Fading Techniques for Error Concealment in Block-Based Video Decoding Systems

Markus Friebe and André Kaup, *Senior Member, IEEE*

***Abstract*—Error concealment has a great importance within block-based video decoding systems especially for mobile transmission systems like DVB-T or DVB-H. In this paper, we propose a spatio-temporal and spatio-bi-temporal fading technique for recovering lost or erroneously received macroblock samples. For spatio-temporal fading based on a boundary error criterion obtained from temporal error concealment, either spatial, temporal, or fading of both methods is used for recovering lost image samples in one macroblock. A weighted absolute difference between well received macroblock boundary samples from the current frame and motion compensated macroblock boundary samples from the previous frame represents the boundary error. The weights for fading are interpolated from the boundary error. For spatio-bi-temporal fading, one lost macroblock can individually be recovered pel-wise spatially from the current or bi-temporally from the previous and the future frame by weighted averaging three different estimations for the lost macroblock. The simulation results have shown that for concealing erroneously received uncompressed and compressed video data the proposed methods outperform the reference methods both in subjective and objective video quality.**

***Index Terms*—Error concealment, video coding and transmission, video restoration.**

I. Introduction

In CASE video signals are transmitted over mobile networks like DVB-T, DVB-H or other systems, error concealment is very important. When transmitting block-based coded video over error-prone networks, packets may be lost in case of transmission errors. Due to entropy coding, image information is lost until a synchronization marker is reached. Here, visually annoying block artifacts are introduced in decoded video frames. Restoration techniques are either usually spatial to conceal Intra-coded frames, temporal or spatio-temporal to conceal Inter-coded frames.

The method frequency selective signal extrapolation [1] is a spatial error concealment method and uses 2D-DFT basis function for approximating surrounding error-free received image samples. The area, which corresponds to the lost macroblock, is copied to the lost image area. H.264 Intra [2] is using surrounding error-free or concealed image samples for spatial bilinear interpolation.

The Boundary Matching Algorithm (BMA) [3] recovers a motion vector of the lost macroblock by using the information of surrounding error-free received motion vectors. This is a temporal technique. A temporal replacement algorithm [4] uses zero motion vectors (Zero MV) to conceal a lost macroblock. Image samples of the same macroblock within the previous frame are copied into the lost area. The Decoder Motion-Vector Estimation Algorithm (DMVE) [5] is a temporal technique and uses image samples, which are immediately neighbored to the lost macroblock. Using these neighbored image samples, the DMVE [5] is looking for the best match in the previous frame. A temporal recursive block-matching (RBM) algorithm [6] uses a two-step block-matching principle and additionally recursion steps for temporal error concealment.

A content-based adaptive spatio-temporal method (CABLR) [7] is using temporal image information for macroblock loss recovery, if the temporal information fits well. Otherwise, correctly received or concealed spatial neighboring macroblocks are used to recover a lost macroblock. Finally, a range constraint is applied on the spatially recovered macroblock. A 3D-deblocking method (3D-DE) [8] is a spatio-temporal technique which first recovers temporally a lost macroblock and further uses a modified deblocking filter [9] for spatial filtering.

The Decoder Motion-Vector Estimation Bidirectional Algorithm (DMVE-BiDir) [5] is a bi-temporal error concealment method and is looking for the best match in the previous and the future frame. A low complexity motion compensated frame interpolation method (OBMC) [10] is used for macroblock interpolation. Overlapped block bi-temporal motion estimation estimates lost motion vectors. A motion vector post-processing eliminates bad estimated motion vectors and overlapped block motion compensation is recovering a lost macroblock.

Frequency selective signal extrapolation is a spatial method and is introducing blurred image areas. H.264 Intra [2] introduces also blurred image areas because of spatial bilinear interpolation. Zero MV [4], BMA [3], DMVE [5], and RBM [6] are all temporal methods and introduce block artifacts in case of object occlusions and uncovering. CABLR conceals either temporal or spatial. Through this switching scheme, either blurred image areas or block artifacts are visible. Also 3D-DE [8] generates some artifacts in case of object occlusions and uncovering. DMVE-BiDir [5] is also introducing some block artifacts in case of object occlusions and uncovering. Blurred image areas along macroblock borders are introduced by OBMC [10] in case of discontinuity of surrounding motion vectors. An overview of different error concealment techniques is given in [11].

In this contribution, we propose a spatio-temporal fading scheme, which can be used for block loss recovery in uncompressed, Intra- and Predictive-coded frames. This fading scheme

Manuscript received June 30, 2006; revised November 11, 2006.

The authors are with the the Department of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, 91058 Erlangen, Germany.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2006.889688

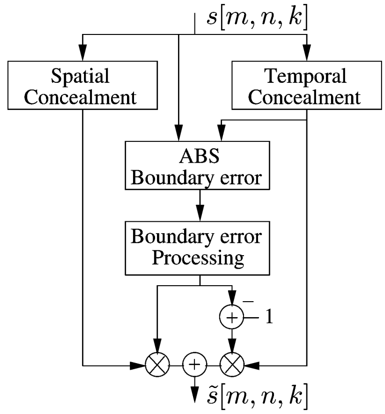

Fig. 1. Block diagram of the spatio-temporal fading scheme.

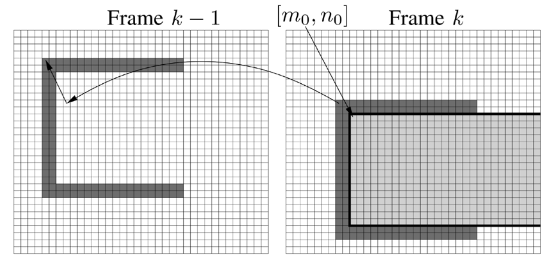

Fig. 2. Left: Frame $k-1$. Right: Frame k. The dark-gray marked area represents the $SAD_w[mv_m, mv_n]$ evaluation area with $w[m,n] = 1$ and the light-gray area represents the lost image samples with $w[m,n] = 0$, both for the left most consecutive macroblock loss.

combines the result obtained from spatial with that obtained from temporal error concealment. Using this fading scheme, block artifacts can be reduced while keeping the reconstructed image area relatively sharp. The proposed spatio-temporal fading scheme will be further described in Section II. Further, we propose a spatio-bi-temporal method for concealing Bidirectional-coded frames. Using image information obtained from both temporal directions, appearing artifacts from object occlusions and uncovering can be further reduced. Different to the spatio-bi-temporal method in [12] compressed video signals are considered. The spatio-bi-temporal fading scheme is further described in Section III. Simulation results are given in Section IV with conclusions following in Section V.

II. Spatio-Temporal Fading Scheme

Fig. 1 shows a block diagram of the spatio-temporal fading scheme (ST-FS). First temporal and spatial concealment estimate the lost image samples in one macroblock. Descriptions about these methods are given in Sections II-A and II-B. Based on the result of temporal concealment, a boundary error is computed. A post-processing unit computes a weighting matrix for the missing macroblock. Using these weighting matrix, pel-wise weighted mean values between spatial and temporal error concealment results are computed. This post-processing unit is further described in Section II-C.

