16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

MOTION COMPENSATED FREQUENCY SELECTIVE EXTRAPOLATION FOR ERROR CONCEALMENT IN VIDEO CODING

Jürgen Seiler and André Kaup

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

ABSTRACT

Although wireless and IP-based access to video content gives a new degree of freedom to the viewers, the risk of severe block losses caused by transmission errors is always present. The purpose of this paper is to present a new method for concealing block losses in erroneously received video sequences. For this, a motion compensated data set is generated around the lost block. Based on this aligned data set, a model of the signal is created that continues the signal into the lost areas. Since spatial as well as temporal informations are used for the model generation, the proposed method is superior to methods that use either spatial or temporal information for concealment. Furthermore it outperforms current state of the art spatio-temporal concealment algorithms by up to 1.4 dB in PSNR.

1. INTRODUCTION

Due to modern video codecs and increased computational power, the transmission and processing of video signals in wireless environments or IP-based access to videos became more and more usual in the past years. Unfortunately, in these cases the risk of data lost in transmission or erroneously received is omnipresent. To cope with this risk, modern video codecs such as H.264/AVC use two strategies. According to [1], the first one is error resilience by protecting the coded video against transmission errors and by minimizing the potential damage produced by incorrectly received bits. In the case that an error occurs, the second strategy has to take place, the concealment of block losses. Although the concealment is not part of the actual video standard, error concealment is widely used in many decoders in order to display a pleasant video and to reduce error propagation. A good overview over this area can be found in [2].

For concealing a lost block, most existent techniques use either spatial or temporal information for extrapolating the signal into the area of the lost block. The spatial methods only use information from the neighborhood of the lost block in the actual frame to extrapolate the signal into the area of the lost block. Two algorithms out of this group are e. g. the Projections onto Convex Sets from [3] and the Sequential Error-Concealment from [4]. On the other hand, the temporal methods extrapolate the signal into the lost area only by means of information from previous or following already correctly received frames. In most cases, these methods try to estimate the motion in a sequence and replace the lost block with one from another frame, shifted according to the estimated motion. Two powerful temporal concealment algorithms are e. g. the Extended Boundary Matching Algorithm from [5] or the Decoder Motion-Vector Estimation from [6]. Unlike these two groups, spatio-temporal algorithms use spatial as well as temporal information for concealing the errors. One state of the art spatio-temporal concealment algorithm is the Three-Dimensional Frequency Selective Extrapolation (3D-FSE) introduced by [7]. This algorithm aims at generating a model of the signal for a data volume centered by the erroneous block. The 3D-FSE is able to implicitly compensate small motion inside the volume, leading to very good objective and subjective extrapolation results.

Now, we propose a new concealment technique, the Motion Compensated Frequency Selective Extrapolation (MC-FSE). It is based on the 3D-FSE but uses an explicit motion estimation and compensation prior to the model generation. By explicitly estimating the motion and aligning the extrapolation volume, significantly better extrapolation results can be obtained compared to 3D-FSE.

2. MOTION COMPENSATED FREQUENCY SELECTIVE EXTRAPOLATION

Fig. 1 shows three consecutive frames of a possible video sequence $v[x, y, t]$ depicted by the two spatial coordinates x and y and the temporal coordinate t. In the frame at time instance $t = \tau$ an isolated block loss occurs with the top left corner at (x_0, y_0). Although isolated block losses are considered for the illustration of the algorithm, the algorithm easily can be adapted to other loss scenarios as well. In order to conceal the lost block, surrounding pixels in the actual frames and pixels in one or more previous and following frames are used. The number of used previous frames is indicated by N_p, of used following frames by N_f. In the case that the considered sequence is not a static one, the areas in the previous or following frame that correspond to the block loss may have been moved. Based on the frame at $t = \tau$, the movement relative to the frame at $t = \tau + \kappa$ is described by the motion vector

$$\mathbf{d}^{(\kappa)} = \left(x_\mathrm{d}^{(\kappa)}; y_\mathrm{d}^{(\kappa)}\right) \tag{1}$$

whereas each motion vector includes the displacement in horizontal direction $x_\mathrm{d}^{(\kappa)}$ and in vertical direction $y_\mathrm{d}^{(\kappa)}$. For the actual concealment the motion has to be revoked in order to generate an aligned three-dimensional data set, the so called extrapolation volume $\mathcal{L}$. The volume $\mathcal{L}$ is depicted by the spatial coordinates m and n and the temporal coordinate p. All in all, it is of size $M \times N \times P$. Further, it contains the lost block subsumed in area $\mathcal{B}$ and all the pixels used to extrapolate the signal into this area. These pixels are subsumed in volume $\mathcal{A}$, called support volume.

