GLOBAL MOTION GUIDED ADAPTIVE TEMPORAL INTER- / EXTRAPOLATION FOR SIDE INFORMATION GENERATION IN DISTRIBUTED VIDEO CODING

Ralph Hänsel, Erika Müller

Institute of Communications Engineering
University of Rostock, Germany
{ralph.haensel,erika.mueller}@uni-rostock.de

ABSTRACT

The ongoing research on Distributed Video Coding (DVC) is focused on flexibility and increased rate distortion (RD) performance. Besides Slepian-Wolf (SW) coding and coder control, a good side information (SI) quality is essential for high RD performance. The SI is typically extracted by temporal inter-/extrapolation. Fast motion is still a challenge.

Two global motion guided adaptive temporal inter-/extrapolation (GTIE/GpTIE) schemes are proposed. They incorporate fast camera motion by global motion estimation and subsequent refinement. The issue of occlusion and revelation at the frame border is solved by an adaptive temporal inter-/extrapolation method. Simulation results show an RD performance increase of up to $3.1\,\mathrm{dB}$.

***Index Terms*—** Wyner-Ziv Coding, Temporal Interpolation, Temporal Extrapolation, Video Coding

1. INTRODUCTION

Modern multimedia communication systems demand efficient video coding algorithm. Therefore, state-of-the-art systems adopt conventional video coding (e.g. MPEG-4, H.264/AVC), which is well suited for broadcasting scenarios. Hence, the complex video encoding is done once, whereas low complexity decoding is applied many times at the client side. That approach is less adequate for emerging application scenarios, requiring low power and low computational complexity encoding.

In contrast, Distributed Video Coding (DVC) gives the ability to design low complexity video encoders. It is based on the theories of Slepian and Wolf [1] as well as Wyner and Ziv [2]. Furthermore, DVC is well suited for video encoding at mobile devices [3],[4].

The input video sequence is split up into H.264intra encoded key frames K and Wyner-Ziv (WZ) frames X, in a low complexity pixel domain DVC system (fig. 1). The WZ frames are coded based on a low complexity encoding and high complexity decoding scheme. The encoding process includes 2^M-step linear quantization and bit plane $q^{(b)}$ extraction. Subsequently, parity symbols are generated for each bit plane in the Turbo Encoder (Slepian-Wolf encoder). The WZ decoder, at first, generates an estimation of the WZ frame (side information) based on the decoded key frames. Subsequently, it requests party symbols to correct the SI in the Turbo decoding (SW decoding) and reconstruction process.

The key parts of a DVC system are the Slepian-Wolf (SW) encoder/decoder and the side information (SI) generation algorithms (sec. 2). The SI is commonly generated by temporal inter- (TI) [3] or extrapolation (TX) [5]. A good estimation quality is essential for high RD performance. Both schemes (TI/TX) show good performance for video sequences with slow motion. But in case of fast motion, the performance of block matching-based TI/TE is significantly reduced. On the one hand, these algorithms mostly cannot handle long motion vectors. Due to the large search range, the motion vector might not represent the true motion. On the other hand, TI cannot handle occlusion or revelation, because a valid block is needed in both reference frames. Furthermore, TX is robust in occluded areas, but not robust against revelation, which frequently occurs at the frame border in case of fast camera motion.

We propose two global motion guided adaptive temporal inter-/extrapolation methods (GTIE, GpTIE, sec. 3). Fast motion is coped with by applying a robust global motion estimator. The extracted global motion vector, on the one hand, is used to switch between TI and forward/backward TX. On the other hand, the temporal inter-/extrapolation process is guided by the extracted global motion. This scheme can cope with fast motion as well as occlusion and revelation. Thus, the side information quality is increased up to $3.5\,\mathrm{dB}$ (sec. 4) and the overall RD performance is improved up to $3.1\,\mathrm{dB}$. Finally, conclusions are given in section 5.

