Long-Term Global Motion Compensation Applying Super-Resolution Mosaics

Aljoscha Smolić, Yuriy Vatis and Thomas Wiegand
Heinrich-Hertz-Institute (HHI)
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany

smolic@hhi.de

Abstract: A new method for global motion compensation is presented significantly improving existing approaches. For each frame to be predicted a super-resolution mosaic is generated that has double spatial resolution in both dimensions. The very accurate and robust mosaicing algorithm uses *N* preceding frames, which are available at the decoder too in a hybrid video coding scenario. Experimental results are presented for uncoded and coded test sequences, which compare our approach to standard global motion compensation. We report improvements of the prediction performance in terms of PSNR of up to 7.4 dB for single frames and 2.22 dB on average. For some frames we get a decrease of the prediction performance suggesting an adaptive algorithm.

Keywords: Global motion compensation, video mosaicing, super-resolution

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. 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 new video coding standard MPEG-4 [3]. The subjective quality at a given bit-rate can be improved even more with a related technology called sprite coding [2], [3], [4], [6], which employs video mosaics for coding. However, usage is quite restricted due to inherent constraints of these approaches.

In this contribution we present a new global motion-based prediction method for inclusion in a hybrid coding scheme that combines elements from both, GMC and sprite coding. It has been shown in [5] that video mosaicing algorithms can be used to exploit redundancy in video sequences in order to extract additional visual information. In [9] 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. This paper shows how these super-resolution mosaics and video can be used for long-term global motion compensation (LT-GMC) within a hybrid codec framework and presents prediction results that are significantly better compared to a standard GMC approach.

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

2.1 Generation of super-resolution mosaics

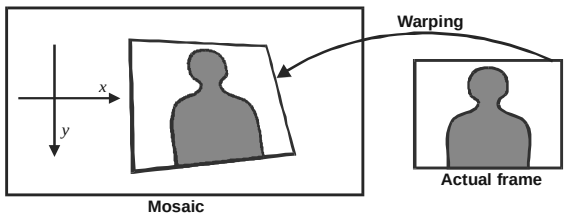

Fig.1 Process of mosaicing: warping and blending all frames of a video sequence towards a common reference system, controlled by estimated global motion parameters

The generation of super-resolution mosaics has already been presented in [9]. As shown in Fig.1, video mosaicing is the process of warping and blending all images of a considered video sequence into a common reference coordinate system. For LT-GMC we use the local pixel coordinate system of the current image to be predicted. In general this can be a further transformed coordinate system, for instance a cylinder or sphere. In any case, accurate warping parameters with respect to the common coordinate system have to be determined. Otherwise artifacts would become visible in the mosaic. Hence, one prerequisite for the generation of mosaics from multiple images is an accurate description and estimation of the global motion.

If the camera operation is restricted to zoom and rotation (i.e. no translation) and the pinhole camera model holds, the global motion can be exactly described by the perspective model:

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

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. Most GMC algorithms apply the simpler affine or even simpler models (also in MPEG-4 [3]). These are mathematically simpler and require less complex implementation. This approximation is justified for a short-term motion description between consecutive frames, but is not suitable for a long-term description with respect to a fixed reference as needed for LT-GMC [8]. However, the perspective model is also an approximation if the camera translates and rotates.

For a fast, reliable and very accurate estimation of the warping parameters, we utilize a combination of a differential and a feature-matching algorithm [6], [7]. After estimation of the warping parameters, the mosaic is constructed by warping the frames towards the common reference. Since most of the visual information is visible in several images, a blending procedure has to be defined, for instance, averaging or median filtering, using only the most recent or first data. In our super-resolution mosaicing algorithm we avoid any filtering or averaging of pixel values. This preserves the original sharpness of the video sequence, through elimination of spatial alias.

Fig.2 illustrates the approach, which is motivated from motion-compensated spatio-temporal filtering and interpolation algorithms [1]. The basic difference is that we apply a global motion model instead of a dense motion vector field.

The pixels of the video frames are transformed into a mosaic of double resolution in both directions, controlled by the estimated warping parameters, which have to be scaled accordingly. First, the frame at time instant t_0 is written into the mosaic of double resolution leaving empty pixel positions in the mosaic at time instant t_0. Second, a half-pixel diagonal shift from frame t_0 to frame t_1 is assumed. In this case, the corresponding pixels in the mosaic are filled since they fall directly onto an integer-pixel position in the mosaic. Note, that we do not interpolate any mosaic pixel. Only mosaic pixels that are directly hit by warped video frame pixel are updated accordingly, using the intensity (or color) value of the video frame pixel. In practice we use a tolerance range of e.g. ± 0.2 pixel units in the mosaic. The remaining empty pixel positions are filled when processing more frames.

