Spatio-Bi-Temporal Error Concealment in Block-Based Video Decoding Systems

Markus Friebe and André Kaup

Chair of Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{friebe,kaup}@LNT.de

***Abstract*— In this paper we present a spatio-bi-temporal fading scheme for block loss recovery in block-based video decoding systems. In the first part of the algorithm, based on two different boundary error criterions obtained from bi-temporal error concealment, either the previous, the future, or fading between both temporal methods is used for bi-temporal macroblock estimation. A weighted absolute difference between motion compensated image samples and macroblock boundary samples of the current frame represents one boundary error. In the second part of the algorithm, based on a boundary error criterion obtained from bi-temporal concealment, spatial, bi-temporal, or fading between both methods is used for recovering a lost macroblock. The advantage of this method is that one lost macroblock can be recovered pelwise spatially from the current or bi-temporally from the previous and the future frame by weighted averaging both error concealment results. The simulation results have shown that for recovering a lost macroblock this method outperforms the reference methods both in subjective and objective video quality.**

I. Introduction

Error concealment has a great importance in case video signals are transmitted over mobile networks. When transmitting block-based coded video over errorprone 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. Generally, restoration techniques are either spatial to conceal Intra-coded frames, temporal, spatio-temporal, or bi-temporal to conceal Inter-coded frames.

H.264 Intra [1] is using surrounding errorfree or concealed image samples for spatial bilinear interpolation. The Decoder Motion-Vector Estimation Algorithm (DMVE) [2] is a temporal technique and uses image samples which are immediately neighbored to the lost macroblock. Using these neighbored image samples, the DMVE is looking for the best match in the previous frame. For Bidirectional-coded frames the Decoder Motion-Vector Estimation Bidirectional Algorithm (DMVE-BiDir) [2] is looking for the best match in the previous and the future frame. A content-based adaptive spatio-temporal method (CABLR) [3] 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) [4] is a spatio-temporal technique which first recovers temporally a lost macroblock and further uses a deblocking filter for spatial filtering. A low complexity motion compensated frame interpolation method (OBMC) [5] is traditionally used for frame interpolation. We use this method for bi-temporal error concealment. Overlapped block bi-directional motion estimation estimates lost motion vectors. Motion vector post-processing eliminates bad estimated motion vectors and overlapped block motion compensation is recovering a lost macroblock.

H.264 Intra introduces burred image areas because of spatial bilinear interpolation. DMVE and DMVE-BiDir are both 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 generate some artifacts in case of object occlusions and uncovering. Blurred image areas along macroblock borders are introduced by OBMC in case of discontinuity of surrounding motion vectors.

We propose a spatio-bi-temporal method for concealing Bidirectional-coded frames. Using this method, appearing artifacts from object occlusions and uncovering can be further reduced. The spatio-bi-temporal fading scheme is further described in the following sections.

II. 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 II-A. The second part is recovering a lost macroblock spatio-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.

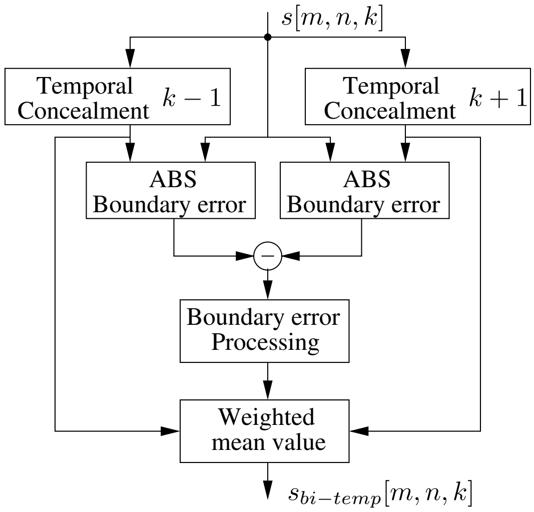

Fig. 1. Block diagram of the bi-temporal fading scheme.

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 II-B.