A. Temporal Error Concealment

The motion vector of the lost macroblock with its vertical $\tilde{mv}_m$ and horizontal $\tilde{mv}_n$ components is estimated similar to DMVE [5] by minimizing a weighted sum of absolute differences $SAD_w[mv_m, mv_n]$. Eqs. (1) and (2) show how $SAD_w[mv_m, mv_n]$ is calculated and how the motion vector is estimated from the minimum $SAD_w[mv_m, mv_n]$. Image samples s in the current frame k are compared to motion compensated image samples from the previous frame $k-1$. A weighting function represents correct ($w[m,n] = 1$) and incorrect ($w[m,n] = 0$) received image samples in the current frame k. In Fig. 2, the dark-gray marked area represents correct and the light-gray area the incorrect received image samples for the left most consecutive macroblock loss. For the right most consecutive macroblock loss, the dark-gray marked area is flipped in horizontal direction. The small arrow in Fig. 2 shows one of the evaluated motion vectors. The size of a lost macroblock is given by M and N in row and column direction respectively. m_0 and n_0 represents the left top image sample in the lost macroblock.

$$SAD_w[mv_m, mv_n] = \sum_{m=m_0-2}^{m_0+M+1} \sum_{n=n_0-2}^{n_0+N+1} w[m,n] \cdot |s[m,n,k] - s[m+mv_m, n+mv_n, k-1]| \quad (1)$$

and

$$[\tilde{mv}_m, \tilde{mv}_n] = \arg \min_{mv_m, mv_n} (SAD_w[mv_m, mv_n]) \quad (2)$$

The difference between DMVE [5] and this method is the weighting function. DMVE [5] treats concealed areas as correctly received image samples by setting $w[m,n] = 1$. In this approach, concealed areas are lower weighted by setting $w[m,n]$ between 0.2 and 0.5. For these values, the best simulation results have been achieved. After estimating the lost motion vector, the temporal error concealment result represents the image samples

$$s_{temp}[m,n] = s[m_0 + m + \tilde{mv}_m, n_0 + n + \tilde{mv}_n, k-1] \quad (3)$$

of the previous frame $k-1$ with $m = 0, \ldots, M-1 \wedge n = 0, \ldots, N-1$.

B. Spatial Error Concealment

For spatial error concealment, a frequency selective signal extrapolation method [1] is used for block loss recovery. This method is extrapolating image samples from surrounding well decoded image samples to the lost macroblock. The known surrounding image samples within a specific area $\mathcal{A}$ are approximated by a set of 2D-DFT basis functions. Such basis functions $\varphi_{u,v}[m,n]$ are selected and scaled by $c_{u,v}$ to minimize an error criterion. The error criterion is computed between the sum of the selected basis functions $g[m,n]$ and the surrounding well received image samples within the specific area. u and v are indices for basis functions in row and column direction.

$$g[m,n] = \sum_{(u,v) \in \mathcal{A}} c_{u,v} \varphi_{u,v}[m,n] \quad (4)$$

This is an iterative method and in each iteration step, one 2D-DFT basis function is selected and scaled. The iteration stops, if the error criterion is smaller than a given threshold or a maximum number of iterations is reached. The area $\mathcal{L}$

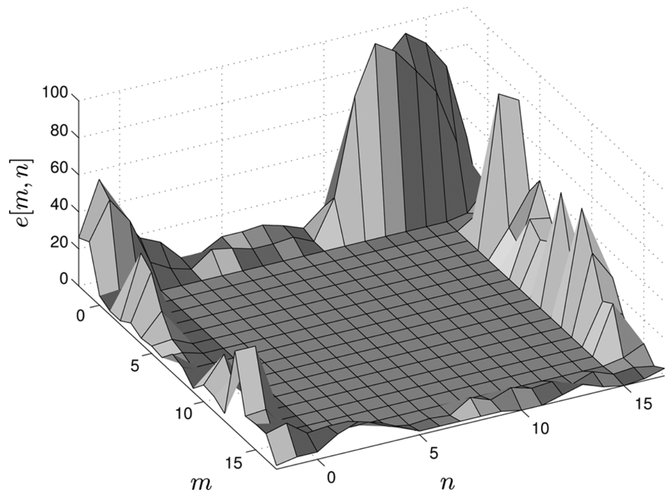

Fig. 3. Boundary error $e[m, n]$ for a 16×16 block loss.

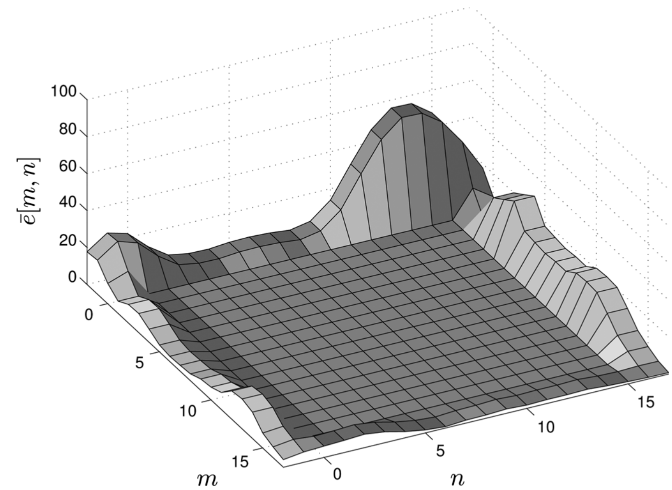

Fig. 4. Mean boundary error $\bar{e}[m, n]$.

within $g[m, n]$, which represents the lost macroblock, is used for spatial error concealment.

$$s_{spatial}[m, m] = g[i, j] \tag{5}$$

with $m = 0, \ldots, M-1 \wedge n = 0, \ldots, N-1 \wedge i, j \in \mathcal{L}$. This method is used because of good extrapolation results. Basically, any other spatial prediction method can be used.

C. Boundary Error Processing and Interpolation

In Fig. 2 the area where the boundary error $e[m, n]$ is computed is represented by the weighted absolute difference between the dark-gray marked area in frame k and the motion compensated dark-gray marked area in frame $k-1$. The weights $w[m, n]$ for computing the boundary error are the same as in Section II-A. The boundary error measures pel-wise, if temporal error concealment fits well to the lost macroblock area.

$$e[m,n] = w[m,n] \cdot |s[m,n,k] - s[m+\tilde{mv}_m, n+\tilde{mv}_n, k-1]| \tag{6}$$

with $m = m_0-2, \ldots, m_0+M+1 \wedge n = n_0-2, \ldots, m_0+N+1$. In Fig. 3 an example for the boundary error $e[m, n]$ is shown. At the bottom macroblock border temporal concealment fits relatively well, but not at the top right macroblock corner. After computing the boundary error, this error is lowpass filtered. For lowpass filtering a mean boundary error $\bar{e}[m, n]$ within a 3×3 window along the dark-gray marked area in Fig. 2 is computed.

$$\bar{e}[m,n] = \frac{1}{\sum_{i=-1}^{1} \sum_{j=-1}^{1} w[m-i, n-j]} \cdot \sum_{i=-1}^{1} \sum_{j=-1}^{1} e[m-i, n-j] \tag{7}$$

Symmetrically boundary extensions are used for filtering. The mean boundary error $\bar{e}[m, n]$ can be seen in Fig. 4. The area $\mathcal{L}$ represents the area $(m = 0, \ldots, M-1 \wedge n = 0, \ldots, N-1)$ encircled by the lowpass filtered boundary error. $\mathcal{L}$ is then bilinear interpolated from the nearest $\bar{e}[m, n]$ neighbors, outside of $\mathcal{L}$, in vertical and horizontal direction. The distances to the nearest neighbors represent the weights for bilinear interpolation.