2. SIDE INFORMATION GENERATION IN DVC

Side information generation in low complexity DVC, commonly exploits the temporal correlation. Block-based temporal interpolation (TI) approaches (e.g. BiMESS [6], MCTI [3]) cannot handle fast motion and produce block artifacts. The later issue is solved by *Pixel-Based Temporal Interpolation* (PBTI, [4]). The challenging fast motion was faced in [7] by applying a global motion model. But, the encoder complexity is increased in [7] due to model estimation. Furthermore, mesh-based TI [8] adapt well to deformation.

On the other hand, fast motion is also very challenging for temporal extrapolation (TX) approaches [5]. Whereas, the TX approach shows good results in revelation areas compared to TI.

The aim of our proposed method is to cope with fast camera motion. Therefore, the strengths of TI and TX are combined as well as a robust global motion estimation is applied. Furthermore, the encoder should stay at very low complexity, by applying the proposed algorithms only at the decoder side.

3. PROPOSED GLOBAL MOTION GUIDED ADAPTIVE TEMPORAL INTER-/EXTRAPOLATION

Our proposed *Global Motion Guided Adaptive Temporal Inter-/Extrapolation* (**GTIE**) method is applied in the pixel domain DVC scheme as shown in figure 1. This codec is chosen due to its very low encoding complexity. Furthermore, the *Global Motion Guided Pixel-Based Adaptive Temporal Inter-/Extrapolation* (**GpTIE**) algorithm is proposed, which significantly reduces block artifacts.

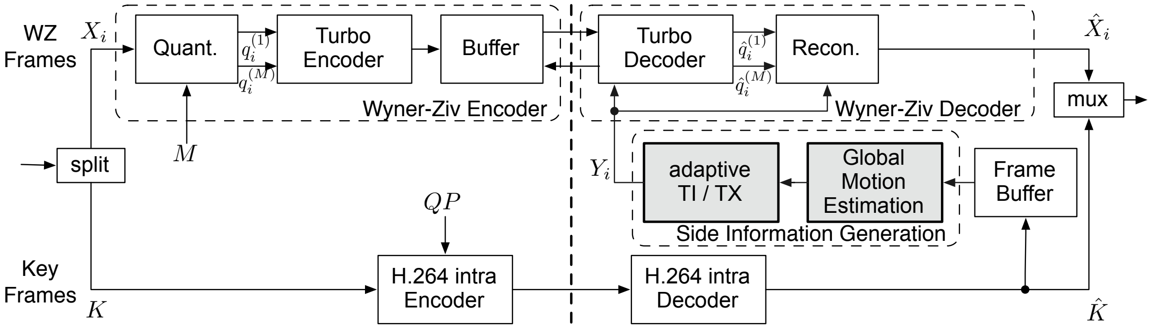
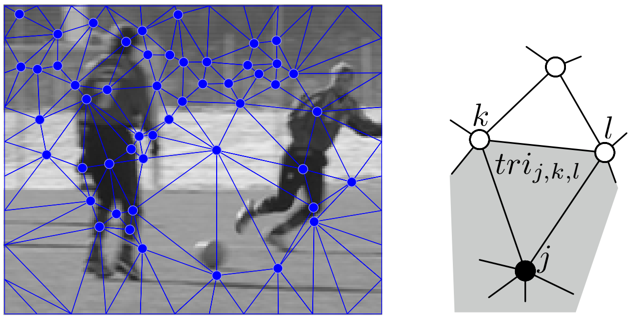


Fig. 1. Proposed pixel domain DVC scheme and *Global Motion Guided Adaptive Temporal Inter-/Extrapolation* (GTIE/GpTIE)



(a) feature point-based triangle mesh T

(b) gray surround area

Fig. 2. Surround area of each feature point build by a triangle mesh

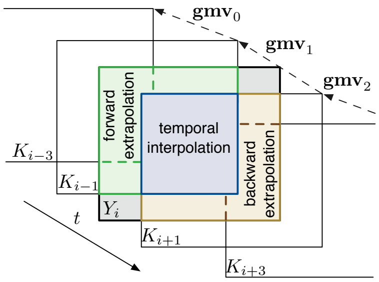


Fig. 3. Mode decision pattern (TI, for-/backward TX)