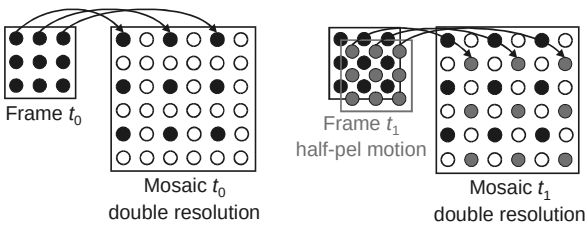

Fig.2 Principle of video mosaicing with super-resolution superposition: transformation of video pixels into mosaic of double resolution in both directions, without interpolation of intermediate intensity or color values

Depending on the global motion in the scene a certain number of pixels can remain unfilled after processing a certain number of frames. For instance if there is no global motion or only full-

pel pure translational motion only ¼ of the super-resolution mosaic pixels could be filled. We have found experimentally, that N=40 frames are clearly sufficient in most cases. Remaining holes in the super-resolution mosaic are filled by standard GMC with interpolation.

In our experiments so far, we have used a priori available segmentation masks, to exclude differently moving foreground objects, which would otherwise cause artifacts in the super-resolution mosaics. These foreground objects can be up-sampled using standard methods, and pasted over the super-resolution mosaic.

2.2 Generation of super-resolution video

The super-resolution mosaic generation can be repeated for every time instant of the video sequence, using for instance past N frames. The result is a sequence of super-resolution mosaics, i.e. super-resolution video. This is illustrated in Fig.3.

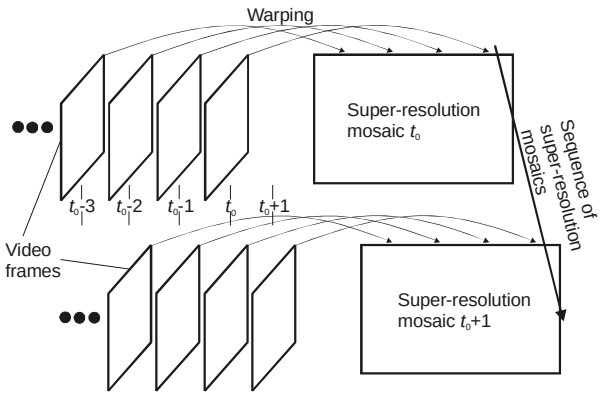

Fig.3 Generation of a sequence of super-resolution mosaics (i.e. super-resolution video) from a video sequence

Fig.4 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 artifacts can be noticed, that result from block structure artifacts, as shown in the left part of Fig.4, changing over time. In the sequence of super-resolution mosaics these block structure artifacts are drastically reduced, as shown in the right part of Fig.4, due to the spatial alias elimination capability of our algorithm. Therefore the resulting video sequence is flicker free.

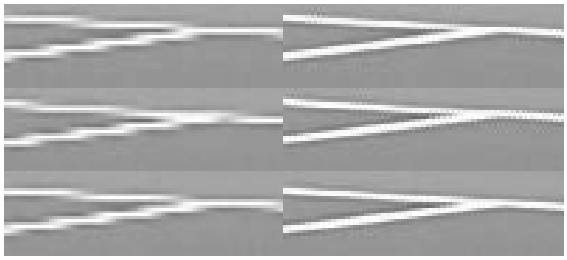

Fig.4 Left: details from consecutive original frames of sequence Stefan, right: corresponding details from consecutive super-resolution mosaics

2.3 Prediction with super-resolution mosaics

For integration into a hybrid codec framework we have developed a suitable prediction scheme. For every frame to be predicted a super-resolution mosaic is generated from N=40 preceding frames as described in the previous sections. The local coordinate system of the current frame is used as common reference, scaled by a factor of 2 in both dimensions.

For each of the N=40 contributing frames global warping parameters are computed with respect to the common reference. This is done by concatenation and scaling of frame-to-frame global warping parameters, which are estimated between consecutive original frames of the video sequence using the algorithm described in [7]. Effects of error accumulation can be neglected since the accurate perspective model and an accurate and robust estimator are applied. Moreover the number of contributing frames is relatively small compared to other mosaicing applications and the super-resolution algorithm favors the most recent frames.

In this scheme only the set or warping parameters between the current and the previous picture needs to be transmitted per frame, which is identical with standard GMC algorithms. This means that LT-GMC does not increase the overhead for transmission of the warping parameters compared to GMC.

