# Improved Adaptive Temporal Inter-/Extrapolation Schemes for Distributed Video Coding

Ralph Hänsel, Erika Müller
University of Rostock, Germany
Institute of Communications Engeneering
{ralph.haensel,erika.mueller}@uni-rostock.de

***Abstract*—Distributed video coding (DVC) gained a lot of research interest in the last decade, due to its high potential for emerging application scenarios. The RD-performance of a DVC coding scheme is mainly influenced by the side information (SI) quality. It is commonly generated by temporal interpolation (TI). TI fails in areas of occlusion or revelation. Therefore, we propose an Adaptive Pixel-based Temporal Interpolation (APBTI) scheme. It adaptively switches between TI and forward-/backward temporal extrapolation. The additionally proposed APBTI2 applies motion vector field extrapolation to solve the problem for non-linear motion. Simulation results show that the SI quality as well as the overall RD-performance is improved by up to $1.2\,\mathrm{dB}$ and $0.5\,\mathrm{dB}$ respectively.**

## I. Introduction

Video coding is an essential element in modern multimedia application scenarios. High compression ratios and low decoding complexity is significant for mobile video retrieval. The conventional video coding standards e.g. MPEG2, H.264 fit well for the mentioned scenario. They are designed for broadcasting scenarios, where low decoding complexity is demanded. Furthermore, the encoding complexity does not have high significance.

Distributed video coding (DVC) aims at low encoding complexity beside some further potential application scenarios [1]. Low encoding complexity is desired for mobile video encoding on devices with limited resources (e.g. processing power, battery life). Basically, the low encoding complexity is achieved by shifting the complex motion estimation algorithm from the encoder to the decoder side. In DVC no motion compensated prediction at the encoder is applied, whereas motion compensated temporal inter-/extrapolation is performed at the decoder to obtain the side information. The theories of Wyner and Ziv [2] as well as Slepian and Wolf [3] build the foundation for DVC by exploiting correlation on the decoder side.

The side information (SI) is an estimation of the current Wyner-Ziv (WZ) frame. A high SI quality results in a higher reconstruction quality and lower coding rate. The side information generation typically exploits temporal correlation, thus temporal inter- or extrapolation is applied (sec. II).

A challenge in temporal inter- and extrapolation are occlusion and revelation areas. In conventional video coding this problem is solved by introducing multiple reference frames (B-frames). Valid temporal interpolation in DVC is only carried out if all areas are visible in both adjacent frames. If revelation or occlusion occur, the performance of the motion estimation and compensation is decreased. The SI may show some ghosting artifacts or blur in that case (sec. III).

In this paper, we propose two methods to solve the challenge of occlusion and revelation for temporal interpolation in DVC (sec. IV). The first approach applies adaptive switching between temporal interpolation as well as forward and backward extrapolation (APBTI). The second uses motion field spatial extrapolation and adaptive motion compensation (APBTI2).

Simulation results (sec. V) show that the side information quality is improved by up to $1.2\,\mathrm{dB}$ and the overall coding performance is improved by up to $0.5\,\mathrm{dB}$. Finally, conclusions and remarks on further work are given in section VI.

## II. Side Information Generation for DVC

The side information quality has strong impact on the rate distortion (RD) performance of the DVC coding scheme. Therefore, the temporal correlation is typically exploited for SI generation. One popular temporal interpolation algorithm is BiMESS [4], which incorporates forward motion estimation (FME), bidirectional motion estimation (BiME) and spatial motion field smoothing at a fixed block size. The powerful MCTI [5] extends the concept of BiMESS by incorporating different block sizes in a hierarchical manner.

The reduction of blocking artifacts is one focus of PBTI [6], which incorporates a dense motion vector field (MVF) for proper temporal interpolation. Global motion estimation (GME) on the decoder side and dense MVF estimation was incorporated in GpTIE [7] to cope with fast camera motion.

Besides temporal interpolation (TI), the temporal extrapolation (TX) was used in [8]. The side information generation by TX compared to TI is typically less accurate. Whereas, TX enables the coding scheme to use larger key frame distances by incorporating the decoded Wyner-Ziv (WZ) frames.