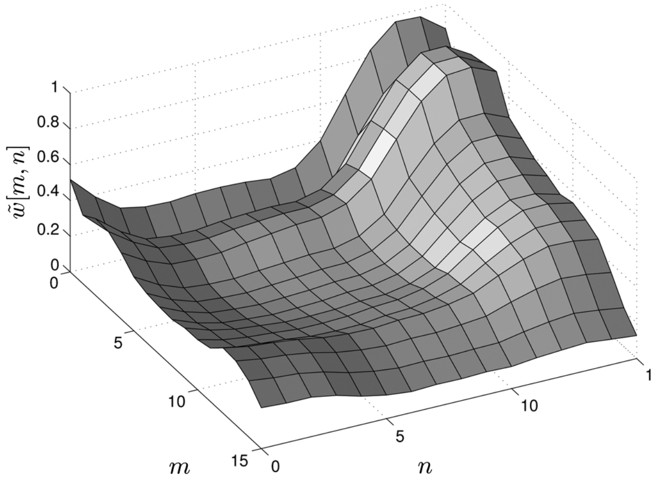

Fig. 5. Weighting matrix $\tilde{w}[m, n]$.

Temporal error concealment is only acceptable in case the boundary error is not to high and therefore the resulting matrix $\tilde{w}[m, n]$ of size $M \times N$ obtained from bilinear interpolation is compared to a threshold T_w and scaled by T_w.

$$\tilde{w}[m,n] = \begin{cases} 1, & \tilde{w}[m,n] > T_w \\ \frac{\tilde{w}[m,n]}{T_w}, & \text{else} \end{cases} \tag{8}$$

In Fig. 5 the weighting matrix $\tilde{w}[m, n]$ for the boundary error in Fig. 3 is shown. This weighting matrix is used for pel-wise fading between spatial and temporal error concealment result. At the bottom macroblock border, the weights are relatively low and temporal concealment is higher weighted than spatial. Here, temporal concealment fits relatively well. Whereas, at the top right macroblock corner, temporal concealment does not fit well, the weights are relatively high and spatial concealment is stronger weighted. Finally, the lost macroblock $\tilde{s}[m, n]$ is estimated from spatial $s_{spatial}[m, n]$ and temporal $s_{temp}[m, n]$ error concealment by

$$\tilde{s}[m,n] = (1 - \tilde{w}[m,n]) \cdot s_{temp}[m,n] + \tilde{w}[m,n] \cdot s_{spatial}[m,n]. \tag{9}$$

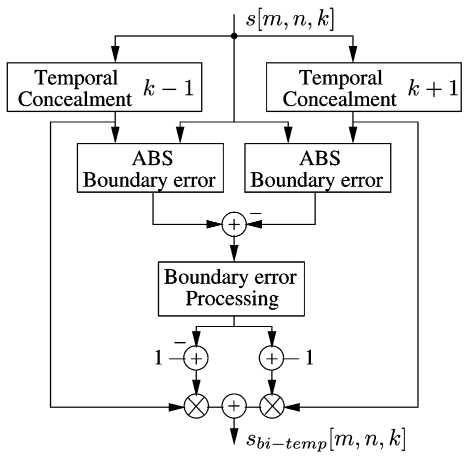

Fig. 6. Block diagram of the bi-temporal fading scheme.

$\tilde{s}[m,n]$ is estimated using a pel-wise weighted mean value between spatial $s_{spatial}[m,n]$ and temporal $s_{temp}[m,n]$ error concealment. In Fig. 1 pel-wise weighted averaging is done by the subtraction, multiplication and addition unit. Lowpass filtering in (7) is necessary to avoid switching between spatial and temporal error concealment for neighboring reconstructed image samples $\tilde{s}[m,n]$.

III. Spatio-Bi-Temporal Fading Scheme

The spatio-bi-temporal fading scheme (SBT-FS) consists of two reconstructing parts. The first part is reestimating a lost macroblock and his macroblock boundary samples temporally from the previous and the future frame. Every reconstructed image sample represents a weighted mean value between the corresponding motion compensated image sample in the previous and the future frame. A weighting matrix is individually averaging each reestimated image sample. This is further described in Section III-A. The second part is recovering a lost macroblock spatio-bi-temporally. The result obtained from bi-temporal fading scheme is used for temporal concealment. Based on a boundary error obtained from bi-temporal fading scheme, spatial, bi-temporal, or fading between both methods is used for recovering lost image samples in one macroblock. The weights for fading are interpolated from the boundary error. A weighted absolute difference between well received macroblock boundary samples from the current frame and bi-temporal boundary samples from the previous and the future frame represent the boundary error. This is further described in Section III-B.

A. *First Reconstructing Part: Bi-Temporal Reconstruction*

A block diagram of the bi-temporal fading scheme is shown in Fig. 6. For bi-temporal reconstruction, temporal error concealment uses the previous frame $k-1$ and the future frame $k+1$. For both directions, a motion vector is estimated. The motion vector of the lost macroblock $\tilde{mv}_m^p$ and $\tilde{mv}_n^p$ regarding the previous frame is estimated like in Section II-A. Eqs. (10) and (11) show how $SAD_w^p[mv_m^p, mv_n^p]$ is calculated and how the motion vector is estimated from the minimum $SAD_w^p[mv_m^p, mv_n^p]$. Hence we calculate the weighted sum of absolute differences by

$$SAD_w^p\left[mv_m^p, mv_n^p\right] = \sum_{m=m_0-2}^{m_0+M+1} \sum_{n=n_0-2}^{n_0+N+1} w[m,n] \cdot \left|s[m,n,k] - s\left[m+mv_m^p, n+mv_n^p, k-1\right]\right| \quad (10)$$

and the best matching motion vector by

$$\left[\tilde{mv}_m^p, \tilde{mv}_n^p\right] = \arg \min_{mv_m^p, mv_n^p} \left(SAD_w^p\left[mv_m^p, mv_n^p\right]\right). \quad (11)$$

The motion vector of the lost macroblock $\tilde{mv}_m^f$ and $\tilde{mv}_n^f$ regarding the future frame is estimated similar to the estimation regarding the previous frame. Instead of $k-1$, mv_m^p, mv_n^p and SAD_w^p we use $k+1$, mv_m^f, mv_n^f and SAD_w^f in (10). The motion vector is estimated from the minimum $SAD_w^f[mv_m^f, mv_n^f]$ by

$$\left[\tilde{mv}_m^f, \tilde{mv}_n^f\right] = \arg \min_{mv_m^f, mv_n^f} \left(SAD_w^f\left[mv_m^f, mv_n^f\right]\right). \quad (12)$$

After estimating the lost motion vectors, the temporal concealment from the previous frame $k-1$ represents the motion compensated image samples

$$s_{temp}^p[m,n,k] = s\left[m+\tilde{mv}_m^p, n+\tilde{mv}_n^p, k-1\right] \quad (13)$$

and from the future frame $k+1$ the motion compensated image samples

$$s_{temp}^f[m,n,k] = s\left[m+\tilde{mv}_m^f, n+\tilde{mv}_n^f, k+1\right] \quad (14)$$

with $m = m_0-2,\ldots,m_0+M+1 \wedge n = n_0-2,\ldots,n_0+N+1$.