The block diagram in Fig. 2 shows the different steps of the MC-FSE that will be carried out in detail in the subsequent subsections. Starting with the lost block, the motion of the sequence around this block is estimated. As this estimation may be inaccurate, the motion estimation's reliability is checked. If the motion estimation is reliable, the extrapolation volume is aligned according to the estimated motion. Afterwards, the Three-Dimensional Frequency Selective Extrapolation [7] enhanced by the fast orthogonality deficiency compensation proposed in [8] is applied to the extrapolation volume. Finally, the area corresponding to the lost block is cut out of the generated model and is used for replacing the lost block.

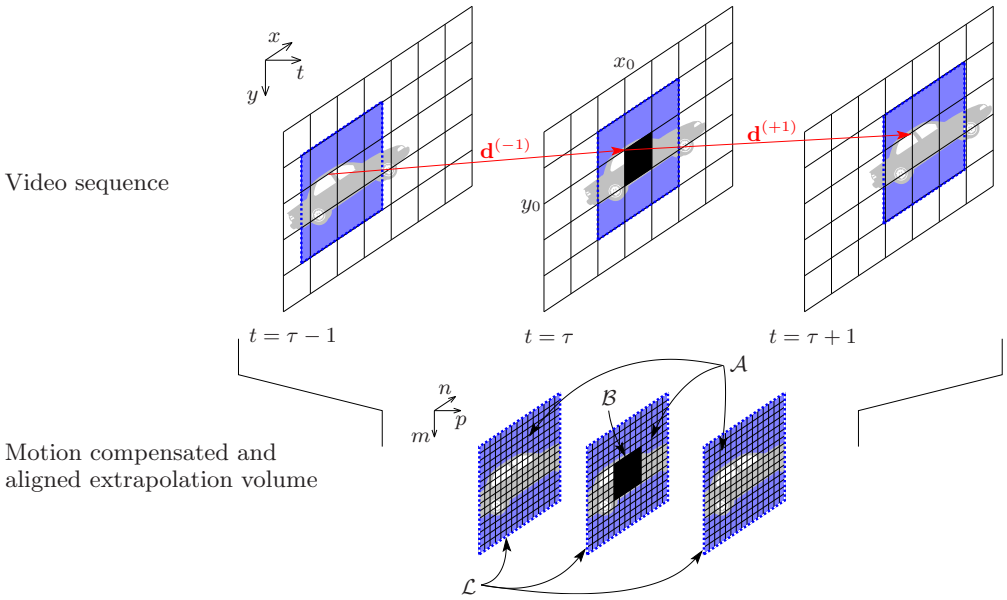

Figure 1: Original video sequence $v\,[x, y, t]$ and aligned extrapolation volume $\mathcal{L}$.

2.1 Motion estimation

As mentioned above, the first step of the MC-FSE is the motion estimation for obtaining an aligned extrapolation volume. Although the 3D-FSE is able to inherently compensate minor motion, large motion cannot be compensated well as in this case the support volume covers inappropriate content of the previous and following frames. Thus, better extrapolation results are obtainable by aligning the extrapolation volume to revoke the motion.

Since the lost block cannot be used to determine the motion vector, the motion is estimated according to the Decoded Motion-Vector Estimation [6]. Thereby, the estimation is performed by using an area around the lost block that here is called decision area $\mathcal{M}$ (see Fig. 3). For estimating the motion from the frame at $t = \tau$ to the frame at $t = \tau+\kappa$ a set of possible motion vectors is analyzed whereas for every possible motion vector $\tilde{\mathbf{d}}^{(\kappa)} = \left(\tilde{x}_\mathrm{d}^{(\kappa)}; \tilde{y}_\mathrm{d}^{(\kappa)}\right)$ the sum of squared errors