A. Bi-Temporal Fading Scheme

A block diagram of the bi-temporal fading scheme is shown in Fig. 1. 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 similar to DMVE [2] by minimizing a weighted sum of absolute differences $SAD_w^p[mv_m^p, mv_n^p]$. Eq. (1) and Eq. (2) show, how $SAD_w^p[mv_m^p, mv_n^p]$ is calculated and how the motion vector is estimated from the minimum $SAD_w^p[mv_m^p, mv_n^p]$. Image samples s in the current frame k are compared to motion compensated image samples in 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 area marked in dark-gray represents correct and the area marked in light-gray the incorrect received image samples for the left most block of a consecutive macroblock loss. For the right most block of a consecutive macroblock loss, the area marked in dark-gray 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. Hence we calculate the weighted sum of absolute differences by

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

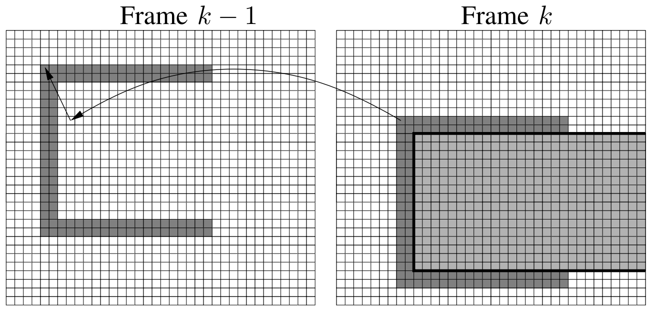

Fig. 2. Left: Frame $k-1$. Right: Frame k. The area marked in dark-gray represents the $SAD_w^p[mv_m^p, mv_n^p]$ evaluation area with $w[m,n] = 1$. The area marked in light-gray represents the lost image samples with $w[m,n] = 0$. Both for the left most consecutive macroblock loss.

and the best matching motion vector by

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

The difference between DMVE and this method is the weighting function. DMVE treats concealed areas as correctly received image samples by setting $w[m,n] = 1$. In this approach, concealed areas are lower weighted by setting $w[m,n] = 0.3$. For this value the best simulation results have been achieved.

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 Eq. (1). The motion vector is estimated from the minimum $SAD_w^f[mv_m^f, mv_n^f]$ by

$$[\tilde{mv}_m^f, \tilde{mv}_n^f] = \arg \min_{mv_m^f, mv_n^f} (SAD_w^f[mv_m^f, mv_n^f]). \quad (3)$$

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

$$s_{temp}^p[m,n,k] = s[m+\tilde{mv}_m^p, n+\tilde{mv}_n^p, k-1] \quad (4)$$

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

$$s_{temp}^f[m,n,k] = s[m+\tilde{mv}_m^f, n+\tilde{mv}_n^f, k+1] \quad (5)$$

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

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 boundary error $e^p[m,n]$. The weights $w[m,n]$ for computing the boundary error are the same as for estimating the motion vectors. The boundary error measures pelwise, 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

$$e^p[m,n] = w[m,n] \cdot |s[m,n,k] - s_{temp}^p[m,n,k]| \quad (6)$$

and for the future frame

$$e^f[m,n] = w[m,n] \cdot |s[m,n,k] - s_{temp}^f[m,n,k]| \quad (7)$$

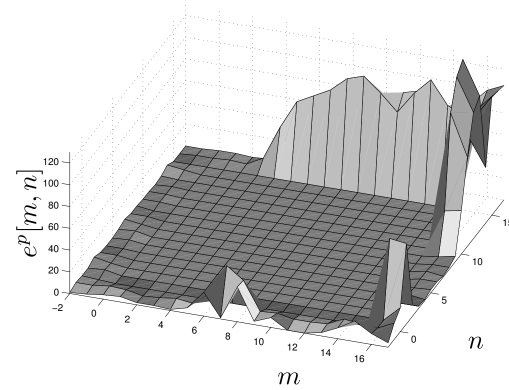

Fig. 3. Boundary error $e^p[m, n]$.

