared to the high resolution examples, which results in a lower probability for assignment of LT-GMC mode.

4 Conclusions and Future Work

We have presented a new approach to video coding. It relies on an automatic detection of global motion regions, which are encoded with a special algorithm. It works best for high resolution sequences that are dominated by global motion. In the worst case, if no LT-GMC macroblocks are assigned, the algorithm falls back to standard H.264/AVC. In such cases the transmission of global motion parameters should be switched off resulting in a neglectable overhead. This still needs to be implemented. Also the integration of LT-GMC for B-frames is still an open issue.

5 References

[1] ISO/IEC 14496, Part 2 (Visual), Ammendment 1 "Information Technology - Coding of Audio-Visual Objects (MPEG-4)", February 2000.

[2] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10 FDIS Advanced Video Coding", Doc. N5555, Pattaya, Thailand, March 2003.

[3] P. Ndjiki-Nya, B. Makai, A. Smolić, H. Schwarz, and T. Wiegand, "Improved H264/AVC Coding Using Texture Analysis and Synthesis", Proc. ICIP'03, IEEE International Conference on Image Processing, Barcelona, Spain, September 2003.

[4] A. Smolić, T. Sikora, and J.-R. Ohm, "Long-Term Global Motion Estimation and its Application for Sprite Coding, Content Description and Segmentation" IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 8, pp. 1227-1242, December 1999.

[5] A. Smolić, and T. Wiegand, "High-Resolution Video Mosaicing", Proc. ICIP'01, IEEE International Conference on Image Processing, Thessaloniki, Greece, October 2001.

[6] A. Smolić, Y. Vatis, and T. Wiegand, "Long-Term Global Motion Compensation Applying Super-Resolution Mosaics", Proc. ISCE'02, IEEE International Symposium on Consumer Electronics, Erfurt, Germany, September 2002.

[7] 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, Feb. 1999.

[8] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards", IEEE Transactions on Circuits and Systems for Video Technology, scheduled July 2003.