The proposed GTIE and GpTIE algorithms are split up into two steps:

1. estimation of global motion vectors $\mathbf{gmv}_{0...2}$ between the four adjacent key frames (fig. 3, sec. 3.1)
2. adaptive TI / TX mode decision and guided block/pixel-based temporal inter-/extrapolation (GTIE/GpTIE)

The block- and pixel-based temporal inter-/extrapolation steps are guided by the extracted global motion vectors. Thus, fast motion is less challenging. Furthermore, the revelation and occlusion problem is solved by an adaptive switching between TI and back-/forward TX. Finally, GpTIE reduces blocking artifacts by estimation of a dense motion vector field (one motion vector for each pixel).

3.1. Global motion estimation

The global motion is estimated based on four adjacent key frames mandatory for TI and for-/backward TX. Incorporation of more key frames is not meaningful due to low temporal correlation. The global motion is approximated by a two parameter model ($x' = x + \Delta x$, $y' = y + \Delta y$), which reflects the translative motion. This model having only two parameters is robust compared to higher order models.

The model parameters are estimated based on the feature points fp_i generated by the SIFT algorithm (Scale-Invariant Feature Transform, [9]). Each feature point pair (fp_i, fp'_i) of adjacent key frames represents a motion vector $\mathbf{mv}_i$.

The global motion vectors central characteristic is to represent the motion of the major area of the frame ($\mathbf{gmv} = (\Delta x, \Delta y)$). Thus, a weighted vector median filter (eq. 1) is applied. The weights (eq. 2) are calculated from the surround area of each feature point pair. The surround area is given by the area of all adjacent triangles $tri_{j,k,l}$ of a feature point fp_j (fig. 2(b)). The triangle mesh T (fig. 2(a)) is constructed by Delaunay triangulation.

The adopted weighted vector median filter assures robust estimation of the global motion vector, which is represented by the major region of the frame.

$$\sum_j w_j ||\mathbf{gmv} - \mathbf{mv}_j||_2 \leq \sum_j w_j ||\mathbf{mv}_i - \mathbf{mv}_j||_2 \qquad (1)$$

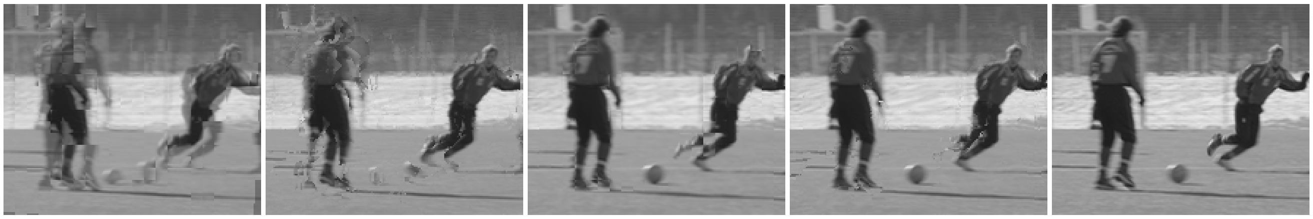
$$w_j = \sum_{tri_{i,k,l} \in T'_j} \text{area}(tri_{i,k,l}) \qquad (2)$$

where $T'_j = \{tri_{s,t,u} | tri_{s,t,u} \in T \wedge j \in \{s,t,u\}\}$ is the set of adjacent triangles of the feature point fp_j.

3.2. Adaptive temporal inter- / extrapolation

3.2.1. Mode decision

The mode for each 8×8 block (GTIE) respectively each pixel (GpTIE) is decided, based on the global motion vectors $\mathbf{gmv}_{0...2}$ (fig. 3). For this reason, the overlap of the global motion compensated key frames and the WZ frame is analyzed. Temporal interpolation is performed in the region, where adjacent key frames (K_{i-1}, K_{i+1}) intersect. Furthermore, temporal forward or backward extrapolation is chosen for regions visible in two preceding (K_{i-3}, K_{i-1}) or successive key frames (K_{i+1}, K_{i+3}). Temporal extrapolation is also chosen, if the region is only visible in one adjacent key frame. In this case, the TX is dominated by the global motion vector. Remaining areas, not visible in any key frame, are estimated by averaging interpolation and extrapolation results.