In Fig. 2 the weighted absolute difference between the area marked in dark-gray in frame k and the motion compensated area marked in dark-gray in frame $k-1$ represents the area where the boundary error $e^p[m,n]$ is computed. The weights $w[m,n]$ for computing the boundary error are the same as for estimating the motion vectors. The boundary error measures pel-wise, if temporal error concealment fits well to the lost macroblock area. Based on $s_{temp}^p[m,n,k]$ and $s_{temp}^f[m,n,k]$, two absolute boundary errors for the previous frame

$$e^p[m,n] = w[m,n] \cdot \left|s[m,n,k] - s_{temp}^p[m,n,k]\right| \quad (15)$$

and for the future frame

$$e^f[m,n] = w[m,n] \cdot \left|s[m,n,k] - s_{temp}^f[m,n,k]\right| \quad (16)$$

are computed for evaluating the fitness to the lost area in the current frame k. An example for the boundary error $e^p[m,n]$ for the previous and $e^f[m,n]$ for the future frame is shown in Figs. 7 and 8. At the bottom left macroblock corner temporal concealment from the future frame does not fit very well, because the error is relatively high. Temporal concealment from the previous

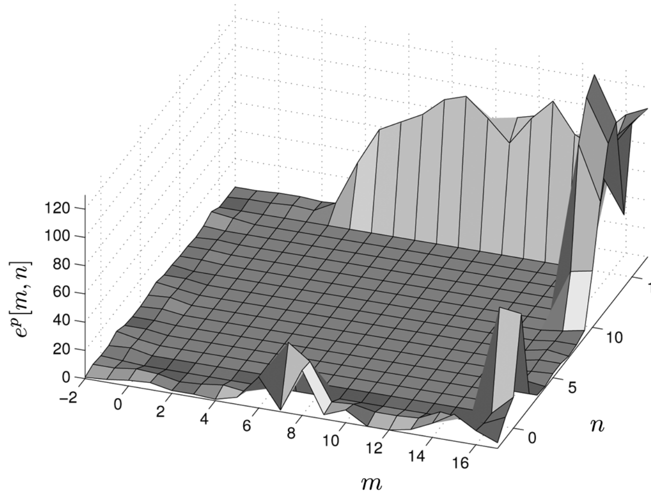

Fig. 7. Boundary error $e^p[m,n]$.

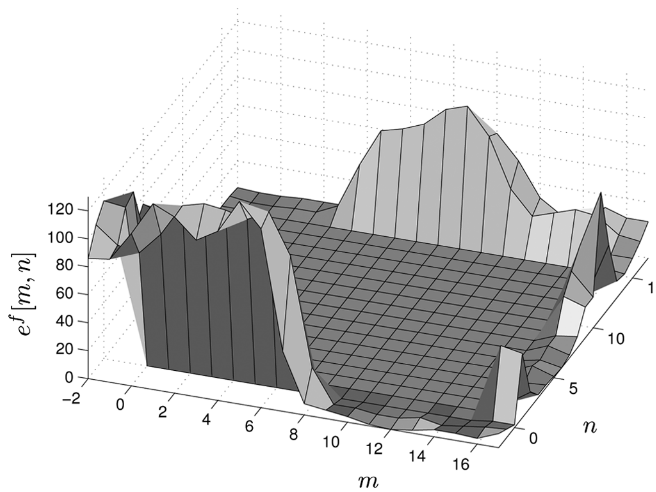

Fig. 8. Boundary error $e^f[m,n]$.

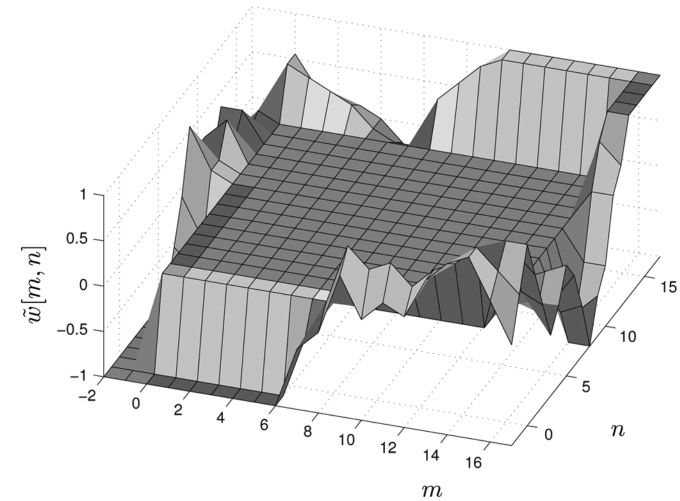

Fig. 9. Lowpass filtered decision signal $\tilde{w}[m,n]$.

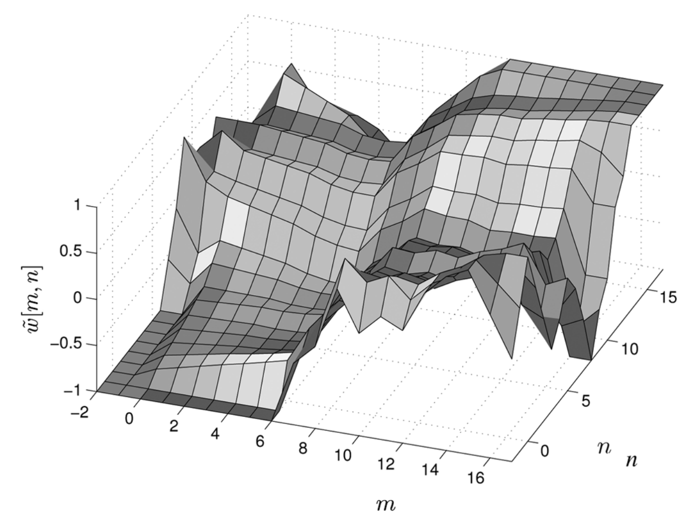

Fig. 10. Bilinearly interpolated lowpass filtered decision signal $\tilde{w}[m,n]$.

frame fits relatively well. After computing both boundary errors, a gradient

$$\bigtriangledown[m,n] = e^p[m,n] - e^f[m,n] \tag{17}$$

between these boundary errors is computed. In the block boundary error processing the $sign$ of $\bigtriangledown[m,n]$ is evaluated, to find pel-wise if temporal concealment from the previous frame fits better than concealment from the future frame.

$$\bigtriangledown_{sign}[m,n] = sign\left(\bigtriangledown[m,n]\right) \tag{18}$$

$\bigtriangledown_{sign}[m,n]$ represents the decision signal. A lowpass filter computes a mean decision signal within a 3×3 window along the corresponding boundary samples.

$$\tilde{w}[m,n] = \frac{1}{\sum_{i=-1}^{1}\sum_{j=-1}^{1} w[m-i,n-j]} \cdot \sum_{i=-1}^{1}\sum_{j=-1}^{1} \bigtriangledown_{sign}[m-i,n-j] \tag{19}$$

For filtering symmetrically boundary extensions are used. The lowpass filtered decision signal $\tilde{w}[m,n]$ can be seen in Fig. 9. The area $\mathcal{L}$ represents the lost macroblock $(m = 0,\ldots,M-1 \wedge n = 0,\ldots,N-1)$ encircled by the lowpass filtered decision signal $\tilde{w}[m,n]$. $\mathcal{L}$ is then bilinearly interpolated from the nearest $\tilde{w}[m,n]$ neighbors, outside of $\mathcal{L}$, in vertical and horizontal direction. The distances to the nearest neighbors represent the weights for bilinear interpolation. The resulting matrix $\tilde{w}[m,n]$ can be seen in Fig. 10 and is used for pel-wise bi-temporal interpolation. Finally, the result for bi-temporal concealment $s_{bi-temp}[m,n,k]$ is estimated from the previous $s^p_{temp}[m,n,k]$ and the future frame $s^f_{temp}[m,n,k]$ by

$$s_{bi-temp}[m,n,k] = \frac{1}{2}\cdot\Big((1-\tilde{w}[m,n])\cdot s^p_{temp}[m,n,k] + (1+\tilde{w}[m,n])\cdot s^f_{temp}[m,n,k]\Big). \tag{20}$$