Side information generation for video sequences having very inhomogeneous and non-linear motion is challenging for the methods mentioned above. Hash-based SI generation algorithm solve that problem by transmitting a small portion of the WZ frame by conventional coding methods. It is incorporated in motion estimation (ME) resulting in higher SI quality [9],[10].

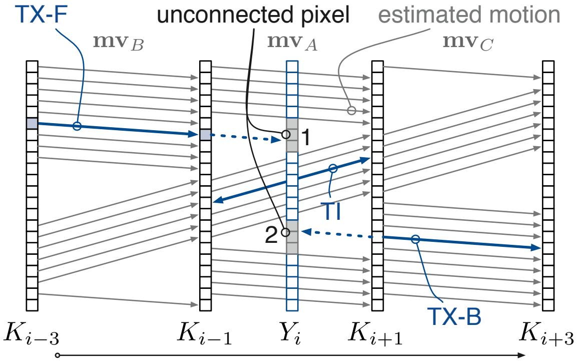


Fig. 1. Unconnected areas in $Y_i$ due to discontinuities the MVF, TI - temporal interpolation, TX-F/B - temporal for-/backward extrapolation

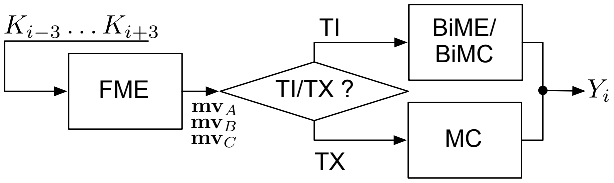


Fig. 2. Structure of Adaptive Pixel-based Temporal Interpolation (APBTI)

## III. Occlusions in Temporal Interpolation

Temporal interpolation works very good for sequences showing linear homogenous motion. TI is likely to produce ghosting artifacts and blur in non-homogenous motion areas, which occurs in case of revelation or occlusion. The reason for this behavior is shown in figure 1, where area 1 represents revelation and area 2 represents occlusion. Motion estimation does not work for occlusion nor revelation and some unconnected pixel are left. This area is only visible in the prior or posterior key frames ($K_{i-1}$, $K_{i+1}$). Temporal interpolation relies on the visibility in both adjacent frames. Typically, neighbor motion vector are filled in here, to solve that problem. The final motion compensation (MC) step merges the foreground and background object, because both key frames are incorporated, which leads to ghosting artifacts and blur.

## IV. Proposed Solution

The basic idea of our proposed solution is to use temporal extrapolation (TX) for unconnected areas as shown in figure 1. Therefore, the final motion compensation incorporates only the key frame, which contains the area to be estimated, avoiding ghosting or blur. In more detail, the proposed algorithm will apply temporal forward extrapolation (TX-F) for revelation (area 1) and temporal backward extrapolation (TX-B) for occlusion areas (area 2). The selection of the proper mode is the key challenge in our proposed *Adaptive Pixel-based Interpolation (APBTI)* scheme.

### *A. Adaptive Pixel-based Temporal Interpolation (APBTI)*

The first approach of Adaptive Pixel-based Temporal Interpolation (APBTI, fig. 2) incorporates four adjacent key frames $K_{i-3} \ldots K_{i+3}$. In the first step FME is performed to estimate the dense motion vector fields $\mathbf{mv}_A, \mathbf{mv}_B, \mathbf{mv}_C$. Therefore, a weighted SAD (sum of absolute difference) matching approach is used (eq. 1).

$$\begin{aligned}\mathbf{mv} \;=\; & \arg\min_{mv_x, mv_y} \sum_{(x,y)\in MW} GW(x,y) \times |K_{i-1}(x,y) \\ & -K_{i+1}(x+mv_x, y+mv_y)| \qquad (1)\end{aligned}$$

where $GW$ is the Gaussian and $MW$ is the matching window.