(a) BiMESS 19.6 dB (b) PBTI 20.5 dB (c) **GTIE 26.2 dB** (d) **GpTIE 26.6 dB** (e) Original

Fig. 4. Soccer side information example, WZ-frame 76, QCIF, 30 fps

3.2.2. *Temporal inter- / extrapolation (GTIE)*

An algorithm very close to BiMESS [6] is chosen for **temporal interpolation** (TI). Forward (eq. 3) and subsequent bidirectional motion estimation (eq. 4, half-pixel accuracy) is performed for each 8×8 block. The global motion vector $\mathbf{gmv}_1$ is incorporated as initial value. Hence, the motion estimation is a refinement based on the global motion vector. Finally, the interpolation result is carried out by bidirectional motion compensation.

$$\begin{aligned}(\delta x_0, \delta y_0) \quad = \quad & \arg\min_{\delta x_0, \delta y_0} \sum_{(x,y)\in MB} |K_{i-1}(x,y) \\ & -K_{i+1}(x+\delta x_0, y+\delta y_0)| \qquad (3)\end{aligned}$$

where MB is the Matching Block. The search interval for the motion vector $(\delta x_0, \delta y_0)$ is given by $-SR_0+\Delta x \leq \delta x_0 \leq SR_0+\Delta x$; $-SR_0 + \Delta y \leq \delta y_0 \leq SR_0 + \Delta y$. The maximum search range is $SR_0 = 10$. Thus, the search interval is guided by the corresponding global motion vector $\mathbf{gmv}_1 = (\Delta x, \Delta y)$.

$$\begin{aligned}(\delta x, \delta y) \quad = \quad & \arg\min_{\delta x, \delta y} \sum_{(x,y)\in MB} |K_{i-1}(x-\delta x, y-\delta y) \\ & -K_{i+1}(x+\delta x, y+\delta y)| \qquad (4)\end{aligned}$$

The subsequent bidirectional motion estimation (eq. 4) is guided by the motion vector $(\delta x_0, \delta y_0)$ extracted in the forward motion estimation step. Therefore, the search interval is given by $-SR+\delta x_0/2 \leq \delta x \leq SR+\delta x_0/2$; $-SR+\delta y_0/2 \leq \delta y \leq SR+\delta y_0/2$. The maximum search range is given by $SR = 1$.

In case of forward **temporal extrapolation** (TX), the motion vector field is obtained by unidirectional motion estimation from K_{i-1} to K_{i-3}. The global motion vector $-\mathbf{gmv}_0$ is treated as initial value. Subsequently, motion compensation is performed based on key frame K_{i-1}. Backward TX is applied in the same way, though using key frames K_{i+1}, K_{i+3} and global motion vector $\mathbf{gmv}_2$.

3.2.3. *Pixel-based temporal inter- / extrapolation (GpTIE)*

An extension of PBTI [4] is applied for **temporal interpolation** (TI) in GpTIE. Unidirectional motion estimation is performed for each pixel by matching its weighted neighborhood (Gaussian Window GW, eq. 5). Subsequent motion compensation is also performed on a pixel-basis by incorporating both adjacent key frames K_{i-1}, K_{i+1}.

$$\begin{aligned}(\delta x, \delta y) \quad = \quad & \arg\min_{\delta x, \delta y} \sum_{(x,y)\in MW} GW(x,y) \times |K_{i-1}(x,y) \\ & -K_{i+1}(x+\delta x, y+\delta y)| \qquad (5)\end{aligned}$$