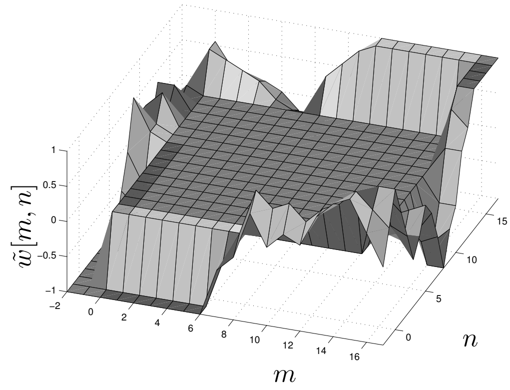

Fig. 5. Lowpass filtered decision signal $\tilde{w}[m, n]$.

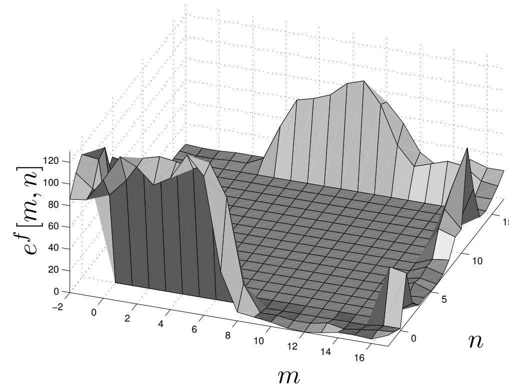

Fig. 4. Boundary error $e^f[m, n]$.

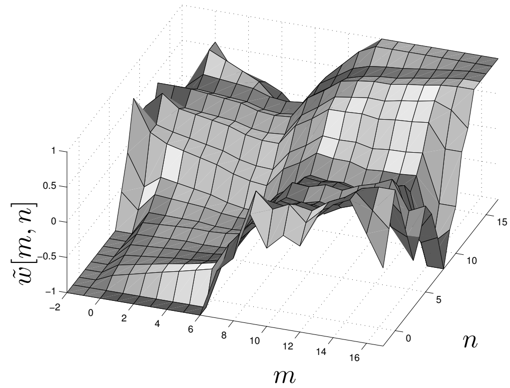

Fig. 6. Bilinearly interpolated lowpass filtered decision signal $\tilde{w}[m, n]$.

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 Fig. 3 and Fig. 4. 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 frame fits relatively well. After computing both boundary errors, a gradient

$$\bigtriangledown[m, n] = e^p[m, n] - e^f[m, n] \quad (8)$$

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

$$\bigtriangledown_{sign}[m, n] = sign(\bigtriangledown[m, n]) \quad (9)$$

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

$$\tilde{w}[m, n] = \frac{1}{\sum_{i=-1}^{1} \sum_{j=-1}^{1} w[m-i, n-j]} \cdot \sum_{i=-1}^{1} \sum_{j=-1}^{1} \bigtriangledown_{sign}[m-i, n-j] \quad (10)$$

For filtering symmetrically boundary extensions are used. The lowpass filtered decision signal $\tilde{w}[m, n]$ can be seen in Fig. 5. The area $\mathcal{L}$ represents here the lost macroblock ($m = 0, ..., M-1 \wedge n = 0, ..., N-1$) 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. 6 and is used for pelwise 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

$$s_{bi-temp}[m, n, k] = \tfrac{1}{2} \cdot \big((1 - \tilde{w}[m, n]) \cdot s^p_{temp}[m, n, k] + (1 + \tilde{w}[m, n]) \cdot s^f_{temp}[m, n, k] \big). \quad (11)$$

$s_{bi-temp}[m, n, k]$ is estimated using a pelwise weighted mean value between $s^p_{temp}[m, n, k]$ and $s^f_{temp}[m, n, k]$. Lowpass filtering of the decision signal is necessary to avoid switching between temporal concealment from the previous and the future frame for neighboring image samples within $s_{bi-temp}[m, n, k]$. The example in Fig. 6 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.

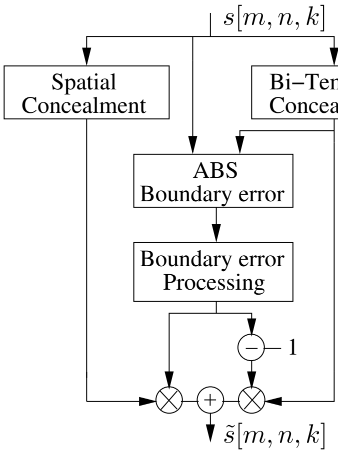

Fig. 7. Block diagram of the spatio-temporal fading scheme.