For each pixel in the SI $Y_i$ the decision between TI, TX-F and TX-B is made based on the mininum corresponding matching error $SAD_A, SAD_B, SAD_C$ (eq. 2). An extra penalty $\delta$ is added for the extrapolation modes, so it is more likely to select the interpolation mode. The TI mode is stronger than the TX modes and thus TX is only chosen if a significant improvement is expected.

$$mode \;=\; \arg\min(\{SAD_A; SAD_B+\delta; SAD_C+\delta\}) \qquad (2)$$

Subsequently, bidirectional motion estimation (BiME) and compensation (BiMC) is performed for all pixels in TI mode. Whereas, unidirectional motion compensation (MC) incorporating the corresponding key frame ($K_{i-1}$ or $K_{i+1}$) is applied for the TX modes.

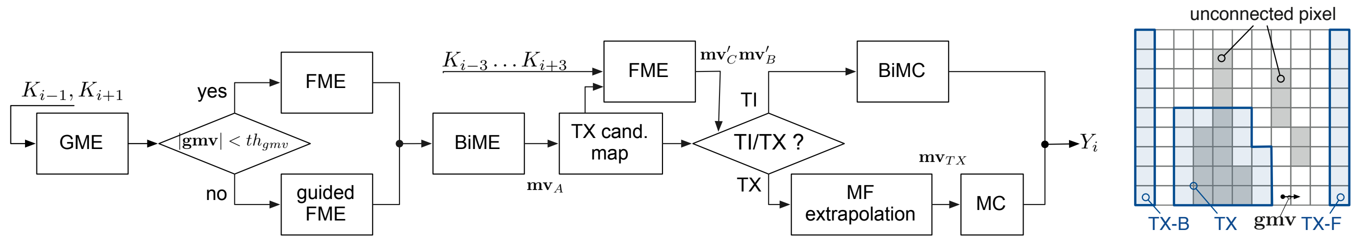
The APBTI method improves the SI quality in areas with non-homogenous motion, which is shown in figure 4(c).

### *B. Adaptive Pixel-based Temporal Interpolation 2 (APBTI2)*

Linear motion is assumed for all 4 adjacent key frames in the APBTI method. This is the major handicap of the first proposed method. APBTI2 aims to solve that problem by extending ABPTI an applying spatial motion field extrapolation. Thus, MVFs $\mathbf{mv}_B$, $\mathbf{mv}_C$ are not used for extrapolation. Furthermore, a global motion estimation (GME, [7]) is incorporated to improve performance for fast camera motion scenes and to identify potential extrapolation areas.

The structure of APBTI2 is shown in figure 3(a). In the first step the global motion vector $\mathbf{gmv}$ is estimated. If the $\mathbf{gmv}$ is insignificant ($|\mathbf{gmv}| < th_{gmv}$) a basic forward motion estimation (FME) is performed. In case of significant global motion a multiple hypotheses FME guided by $\mathbf{gmv}$ is performed, applying a reduced search range (20% reduction) to level computing complexity. Afterwards, MVF refinement by bidirectional motion estimation (BiME) is performed.

The TX is only applied if occlusion or revelation occur. Therefore, a map indicating candidates for TX is generated based on discontinuities in the MVF (fig. 3(b)). Strong camera motion leads to occlusion and revelation at the frame border [7]. Hence, pixels at the frame border are added to the TX candidate map according to the $\mathbf{gmv}$ (fig. 3(b)). An area of unconnected pixels is added to TX candidate map only if the width is greater than one pixel. Hence, small areas of unconnected pixels are not taken into account for TX, because they are the result of a gradient in the MVF and not of occlusion nor revelation. Finally, the TX candidate regions



(a)

(b)

Fig. 3. (a) Structure of Adaptive Pixel-based Temporal Interpolation 2 (APBTI2) and
(b) Candidate TX areas selected based on unconnected pixels (discontinuities in MVF) and global motion $\mathbf{gmv} = (1, 0)$

are extended by a one pixel border, because TI may not work well close to unconnected areas.