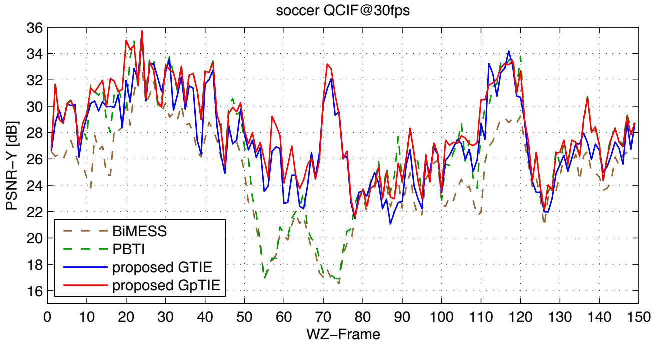


Fig. 5. Side information quality, Soccer, QCIF@30 fps

where MW is the Matching Window (neighborhood) of the current pixel. The motion vector $(\delta x, \delta y)$ is searched in the interval $-SR + \Delta x \leq \delta x \leq SR + \Delta x$; $-SR + \Delta y \leq \delta y \leq SR + \Delta y$ with a maximum search range of $SR = 10$. Thus, pixel-based unidirectional motion estimation is guided by the global motion vector $\mathbf{gmv}_1 = (\Delta x, \Delta y)$.

Unidirectional guided motion estimation is also performed for back-/forward **temporal extrapolation** (TX) applying the corresponding global motion vector $(-\mathbf{gmv}_0, \mathbf{gmv}_2)$. The subsequent motion compensation incorporates only one adjacent key frame.

Finally, pixel-based bidirectional motion estimation/compensation is applied to remove uncovered areas.

4. SIMULATION RESULTS

Simulation results were carried out for the QCIF sequences Coastguard, Foreman, Soccer and Stefan at 15 fps as well as 30 fps and a GOP size of 2. Key frames are H.264/AVC intra encoded.

4.1. Side information quality

An increased side information quality is expected especially for sequences with high camera motion (e.g. Soccer, Stefan). Figure 4 shows the estimated side information (76th WZ frame, 30 fps, Soccer). The proposed GTIE (c) and GpTIE (d) algorithm outperform the BiMESS (a) and PBTI (b) side information quality visually and by PSNR. One should note, that the frame quality at the right border is significantly higher (fig. 4(c),(d)). Backward TX is chosen here due to the right panning camera. GTIE and GpTIE show high gain especially for frames with high camera motion (fig. 5, frames 50 . . . 80). GTIE commonly outperforms BiMESS but not PBTI in case of slow global motion. Comparing the average SI

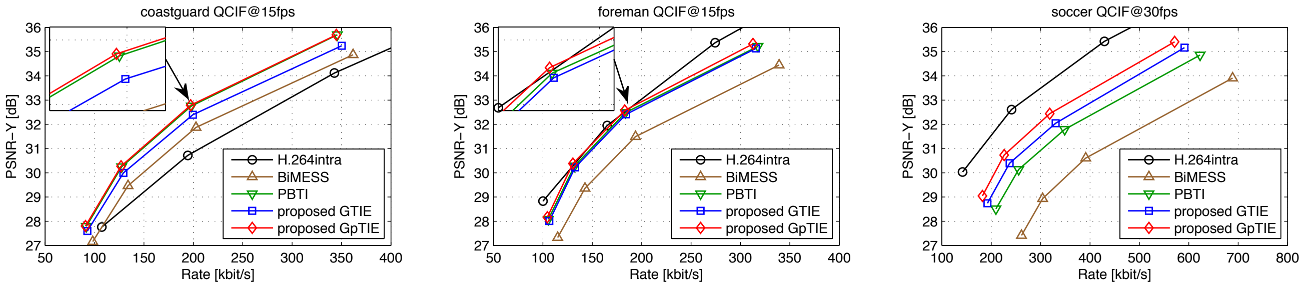


Fig. 6. RD performance, quantization parameter: $(QP, M) = \{(40, 1), (35, 1), (30, 1), (25, 2)\}$