B. Spatio-Temporal Fading Scheme

The spatio-temporal fading scheme uses instead of temporal error concealment the before mentioned bi-temporal error concealment as seen in Fig. 7. The boundary error $e[m,n]$ 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 pelwise, if bi-temporal error concealment fits well to the lost macroblock area.

$$e[m,n] = w[m,n] \cdot |s[m,n,k] - s_{bi-temp}[m,n,k]| \qquad (12)$$

For spatial concealment, a bilinear interpolation method H.264 Intra [1] is used to obtain the spatially estimated image samples $s_{spatial}[m,n,k]$ for the lost macroblock. The boundary error processing unit for spatio-temporal fading scheme is different to that in Section II-A. After computing the boundary error, a lowpass filter computes a mean boundary error $\bar{e}[m,n]$ within a 3×3 window along the macroblock boundary samples.

$$\bar{e}[m,n] = \frac{1}{\sum_{i=-1}^{1}\sum_{j=-1}^{1} w[m-i,n-j]} \cdot \sum_{i=-1}^{1}\sum_{j=-1}^{1} e[m-i,n-j] \qquad (13)$$

The area $\mathcal{L}$ represents the area ($m = 0,...,M-1 \wedge n = 0,...,N-1$) encircled by the lowpass filtered boundary error. $\mathcal{L}$ is then bilinearly interpolated from the nearest neighbors $\bar{e}[m,n]$, outside of $\mathcal{L}$, in vertical and horizontal direction. The distances to the nearest neighbors represent the weights for bilinear interpolation. Bi-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.

$$\tilde{w}[m,n] = \begin{cases} 1 & , \quad \tilde{w}[m,n] > T_w \\ \frac{\tilde{w}[m,n]}{T_w} & , \quad \text{else} \end{cases} \qquad (14)$$

Fig. 8. Original image.

Fig. 9. Consecutive 16×16 macroblock loss.

These weighting matrix $\tilde{w}[m,n]$ is used for pelwise weighted averaging between spatially $s_{spatial}[m,n,k]$ and bi-temporally $s_{bi-temp}[m,n,k]$ estimated lost macroblock samples.

$$\tilde{s}[m,n,k] = (1-\tilde{w}[m,n]) \cdot s_{bi-temp}[m,n,k] + \tilde{w}[m,n] \cdot s_{spatial}[m,n,k]. \qquad (15)$$

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. To avoid high frequencies in the resulting weighting matrix $\tilde{w}[m,n]$ for fading between bi-temporal and spatial error concealment, lowpass filtering of $e[m,n]$ is necessary.

III. Simulation Results

For simulations we consider uncompressed video frames 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

Fig. 10. Left: H.264 Intra [1] with Y PSNR=16.68 dB. Right: 3D-DE [4] with Y PSNR=19.65 dB.

Fig. 11. Left: ST-FS with Y PSNR=20.74 dB. Right: DMVE-BiDir [2] with Y PSNR=20.12 dB.

Fig. 12. Left: OBMC [5] with Y PSNR=21.65 dB. Right: SBT-FS with Y PSNR=23.06 dB.

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 and DMVE-BiDir use both 2 line encirclement for computing SAD. A spatio-temporal fading scheme (ST-FS) uses a temporal concealment unit instead of the bi-temporal concealment unit in Fig. 7. The temporal concealment unit reestimates a lost macroblock only from the previous frame.