$s_{bi-temp}[m,n,k]$ is estimated using a pel-wise weighted mean value between $s^p_{temp}[m,n,k]$ and $s^f_{temp}[m,n,k]$. Lowpass filtering in (19) is necessary to avoid switching between temporal concealment from the previous and the future frame for

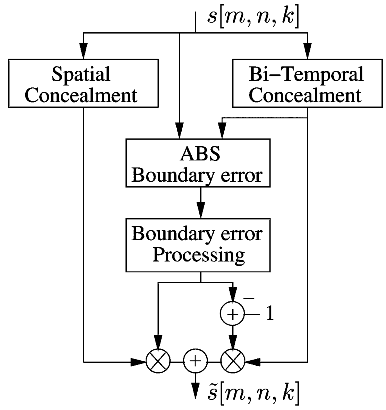

Fig. 11. Block diagram of the spatio-bi-temporal fading scheme.

neighboring reconstructed image samples $s_{bi-temp}[m,n,k]$. The example in Fig. 10 shows that at the bottom left macroblock corner, the weights are quite near to -1 and temporal concealment from the previous frame will be stronger weighted than from the future frame. Here, temporal concealment from the previous frame is better suited. Whereas, at the top right macroblock corner, temporal concealment from the future frame yields a lower boundary error. Here, the weights are quite near to 1 and temporal concealment from the future frame is stronger weighted.

B. Second Reconstructing Part: Spatio-Bi-Temporal Reconstruction

The spatio-temporal fading scheme uses instead of temporal error concealment the before mentioned bi-temporal error concealment as seen in Fig. 11. The area where the boundary error $e[m,n]$ is computed is represented by the weighted absolute difference between the area marked in dark-gray in frame k (Fig. 2) and the result obtained from bi-temporal error concealment. The weights $w[m,n]$ for computing the boundary error are the same as in Section II-A. The boundary error measures pel-wise, if bi-temporal error concealment fits well to the lost macroblock area.

$$e[m,n] = w[m,n] \cdot |s[m,n,k] - s_{bi-temp}[m,n,k]| \quad (21)$$

To reduce computational complexity, a bilinear interpolation method H.264 Intra [2] is used to obtain the spatially estimated image samples $s_{spatial}[m,n,k]$ for the lost macroblock. The description for the boundary error processing unit follows Section II-C. The weighting matrix $\tilde{w}[m,n]$ obtained from boundary error processing as seen in Fig. 5 is used for pel-wise weighted averaging between spatially $s_{spatial}[m,n,k]$ and bi-temporally $s_{bi-temp}[m,n,k]$ estimated lost macroblock samples.

$$\tilde{s}[m,n,k] = (1 - \tilde{w}[m,n]) \cdot s_{bi-temp}[m,n,k] + \tilde{w}[m,n] \cdot s_{spatial}[m,n,k]. \quad (22)$$

For macroblock regions where bi-temporal concealment fits well to the lost macroblock, bi-temporal error concealment is stronger weighted than spatial. For macroblock regions where bi-temporal error concealment does not fit well, spatial compensation is stronger weighted than bi-temporal.

Fig. 12. Left: Original image. Right: Consecutive 16×16 macroblock loss.

Fig. 13. Left: H.264 Intra [2] with $\mathrm{Y\ PSNR} = 16.68$ dB. Right: 3D-DE [8] with $\mathrm{Y\ PSNR} = 19.65$ dB.

IV. Simulation Results

For simulations, both uncompressed and compressed video signals are considered in the following sections. Using uncompressed video signals our proposed methods can be evaluated on signals without error propagation. Further compressed video signal are used for evaluating our proposed methods with the impact of error propagation. The proposed error concealment methods are subjectively and objectively compared to different reference methods.

A. Results for Uncompressed Video Signals

For uncompressed video frames, we consider videos of size CIF and 25 frames per second. Consecutive macroblocks are lost in every 5th video frame and in each YUV colorspace. In each third macroblock row, 18 consecutive macroblocks are lost starting with the third macroblock in one row. The colorspaces Y, U, and V are processed separately for all methods. The motion vectors are only estimated for the luminance and further used for the chrominances. If the luminance in CABLR is recovered spatially, then also the chrominances are recovered spatially. DMVE [5] and DMVE-BiDir [5] use both 2 line encirclement for computing SAD.

In Fig. 12 at the left hand side a frame of the sequence *Basketball* is shown without errors and with simulated consecutive macroblock loss at the right hand side. The result for H.264 Intra [2] is shown at the left hand side and for 3D-DE [8] at the right hand side in Fig. 13. H.264 Intra [2] is introducing blurred image areas because of the bilinear interpolation. 3D-DE [8] introduces some block artifacts in case of object occlusions and uncovering, as seen at the shoes from the basketball players. The result for the proposed spatio-temporal fading scheme

Fig. 14. Left: ST-FS with $\mathrm{Y\ PSNR} = 20.74\ \mathrm{dB}$. Right: DMVE-BiDir [5] with $\mathrm{Y\ PSNR} = 20.12\ \mathrm{dB}$.

Fig. 15. Left: OBMC [10] with $\mathrm{Y\ PSNR} = 21.65\ \mathrm{dB}$. Right: SBT-FS with $\mathrm{Y\ PSNR} = 23.06\ \mathrm{dB}$.

(ST-FS) is shown at the left hand side and for DMVE-BiDir [5] at the right hand side in Fig. 14. Also, ST-FS is introducing some block artifacts, as seen at the legs of the basketball player next to the camera. Block artifacts are also visible in the result for DMVE-BiDir [5] at the head of the basketball player next to the camera. At the left hand side in Fig. 15, the result for OBMC [10] and at the right hand side the result for the proposed spatio-bi-temporal fading scheme (SBT-FS) can be seen. Strong blurred image areas are visible between the black shoes next to the camera for OBMC [10]. For the proposed SBT-FS, the recovered areas look visually better than the recovered areas from the mentioned reference methods. Block artifacts are not visible and only a few reconstructed areas look blurry.

In Table I mean luminance PSNR results in dB for different videos and error concealment methods are shown. The PSNR values are only evaluated at the lost macroblock areas. Spatial methods like H.264 Intra [2] yields the lowest PSNR values. Temporal methods like DMVE [5] and spatio-temporal like CABLR, 3D-DE [8] and ST-FS perform much better than spatial methods. Bi-temporal methods do not always perform better than spatio-temporal methods as seen for the videos *Foreman* and *Basketball*. E.g. ST-FS yields higher PSNR results than DMVE-BiDir [5] for *Foreman* and *Basketball* and 3D-DE [8] better results than DMVE-BiDir [5] for *Foreman*. OBMC [10] achieves for all video sequences higher PSNR results than spatio-temporal methods. The proposed SBT-FS outperforms all other reference methods for all video sequences.