QCIF 15 fps	BiMESS [6]	PBTI [4]	MCTI [7]	**GTIE**	**GpTIE**
Coastguard	30.3	32.4	31.4	**31.3**	**32.5**
Foreman	28.0	**29.8**	29.3	**29.3**	**29.7**
Soccer	21.0	21.4	22.1	**22.4**	**22.8**
Stefan	21.7	22.7	–	**23.4**	**23.6**
QCIF 30 fps	BiMESS [6]	PBTI [4]	MI-2 [5]	**GTIE**	**GpTIE**
Coastguard	36.3	**39.1**	37.5	**37.7**	**38.8**
Foreman	33.6	**36.8**	34.9	**35.2**	**36.5**
Soccer	24.8	26.9	–	**27.4**	**28.3**
Stefan	25.3	27.3	–	**27.8**	**28.6**

Table 1. Average side information quality, PSNR [dB]

quality (tab. 1), GTIE and GpTIE outperform state-of-the-art side information generation at the decoder for sequences with fast motion. Only PBTI performs slightly better than GTIE and GpTIE for low motion sequences and small key frame distance (QCIF@30fps).

4.2. Computational decoding complexity

An advantage of the GTIE is its low computational complexity compared to PBTI and GpTIE. The execution time of GTIE for one frame is 3.0 s (QCIF, Matlab M-Code, 1 core, 2.66GHz Xeon). PBTI and GpTIE are far more complex with 9.0 s and 13.6 s, in contrast to BiMESS with 1.1 s. The extra execution time of GTIE compared to BiMESS is consumed by the global motion estimation and additional extrapolation steps. At least, the execution time can be significantly reduced by a native implementation of these algorithms.

4.3. Overall RD performance

The proposed GpTIE outperforms all mentioned algorithm (fig. 6). The highest gain of 3.1 dB is achieved for the sequence Soccer. Furthermore, the less complex GTIE method outperforms BiMESS for all sequences. It also outperforms PBTI for the sequence Soccer, due to the fast motion characteristic of Soccer. GTIE does not perform better than PBTI for slow motion sequences (e.g. Coastguard and Foreman) due to its block basis. Finally, the proposed GTIE approach shows similar RD performance compared to PBTI at a significantly lower decoding complexity. If decoding complexity is not critical, GpTIE is the first choice. The RD performance of our proposed DVC coding scheme is very close to the RD performance of H.264intra for Foreman (QCIF, 15fps) – which is remarkable for a pixel domain DVC coding scheme.

5. CONCLUSIONS

Two novel global motion guided adaptive temporal inter-/extrapolation methods are proposed (GTIE, GpTIE). The proposed algorithms incorporate global motion estimation, at the decoder, to switch between temporal interpolation (TI) and back-/forward extrapolation (TX) as well as guide the subsequent motion estimation. This approach significantly increases the performance in case of fast motion. The proposed GpTIE outperforms state-of-the-art side information generation at the decoder. Finally, the overall RD performance is increased up to 3.1 dB (Soccer).

6. REFERENCES

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, pp. 471–480, 1973.

[2] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *Information Theory, IEEE Transactions on*, vol. 22, pp. 1–10, 1976.

[3] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The discover codec: Architecture, techniques and evaluation," in *Proc. Picture Coding Symposium (PCS)*, 2007.

[4] S. Sofke, R. Hänsel, and E. Müller, "Human visual system aware decoding strategies for distributed video coding," in *Proc. Picture Coding Symposium (PCS)*, 2009.

[5] S. Borchert, R.P. Westerlaken, R.K. Gunnewiek, and R.L. Lagendijk, "On extrapolating side information in distributed video coding," in *Proc. Picture Coding Symposium (PCS)*, 2007.

[6] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proc. Speech and Image Processing, Multimedia Communications and Services, Conference on (EC-SIP-M, EURASIP)*, 2005.

[7] F. Dufaux and T. Ebrahimi, "Encoder and decoder side global and local motion estimation for distributed video coding," in *Proc. Multimedia and Signal Processing (MMSP), IEEE Workshop. on*, 2010.

[8] D. Kubasov and C. Guillemot, "Mesh-based motion-compensated interpolation for side information extraction in distributed video coding," in *Proc. Image Processing (ICIP), IEEE Int. Conf. on*, 2006.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.