For all TX candidate pixels the MVFs $\mathbf{mv}'_B$ and/or $\mathbf{mv}'_C$ are estimated incorporating the outer key frames ($K_{i-3}$, $K_{i+3}$). The estimated MVFs are not used for temporal extrapolation due to its low accuracy in case of non-linear motion. Only, the matching error $SAD$ is used for mode decision, as shown in equation 2, if the pixel is in the candidate map.

The estimated motion vectors $\mathbf{mv}'_B$, $\mathbf{mv}'_C$ are used for MF extrapolation as an initial value for the motion of a pixel in TX-F or TX-B mode. Motion field extrapolation is performed in an iterative algorithm. The $7 \times 7$ neighborhood of a pixel in TX mode is selected. If the neighborhood contains more than nine valid motion vectors (TI mode, or already filled in TX mode), the motion vector for TX $\mathbf{mv}_{TX}$ is selected from the set of MV in the neighborhood $\mathbf{mv}_{NH}$. Therefore, equation 3 is used for the TX-F mode ($\mathbf{mv}'_B$ is replaced by $\mathbf{mv}'_C$ for TX-B mode).

$$\mathbf{mv}_{TX} \;=\; \arg\min_{\mathbf{mv}_i} \left[\mathrm{d}(\mathbf{mv}'_B, \mathbf{mv}_i) + \mathrm{d}(\overline{\mathbf{mv}_{NH}}, \mathbf{mv}_i)\right] \quad (3)$$

where $\mathrm{d}(\ldots)$ is the Euclidian distance and $\mathbf{mv}_i \in \mathbf{mv}_{NH}$.

A motion vector $\mathbf{mv}_{TX}$ close to the initial value $\mathbf{mv}'_B$ or $\mathbf{mv}'_C$ respectively as well as close to the mean motion of the neighborhood is selected. This cancels out non-linear motion and smoothes the MF extrapolation result. This process is repeated until all motion vectors of TX mode pixels are filled in.

At least, bidirectional MC is performed for TI pixels incorporating $\mathbf{mv}_A$ and unidirectional MC for TX pixels incorporating $\mathbf{mv}_{TX}$.

## V. Simulation Results

Simulation results on side information quality and RD-performance are carried out for CIF, SIF (NTSC) and QCIF resolution sequences. A maximum search range of 16 pixels ($32 \times 32$ window) is applied for MCTI. Whereas, the search range for GpTIE, APBTI and APBTI2 is set to 10 pixels for QCIF and 15 pixels for SIF and CIF sequences. ME and MC are performed on half-pixel accuracy.

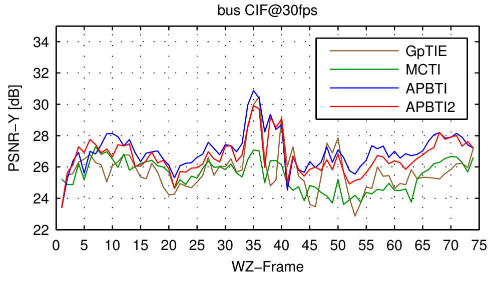


Fig. 5. Side information quality, bus, CIF, 30 fps

### A. Side Information Quality

The side information quality incorporating key frames coded with H.264intra at QP25 is shown in figure 4. This frame of the sequence bus shows a lot of occlusion and revelation areas. Especially details on the left side of the bus are well visible in the side information generated by APBTI and APBTI2. The proposed methods also show good results in terms of PSNR (fig. 5). Only GpTIE outperforms APBTI for some frames of the sequence bus.

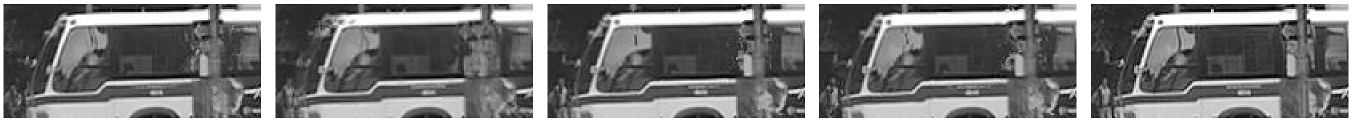
The proposed algorithm APBTI and APBTI2 also show good results for sequences which are less influenced by occlusion or revelation (table I). The proposed MVF extrapolation is the main reason why APBTI2 shows significantly better results than APBTI for common sequences with non-linear motion.