In Fig. 16 luminance PSNR results for evaluated *Basketball* video frames are shown. At the very beginning of the sequence up to the 7th evaluated frame, OBMC [10] and SBT-FS achieve quite similar results. In this part of the sequence, less motion is visible. From the 8th up to the 20th evaluated frame, motion increases and object occlusions and uncovering appear. In this part of the sequence, SBT-FS outperforms all other reference

TABLE I
MEAN Y PSNR RESULTS

Video	*Fore-man*	*Basket-ball*	*Sales-man*	*Flower garden*	*Coast-guard*
H.264 Intra [2]	21.99	15.93	22.26	15.44	17.74
DMVE [5]	32.75	21.16	32.60	26.51	28.46
CABLR [7]	33.15	20.58	32.63	25.86	27.68
3D-DE [8]	33.46	21.73	33.49	27.23	27.86
ST-FS	33.51	22.37	33.87	27.14	28.28
DMVE-BiDir [5]	33.43	21.81	32.89	27.42	29.40
OBMC [10]	34.63	23.52	35.42	30.93	30.08
SBT-FS	**36.10**	**24.38**	**35.57**	**31.45**	**31.97**

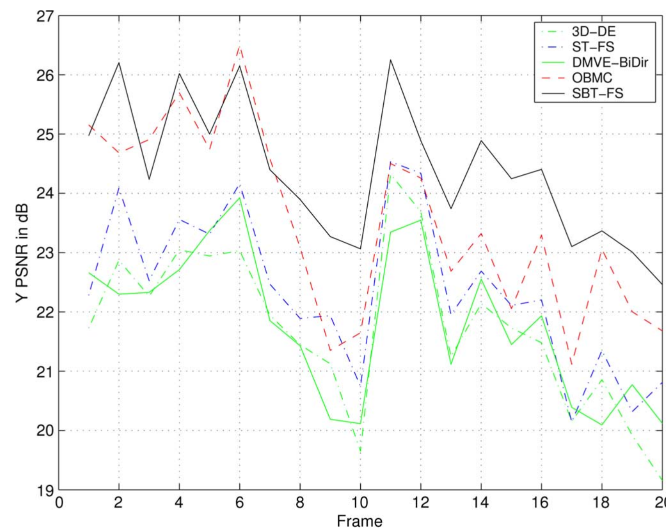

Fig. 16. PSNR results for evaluated *Basketball* frames.

methods because SBT-FS is able to recover additionally spatially to bi-temporal concealment in case bi-temporal error concealment fits not well to the lost macroblock area. For 16 out of 20 evaluated *Basketball* frames SBT-FS achieves the highest PSNR results.

B. Results for Compressed Video Signals

For compressed video sequences MPEG-2 coded video signals of size 720×576, 25 frames per second at 4 Mbit/s are considered. Group of picture size is 13 and B-frames are used. The packet size within the packetized elementary streams is 188 Byte. Packet errors with different packet error rate (PER) are introduced in coded video streams. Four different video sequences from the $VQEG$ group [13] with three different packet error rates are used in our simulations. For all methods, motion estimation is done for the luminance component and further used for the chrominance components. Each algorithm is used for concealing errors in I-, P- and B-frames. The method SBT-FS uses for I- and P-frames ST-FS and for B-frames the SBT-FS. ST-FS uses for compressed video data H.264 Intra [2] within the block spatial error concealment. A search range of ±96 image samples is used for motion estimation in all algorithms.

In Tables II–V mean Y PSNR results in dB for different video sequences and PER are shown. PSNR values are only computed for concealed video frames. For all PER of each video sequence, SBT-FS yields on average the highest and ST-FS the second highest PSNR results. In Table VI mean Y PSNR results for

TABLE II
MEAN Y PSNR RESULTS FOR $vqeg2$

PER	1E-3	3E-3	5E-3	Mean
Zero MV [4]	38.57	27.83	27.07	31.15
DMVE [5]	**41.68**	28.35	27.92	32.65
RBM [6]	40.90	28.25	28.04	32.40
3D-DE [8]	41.50	**28.35**	28.02	32.62
ST-FS	41.44	28.34	28.10	32.63
SBT-FS	41.51	28.33	**28.16**	**32.67**

TABLE III
MEAN Y PSNR RESULTS FOR $vqeg5$

PER	1E-3	3E-3	5E-3	Mean
Zero MV [4]	35.37	31.04	22.37	29.60
DMVE [5]	37.20	32.38	22.81	30.80
RBM [6]	37.46	33.22	**23.31**	31.33
3D-DE [8]	37.44	32.75	23.05	31.08
ST-FS	38.33	33.24	23.18	31.58
SBT-FS	**38.37**	**33.42**	23.24	**31.68**

TABLE IV
MEAN Y PSNR RESULTS FOR $vqeg6$

PER	1E-3	3E-3	5E-3	Mean
Zero MV [4]	35.16	30.34	23.70	29.73
DMVE [5]	37.36	**36.10**	25.43	32.96
RBM [6]	38.14	34.80	**26.52**	33.15
3D-DE [8]	40.01	35.15	25.61	33.59
ST-FS	40.51	35.20	25.84	33.85
SBT-FS	**40.64**	35.23	25.97	**33.95**

TABLE V
MEAN Y PSNR RESULTS FOR $vqeg7$

PER	1E-3	3E-3	5E-3	Mean
Zero MV [4]	33.46	27.44	22.93	27.94
DMVE [5]	39.12	**28.86**	22.49	30.16
RBM [6]	40.30	26.84	**23.29**	30.14
3D-DE [8]	39.48	28.16	22.69	30.11
ST-FS	**40.50**	28.34	23.03	30.62
SBT-FS	40.49	28.40	23.06	**30.65**

TABLE VI
MEAN Y PSNR RESULTS FOR ALL SEQUENCES AND PER

Method	Zero MV	DMVE	RBM	3D-DE	ST-FS	SBT-FS
PSNR	29.61	31.64	31.76	31.85	32.17	**32.24**

Fig. 17. Left: Original $vqeg5$ B-frame. Right: Erroneous PER 3E-3.

Fig. 18. Left: Zero MV [4] with $\mathrm{Y\ PSNR} = 26.14\ \mathrm{dB}$. Right: RBM [6] with $\mathrm{Y\ PSNR} = 26.53\ \mathrm{dB}$.

Fig. 19. Left: DMVE [5] with $\mathrm{Y\ PSNR} = 26.86\ \mathrm{dB}$. Right: 3D-DE [8] with $\mathrm{Y\ PSNR} = 27.04\ \mathrm{dB}$.

Fig. 20. Left: ST-FS with $\mathrm{Y\ PSNR} = 27.64\ \mathrm{dB}$. Right: SBT-FS with $\mathrm{Y\ PSNR} = 27.73\ \mathrm{dB}$.

all sequences and PER are shown. The proposed SBT-FS and ST-FS achieve on average the highest and the second highest PSNR results followed by 3D-DE [8].