The residual prediction error is calculated as difference of a background pixel from the original frame and the corresponding pixel from the generated super-resolution mosaic, i.e. with coordinates scaled by a factor of 2.

3 Experimental results

In our experiments we have compared the performance of GMC and LT-GMC in terms of prediction accuracy. We performed standard GMC and LT-GMC for several test sequences and calculated the PSNR of the predictable background pixels.

3.1 Results with uncoded test sequences

In these experiments we generated the super-resolution mosaics from uncoded video and we also performed GMC from uncoded sequences both using the same set of motion parameters. Results are shown in Fig.5 and Fig.6 for the test sequences *Mobile & Calendar* and *Stefan* respectively. Significant improvements of up to 7.4 dB for *Mobile & Calendar* and up to 5.5 dB for *Stefan* are achieved by LT-GMC. The mean PSNR over all frames is 31.22 dB (LT-GMC) vs. 29.00 dB (GMC) for *Mobile & Calendar* and 30.76 dB (LT-GMC) vs. 30.00 dB (GMC) for *Stefan*. We achieved a mean gain of 2.22 dB for *Mobile & Calendar* and 0.76 dB for *Stefan*.

However, for some frames the performance drops (up to -1.9 dB for *Mobile & Calendar* and -3.8 dB for *Stefan*). One reason is for instance the rapid camera pan in parts of the *Stefan* sequence that results in a lot of motion blur. This has a bad impact on estimation and prediction accuracy especially if multiple frames are used. One possibility for further improvement of our approach would be to trade-off GMC and LT-GMC, e.g. to switch on a frame or macro-block basis. Standard GMC approaches use such switches anyway in order to signal local MC or GMC. This can also be used to avoid the transmission of segmentation masks. For example, when switching on a frame-by-frame basis, the gain increases to 2.35 dB for *Mobile & Calendar* and 1.16 dB for *Stefan.*

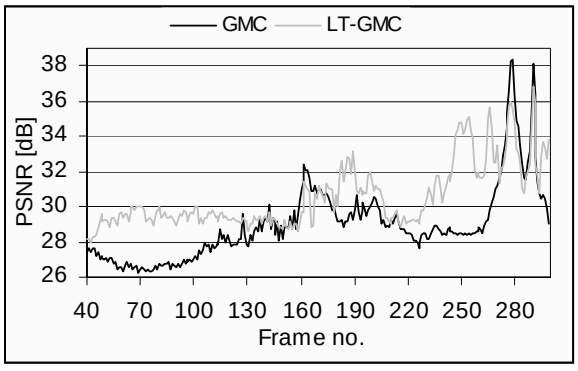

Fig.5 Prediction performance of GMC and LT-GMC for sequence Mobile & Calendar, uncoded

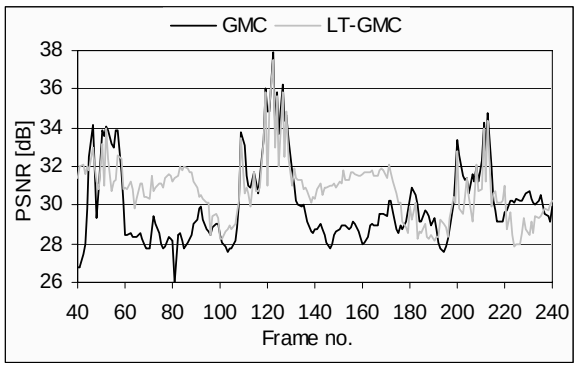

Fig.6 Prediction performance of GMC and LT-GMC for sequence Stefan, uncoded

3.2 Results with coded test sequences

In order to get a first evaluation of how our algorithm would perform in a real hybrid coding scheme we conducted some experiments with precoded test sequences. We encoded them using H.26L (TML.8.0) which is currently the best available video codec and is also the target platform of our algorithm. The test sequences were encoded using a number of fixed quantizer scales and typical encoder settings.

Global motion estimation was carried out as before using the uncoded test sequences, but the super-resolution mosaics were generated from the different coded versions and GMC was also carried out using coded video. Then the prediction performance in terms of background PSNR was measured as before in comparison to the original video. Of course these results do not represent exactly the performance of our algorithm integrated into H.26L, but they give some first insights in comparison to standard GMC at different bit-rates.