### B. RD Performance

The proposed methods APBTI and APBTI2 show the best overall RD-performance for the sequence bus in a pixel-domain DVC coding scheme (+0.47 dB comp. to H.264intra, fig. 6). Furthermore, APBTI2 performs equal or better in a pixel-domain DVC compared to GpTIE or MCTI for a common set of sequences (tab. II).

## VI. Conclusion

Two side information generation algorithms for DVC were proposed, where the problem of temporal interpolation in occlusion and revelation areas was faced. On the one hand, adaptive temporal inter-/extrapolation is designed for sequences with very linear motion (APBTI, up to +1.2 dB SI quality). On the other hand, the secondly proposed method APBTI2 based on motion field extrapolation and adaptive MC shows better results for a general set of sequences.



(a) GpTIE - 26.01 dB (b) MCTI - 25.84 dB (c) **APBTI** - 27.42 dB (d) **APBTI2** - 27.32 dB (e) original

Fig. 4. Visual side information quality, `bus`, CIF, 30 fps, cropped WZ-frame 30, PSNR of full frame, key frames: H.264intra, QP25

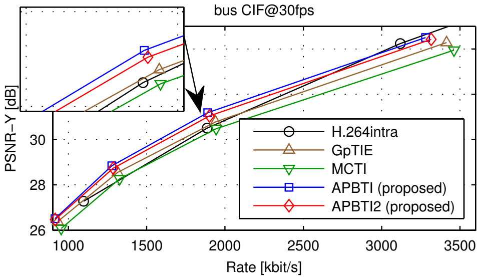


Fig. 6. RD-performance, `bus`, CIF, 30 fps, GOP = 2

TABLE I
AVERAGE SIDE INFORMATION QUALITY, PSNR-Y [DB], KEY FRAMES: H.264INTRA, QP25

| | GpTIE [7] | MCTI [5] | **APBTI** proposed | **APBTI2** proposed |
|---|---|---|---|---|
| CIF 30 fps | | | | |
| `bus` | 25.8 | 25.5 | **27.0** | **26.7** |
| SIF 30 fps | | | | |
| `sflowg` | 28.8 | 27.4 | 28.2 | 28.5 |
| QCIF 15 fps | | | | |
| `coastguard` | 32.1 | 30.0 | 31.9 | **32.2** |
| `foreman` | 29.4 | 28.5 | 28.4 | **29.5** |
| `soccer` | 22.8 | 22.3 | 20.8 | 22.4 |
| `stefan` | 23.6 | 22.9 | 22.1 | **23.9** |

TABLE II
RD-PERFORMANCE PSNR-Y GAIN [DB] AND RATE DIFFERENCE [%] (BD [11]) OF APBTI AND APBTI2 COMPARED TO H.264INTRA AS WELL AS GPTIE [7] AND MCTI [5] IN A PIXEL-DOMAIN DVC, GOP=2