In Fig. 17 at the left hand side one part of a B-frame of the sequence $vqeg5$ and at the right hand side the part of the B-frame including errors for PER 3E-3 is shown. In the middle image area, the small black marked regions represent errors from error propagation. In the lower image area, the big black marked region represents errors in the current frame. The results obtained from temporal error concealment methods Zero MV [4], RBM [6] and DMVE [5] can be seen in Fig. 18 at the left and right hand side and Fig. 19 at the left hand side. All temporal methods Zero MV [4], RBM [6] and DMVE [5] introduce strong block artifacts on some reconstructed regions. Spatio-temporal methods like 3D-DE [8] and the proposed ST-FS recovers both lost image regions better than the mentioned reference methods as seen in Fig. 19 at the right and in Fig. 20 at the left hand side. These methods additionally reduce appearing block artifacts on reconstructed regions. The result obtained from the proposed spatio-bi-temporal method SBT-FS can be seen in Fig. 20 at the right hand side. SBT-FS uses for B-frames both temporal directions for estimating the lost image regions. The concealed

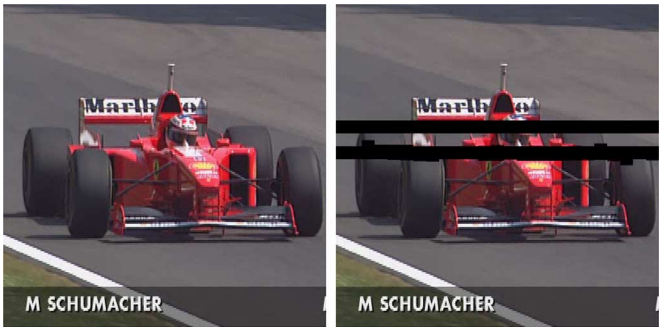

Fig. 21. Left: Original $vqeg6$ P-frame. Right: Erroneous PER 1E-3.

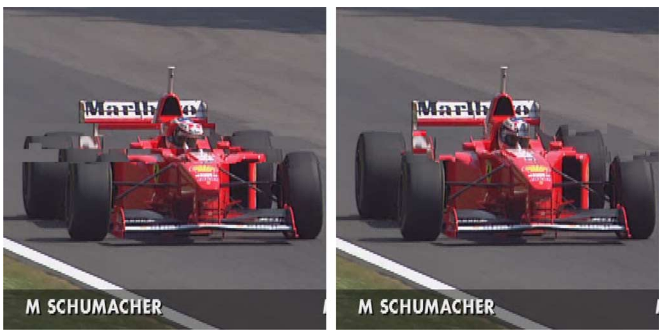

Fig. 22. Left: Zero MV [4] with $\mathrm{Y\ PSNR} = 31.08\ \mathrm{dB}$. Right: RBM [6] with $\mathrm{Y\ PSNR} = 35.94\ \mathrm{dB}$.

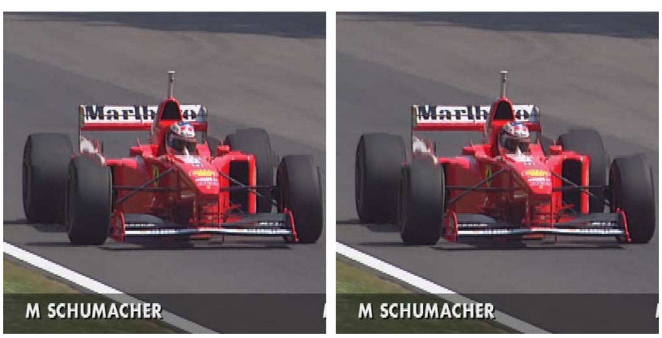

Fig. 23. Left: DMVE [5] with $\mathrm{Y\ PSNR} = 37.09\ \mathrm{dB}$. Right: 3D-DE [8] with $\mathrm{Y\ PSNR} = 38.31\ \mathrm{dB}$.

regions look visually better and the PSNR value for this B-frame is higher than the PSNR values for the reference methods. Also in areas where errors from error propagation appear, SBT-FS conceals these regions better than the reference methods.

In Fig. 21 at the left hand side one part of a P-frame of the sequence $vqeg6$ and at the right hand side the part of the P-frame including errors for PER 1E-3 is shown. The results for temporal error concealment methods Zero MV [4], RBM [6] and DMVE [5] can be seen in Fig. 22 at the left and right hand side and Fig. 23 at the left hand side. Also here introduce temporal methods annoying block artifacts. 3D-DE [8] can reduce appearing block artifacts as seen in Fig. 23 at the right hand side. The result for ST-FS is shown in Fig. 24 at the left hand side. ST-FS also reduces appearing block artifacts and the result looks similar to 3D-DE [8]. The proposed SBT-FS uses ST-FS for concealing P-frames and therefore is the result obtained from SBT-FS the same as for ST-FS which can be seen in Fig. 24 at the right hand side.

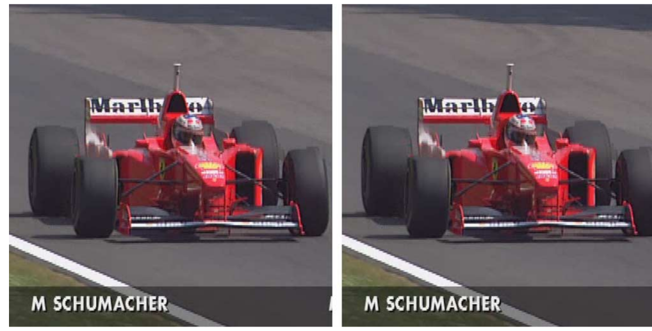

Fig. 24. Left: ST-FS with $\mathrm{Y\ PSNR} = 37.87\ \mathrm{dB}$. Right: SBT-FS with $\mathrm{Y\ PSNR} = 37.87\ \mathrm{dB}$.

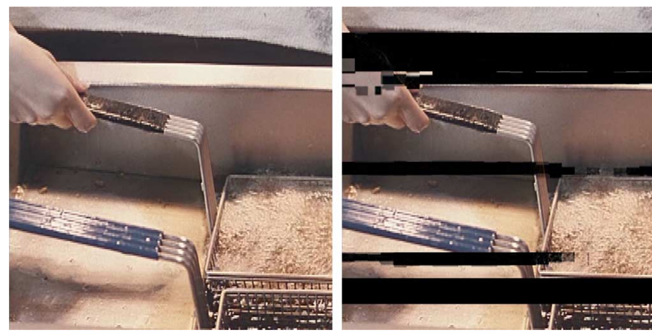

Fig. 25. Left: Original $vqeg7$ B-frame. Right: Erroneous PER 5E-3.

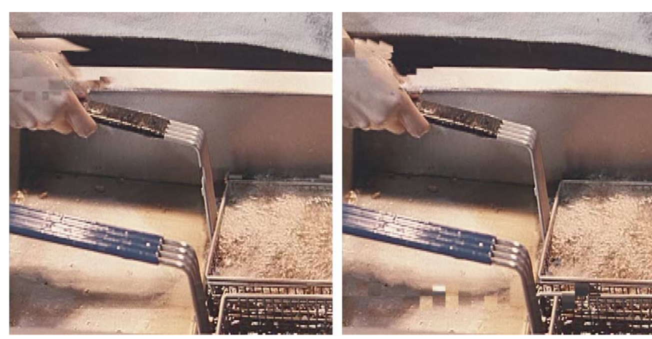

Fig. 26. Left: Zero MV [4] with $\mathrm{Y\ PSNR} = 26.93\ \mathrm{dB}$. Right: RBM [6] with $\mathrm{Y\ PSNR} = 25.62\ \mathrm{dB}$.