Fig.5 and Fig.6 show the mean background PSNR achieved with GMC and LT-GMC for the different quantizer values and test sequences *Mobile & Calendar* and *Stefan* respectively. LT-GMC outperforms GMC for all quantizer scales. The large gain of e.g. more than 2 dB for *Mobile & Calendar* is achieved over a wide range of the larger quantizer scales. For the smaller quantizer scales, the gain decreases linearly with the quality of the video. The figures also include the results of a combination of both methods using frame-by-frame switching. This leads to further significant improvements especially for test sequence *Stefan*.

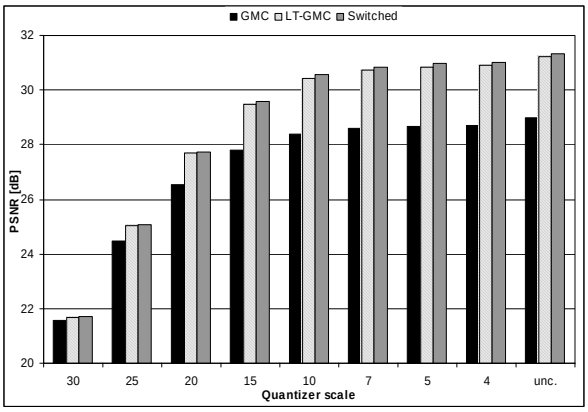

Fig.7 Prediction performance of GMC, LT-GMC and combination for sequence Mobile & Calendar, coded

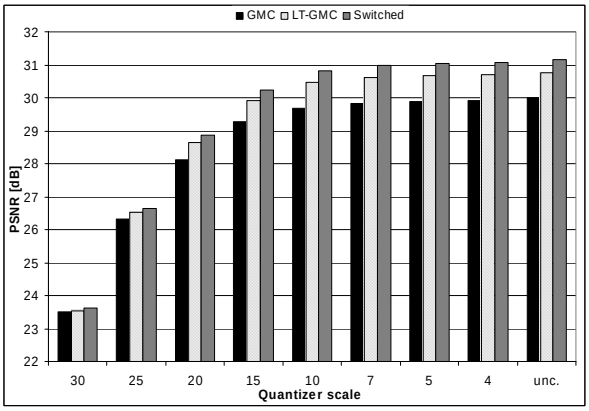

Fig.8 Prediction performance of GMC, LT-GMC and combination for sequence Stefan, coded

4 Conclusions and future work

We have presented a new approach for LT-GMC that significantly outperforms standard GMC algorithms. Super-resolution mosaics are used for prediction instead of only the last decoded frame. We have reported PSNR improvements of up to 7.4 dB for single frames and up to 2.22 dB in average. However, for some frames GMC provides better results than LT-GMC. Future optimization work will therefore include an integration and trade-off of both approaches. Finally, LT-GMC will be integrated into a hybrid codec such as H.26L.

References

[1] G. De Haan, "Progress in motion estimation for video format conversion", IEEE Transactions on Consumer Electronics, Vol. 46, No. 3, pp. 449-459, August 2000.

[2] F. Dufaux and F. Moscheni, "Background Mosaicing for Low Bit Rate Video Coding", Proc. ICIP'96, IEEE International Conference on Image Processing, Lausanne, Switzerland, September 1996.

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

[4] M.-C. Lee, W. Chen, C.B. Lin, C. Gu, T. Markoc, S.I. Zarbinsky and R. Szeliski, "A Layered Video Object Coding System Using Sprite and Affine Motion Model", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 1, February 1997.

[5] Y. Schechner and S. Nayar, "Generalized Mosaicing" Proc. ICCV'2001, International Conference on Computer Vision, Vancouver, Canada, July 2001.

[6] A. Smolic, T. Sikora and J.-R. Ohm, "Long-Term Global Motion Estimation and its Application for Sprite Coding, Content Description and Segmentation", IEEE Trans. on CSVT, Vol. 9, No.8, pp. 1227-1242, December 1999.

[7] A. Smolic and J.-R. Ohm, "Robust Global Motion Estimation Using a Simplified M-Estimator Approach", ICIP'2000, IEEE International Conference on Image Processing, Vancouver, Canada, September 2000.

[8] A. Smolic, K. Müller and J.-R. Ohm, "Global Motion Compensation and Video Mosaicing using different 2-D MotionModels", Proc. PCS'2001, Picture Coding Symposium, 25.-27. April 2001, Seoul, Korea.

[9] A. Smolic and T. Wiegand, "High-Resolution Video Mosaicing", Proc. ICIP'2001, IEEE International Conference on Image Processing, Thessaloniki, Greece, October 7.-10. 2001.