| | APBTI | | APBTI2 | |
|---|---|---|---|---|
| | ΔPSNR | Δrate | ΔPSNR | Δrate |
| `bus`, CIF, 30 fps | | | | |
| H.264intra | 0.47 dB | −7.8% | 0.31 dB | −5.5% |
| GpTIE | 0.50 dB | −7.5% | 0.34 dB | −5.3% |
| MCTI | 0.86 dB | −12.7% | 0.70 dB | −10.6% |
| `sflowg`, SIF, 30 fps | | | | |
| H.264intra | 3.40 dB | −32.7% | 3.57 dB | −34.0% |
| GpTIE | −0.34 dB | 3.7% | −0.17 dB | 1.7% |
| MCTI | 0.02 dB | 0.1% | 0.16 dB | −1.8% |
| `coastguard`, QCIF, 15 fps | | | | |
| H.264intra | 1.66 dB | −25.6% | 1.8 dB | −27.2% |
| GpTIE | −0.12 dB | 2.0% | 0.02 dB | −0.2% |
| MCTI | 0.86 dB | −13.7% | 1.0 dB | −15.6% |
| `foreman`, QCIF, 15 fps | | | | |
| H.264intra | −0.93 dB | 15.0% | −0.27 dB | 3.9% |
| GpTIE | −0.62 dB | 10.0% | 0.04 dB | −0.6% |
| MCTI | 0.04 dB | −1.0% | 0.70 dB | −10.6% |
| `soccer`, QCIF, 15 fps | | | | |
| H.264intra | −6.34 dB | 234.2% | −5.10 dB | 177.6% |
| GpTIE | −1.37 dB | 23.5% | −0.13 dB | 2.6% |
| MCTI | −0.77 dB | 10.4% | 0.47 dB | −8.3% |
| `stefan`, QCIF, 15 fps | | | | |
| H.264intra | −2.18 dB | 30.9% | −0.86 dB | 11.2% |
| GpTIE | −1.16 dB | 15.5% | 0.16 dB | −2.0% |
| MCTI | −0.59 dB | 7.3% | 0.74 dB | −8.9% |

The proposed APBTI and APBTI2 improves the side information quality for most sequences by up 0.9 dB. Furthermore, the overall RD-performance is improved by up to 0.50 dB (APBTI) and 0.34 dB (APBTI2) respectively.

Further work may include an additional MVF refinement step based on the decoded bit planes as well as adaptive MC. Furthermore, the motion field extrapolation should include depth estimation results. Knowledge of the depth of objects helps to find the covered object. Therefore, monocular depth estimation [12] based on motion and scene structure might be adopted.

## REFERENCES

[1] F. Dufaux, W. Gao, S. Tubaro, and A. Vetro, "Distributed Video Coding: Trends and Perspectives," *EURASIP Journal on Image and Video Proc.*, vol. 2009, p. 13, 2009.

[2] A. Wyner and J. Ziv, "The Rate-distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. on Information Theory*, vol. 22, pp. 1–10, 1976.

[3] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. on Information Theory*, vol. 19, pp. 471–480, 1973.

[4] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," in *Proc. Conf. on Speech and Image Proc., Multimedia Comm. and Services (EURASIP)*, 2005.

[5] J. Ascenso and F. Pereira, "Hierarchical Motion Estimation for Side Information Creation in Wyner-Ziv Video Coding," in *Proc. Int. Conf. on Ubiquitous Information Management and Comm. (ICUIMC)*, 2008.

[6] S. Sofke, R. Hänsel, and E. Müller, "Human Visual System Aware Decoding Stratgies for Distributed Video Coding," in *Proc. Picture Coding Symposium (PCS)*, 2009.

[7] R. Hänsel and E. Müller, "Global Motion Guided Adaptive Temporal Inter-/Extrapolation for Side Information Generation in Distributed Video Coding," in *Proc. Int. Conf. on Image Proc. (ICIP)*, 2011.

[8] S. Borchert, R. Westerlaken, R. Gunnewiek, and R. Lagendijk, "On Extrapolating Side Information in Distributed Video Coding," in *Proc. Picture Coding Symposium (PCS)*, 2007.

[9] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv Video Coding with Hash-based Motion Compensation at the Receiver," in *Proc. IEEE Int. Conf. on Image Proc. (ICIP)*, 2004.

[10] N. Deligiannis, F. Verbist, J. Barbarien, J. Slowack, R. V. de Walle, P. Schelkens, and A. Munteanu, "Distributed Coding of Endoscopic Video," in *Proc. IEEE Int. Conf. on Image Proc. (ICIP)*, 2011.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-curves," Telenor Satellite Services, Tech. Rep. VCEG-M33, 2001.

[12] E. Meinhardt-Llopis, O. D'Hondt, G. Facciolo, and V. Caselles, "Relative Depth from Monocular Optical Flow," in *Proc. Int. Conf. on Image Proc. (ICIP)*, 2011.