In Fig. 25 at the left hand side one part of a B-frame of the sequence $vqeg7$ and at the right hand side the part of the B-frame including errors for PER 5E-3 is shown. The results for Zero MV [4], RBM [6] and DMVE [5] can be seen in Fig. 26 at the left and right hand side and Fig. 27 at the left hand side. Using these methods block artifacts appear as seen at the hand or at the French fries basket. Zero MV [4] is recovering the French fries basket well because for this region no motion appears regarding the previous frame. In the result for 3D-DE [8] is shown in Fig. 27 at the right hand side. At the French fries basket, the motion estimation is estimating a wrong motion vector for some macroblocks and temporal concealment fits not well in the lost image area. 3D-DE [8] is only spatially smoothing macroblock boundary samples and therefore the inner part of the reconstructed macroblocks is still recovered temporally. The results for the proposed ST-FS and SBT-FS can be seen in Fig. 28 at the

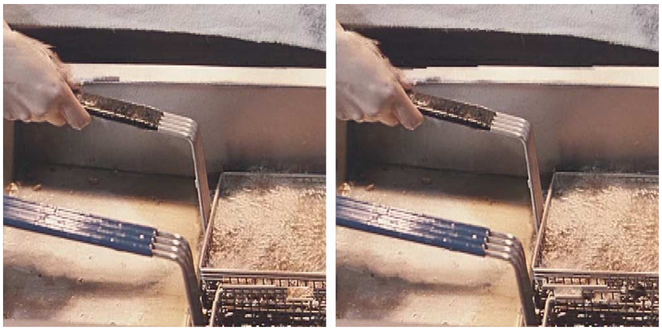

Fig. 27. Left: DMVE [5] with $\mathrm{Y\ PSNR} = 23.15\ \mathrm{dB}$. Right: 3D-DE [8] with $\mathrm{Y\ PSNR} = 22.90\ \mathrm{dB}$.

Fig. 28. Left: ST-FS with $\mathrm{Y\ PSNR} = 24.08\ \mathrm{dB}$. Right: SBT-FS with $\mathrm{Y\ PSNR} = 24.40\ \mathrm{dB}$.

left and the right hand side respectively. ST-FS is using the same motion estimation as 3D-DE [8]. For the above considered macroblocks the temporal recovered macroblock fits not well in the lost image area, spatial error concealment is higher weighted for the whole reconstructed macroblock samples. This area looks some what blurry but visually better than the results for DMVE [5], RBM [6] and 3D-DE [8]. For this B-frame SBT-FS can further enhance the visual quality as seen at the French fries basket by using additionally the future frame for concealing lost macroblock samples.

V. Conclusion

In this paper, we proposed fading techniques for error concealment in block-based video decoding systems. As seen in the simulation results, the spatio-temporal fading scheme (ST-FS) and the spatio-bi-temporal fading scheme (SBT-FS) has been evaluated for uncompressed and compressed video data. It has been shown that ST-FS and SBT-FS outperform the mentioned reference methods both in subjective and objective video quality. The main advantage of the proposed fading schemes is that each lost macroblock can individually be recovered pel-wise spatially from the current or bi-temporally from the previous and the future frame by weighted averaging three different estimations for the lost macroblock. This is very successful for error concealment in video decoding systems, especially in case the video signals are transmitted over mobile networks like DVB-T, DVB-H or other systems. In future work we will investigate methods for reducing computational complexity for the most time consuming part the motion estimation within ST-FS and SBT-FS.

Acknowledgment

The authors would like to thank Mirja Kühlewind and Daniel Wilhelm for their programming help and inspiring discussions. Additionally we would like to thank the European Regional Development Fund (ERDF) for the financial support and Prof. Sanjit K. Mitra (University of California in Santa Barbara) for cooperation, which is supported by Bavaria California Technology Center (BaCaTeC).

References

[1] A. Kaup, K. Meisinger, and T. Aach, "Frequency selective signal extrapolation with applications to error concealment in image communication," *Int. J. Electron. Commun. (AEÜ)*, vol. 59, pp. 147–156, May 2005.

[2] Y.-K. Wang, M. M. Hannuksela, and V. Varsa, "The error concealment feature in the H.26L test model," in *Proc. Int. Conf. on Image Processing (ICIP)*, Rochester, 2002, pp. 729–732.

[3] W.-M. Lam, A.-R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis, April 1993, pp. V417–V420.

[4] J.-W. Suh and Y.-S. Ho, "Error concealment based on directional interpolation," *IEEE Trans. on Consumer Electronics*, vol. 43, no. 3, pp. 295–302, Aug. 1997.

[5] J. Zhang, J. F. Arnold, and M. R. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 659–665, June 2000.

[6] M.-J. Chen, C.-S. Chen, and M.-C. Chi, "Temporal error concealment algorithm by recursive block-matching principle," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 11, pp. 1385–1393, Nov. 2005.

[7] J. Park, R. J. M. D.-C. Park, and M. A. El-Sharkawi, "Content-based adaptive spatio-temporal methods for MPEG-repair," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1066–1077, August 2004.

[8] M. Friebe and A. Kaup, "3D-deblocking for error concealment in block-based video decoding systems," in *Proc. Picture Coding Symposium (PCS)*, Beijing, April 2006.

[9] P. Piastowski, "System for decoder-independent reduction of block artifacts (system zur decoder-unabhängigen reduktion von blockartefakten)," in *ITG Fachbericht Elektronische Medien: 11 Dortmunder Fernsehseminar*. Berlin, Germany: VDE-Verlag, September 2005, pp. 215–218, no. 188.

[10] J. Zhai, K. Yu, J. Li, and S. Li, "A low complexity motion compensated frame interpolation method," *IEEE Int. Symposium on Circuits and Systems*, pp. 4927–4930, May 2005.

[11] Y. Wang and Q.-F. Zhu, "Error control and concealment in video communications—a review," *Proceedings of the IEEE*, vol. 86, no. 974–997, May 1998.

[12] M. Friebe and A. Kaup, "Spatio-bi-temporal error concealment in block-based video decoding systems," in *Proc. IEEE Int. Workshop on Multimedia and Signal Processing*, Victoria, Oct. 2006, pp. 296–301.

[13] VQEG Group, "Video Sequences," Aug. 2006 [Online]. Available: http://www.its.bldrdoc.gov/vqeg/

Markus Friebe received his Dipl.-Ing. (FH) degree in Electrical Engineering from the Georg-Simon-Ohm University of Applied Sciences in Nuremberg, Germany, in 2002. In 2002, he visited the Image Processing Lab from Prof. Sanjit K. Mitra at the University of California in Santa Barbara, USA, for writing his diploma thesis. He holds currently a position as a research scientist at the Chair of Multimedia Communications and Signal Processing of the University of Erlangen-Nuremberg, Germany, where he is working towards the Dr.-Ing. degree. His research interests include image and video signal processing with special focus on error concealment within video signals.

André Kaup received the Dipl.-Ing. and Dr.-Ing. degrees in Electrical Engineering from RWTH Aachen University, Germany, in 1989 and 1995, respectively. From 1989 to 1995 he was with the Institute for Communication Engineering at Aachen University of Technology, where he was responsible for industrial as well as academic research projects in the area of high resolution printed image compression, object-based image analysis and coding, and models for human perception. In 1995 he joined the Networks and Multimedia Communications department at Siemens Corporate Technology in Munich, where he chaired work packages in several European research projects in the area of very low bit rate video coding, image quality enhancement and mobile multimedia communications. In 1999 he was appointed head of the mobile applications and services group in the same department, with research focusing on multimedia adaptation for heterogeneous communication networks. Since 2001 he is full professor and head of the chair of multimedia communications and signal processing at Friedrich-Alexander-University of Erlangen-Nuremberg.

Prof. Kaup is member of the German ITG and senior member of the IEEE. He was elected Siemens inventor of the year 1998 and is recipient of the 1999 ITG Award. From 1998 to 2001 he served as adjunct professor at Technical University of Munich and the University of Erlangen-Nuremberg, teaching courses on image and video communication. Since 2005 he is vice-speaker of the DFG Collaborative Research Center 603 “Modeling and Analysis of Complex Scenes and Sensor Data”.