In Fig. 8 a frame of the sequence *Basketball* is shown without errors and with consecutive macroblock loss in Fig. 9. The result for H.264 Intra is shown at the left hand side and for 3D-DE at the right hand side in Fig. 10. H.264 Intra is introducing blurred image areas because of the bilinear interpolation. 3D-DE introduces some block artifacts in case of object occlusions and uncovering, as seen at the shoes from the basketball players. The result for ST-FS is shown at the left hand side and for DMVE-BiDir at the right hand side in Fig. 11. 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 at the head of the basketball player next to the camera. At the left hand side in Fig. 12, the result for OBMC and at the right hand side the result for the proposed SBT-FS method can be seen. Strong blurred image areas are visible between the black shoes next to the camera for OBMC. For the proposed SBT-FS method, the recovered areas look visually better than the recovered areas from the 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 yields the lowest PSNR values. Temporal methods like DMVE and spatio-temporal like CABLR, 3D-DE 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 for *Foreman* and *Basketball* and 3D-DE better results than DMVE-BiDir for *Foreman*. OBMC 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. 13 luminance PSNR results for evaluated *Basketball* video frames are shown. At the very beginning of the sequence up to the 7th evaluated frame, OBMC and SBT-FS achieves 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 methods because SBT-FS is able to recover additionally spatially in case bi-temporal error concealment fits not well to the lost macroblock area. SBT-FS achieves for 16 out of 20 evaluated *Basketball* frames the highest PSNR results.

TABLE I
MEAN Y PSNR RESULTS.

Video	*Fore-man*	*Basket-ball*	*Sales-man*	*Flower garden*	*Coast-guard*
H.264 Intra [1]	21.99	15.93	22.26	15.44	17.74
DMVE [2]	32.75	21.16	32.60	26.51	28.46
CABLR [3]	33.15	20.58	32.63	25.86	27.68
3D-DE [4]	33.46	21.73	33.49	27.23	27.86
ST-FS	33.51	22.37	33.87	27.14	28.28
DMVE-BiDir [2]	33.43	21.81	32.89	27.42	29.40
OBMC [5]	34.63	23.52	35.42	30.93	30.08
SBT-FS	**36.10**	**24.38**	**35.57**	**31.45**	**31.97**

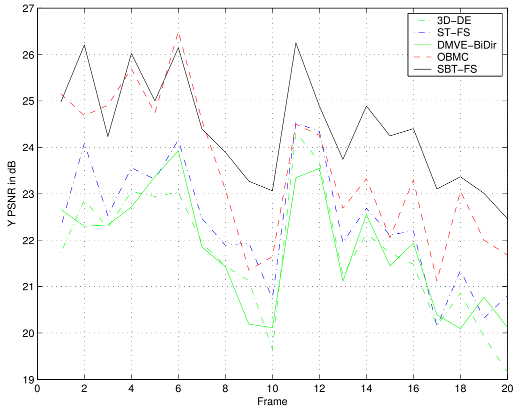

Fig. 13. PSNR results for evaluated *Basketball* frames.

IV. CONCLUSION

In this paper we presented a spatio-bi-temporal fading scheme for macroblock loss recovery in block-based video decoding systems. The simulation results show that for macroblock loss in uncompressed video frames this method outperforms the reference methods both in subjective and objective video quality. For compressed video data, appearing artifacts are propagated from erroneously decoded Intra-coded frames to Inter-coded frames in case of transmission errors. The spatio-temporal fading scheme (ST-FS) as seen in the simulation results is well suited for recovering lost macroblocks in case only a previous frame in known. ST-FS can be used in Intra-coded and Predictive-coded frames to reduce error propagation. In case a previous and a future frame is known, spatio-bi-temporal fading scheme (SBT-FS) performs best for macroblock loss recovery. Using SBT-FS for Bidirectional-coded frames, lost macroblocks will be better reconstructed than using the reference bidirectional methods.

REFERENCES

[1] Y.-K. Wang, M. M. Hannuksela, and V. Varsa, The Error Concealment Feature in the H.26L Test Model, Proc. Int. Conf. on Image Processing (ICIP), pp. 729-732, Rochester, Sep. 2002.
[2] 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, pp. 659-665, Jun. 2000.
[3] J. Park, D.-C. Park, R. J. Marks, and M. A. El-Sharkawi, Content-Based Adaptive Spatio-Temporal Methods for MPEG-Repair, IEEE Trans. Image Process., Vol. 13, pp. 1066-1077, Aug. 2004.
[4] M. Friebe, and A. Kaup, 3D-Deblocking for Error Concealment in Block-Based Video Decoding Systems, Proc. Picture Coding Symposium (PCS), Beijing, Apr. 2006.
[5] 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, Kobe, May 2005.