$$E^{(\kappa)}\left(\tilde{\mathbf{d}}^{(\kappa)}\right) = \sum_{\forall (x,y)\in\mathcal{M}} \left(v\,[x, y, \tau] - v\left[x + \tilde{x}_\mathrm{d}^{(\kappa)}, y + \tilde{y}_\mathrm{d}^{(\kappa)}, \tau + \kappa\right]\right)^2 \tag{2}$$

is calculated for all pixels in area $\mathcal{M}$. After this, the motion vector is chosen that minimizes the sum of squared errors

$$\hat{\mathbf{d}}^{(\kappa)} = \underset{\tilde{x}_\mathrm{d}^{(\kappa)}, \tilde{y}_\mathrm{d}^{(\kappa)} = -d_\mathrm{max}, \ldots, d_\mathrm{max}}{\operatorname{argmin}} E^{(\kappa)}\left(\tilde{x}_\mathrm{d}^{(\kappa)}, \tilde{y}_\mathrm{d}^{(\kappa)}\right). \tag{3}$$

Here, the set of motion vectors to be tested covers all vectors with a maximum displacement of d_max in horizontal and vertical direction.

The estimation error $\check{E}^{(\kappa)}$ for the chosen motion vector is determined by

$$\check{E}^{(\kappa)} = E^{(\kappa)}\left(\hat{\mathbf{d}}^{(\kappa)}\right). \tag{4}$$

2.2 Estimation reliability check

Unfortunately, in some cases the motion cannot be estimated well. This may happen e. g. in the event of non-translational movement or changes in the content. As no reliable motion

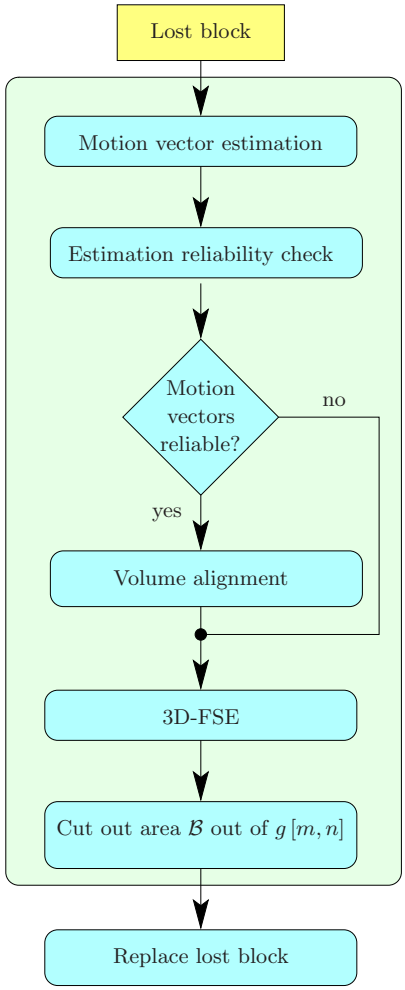

Figure 2: Block diagram for MC-FSE

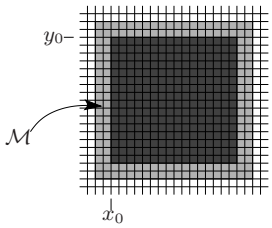

Figure 3: Decision area $\mathcal{M}$ for motion estimation of block at (x_0, y_0)

vectors can be derived then, a compensated volume may be fitted less well for the extrapolation than an uncompensated one. To protect the algorithm from suffering from this circumstance, two thresholds are used to evaluate the motion vector quality. If at least one of these thresholds is exceeded, the motion vectors are discarded and no alignment is applied to the extrapolation volume.

The first threshold applies to the absolute motion vector quality. Therefore the maximum of the square root of the estimation error $\check{E}^{(\kappa)}$ for all preliminary calculated motion vectors is determined. This is normalized by the cardinality of $\mathcal{M}$ in order to get the mean error per pixel. The threshold for this criterion is denoted by T_{abs} and if

$$\frac{1}{|\mathcal{M}|} \max_{\kappa=-N_{\text{p}},\dots,N_{\text{f}}\setminus 0} \sqrt{\check{E}^{(\kappa)}} > T_{\text{abs}} \tag{5}$$

holds, all the estimated motion vectors are discarded.

The second criterion is the homogeneity of the motion vector quality. For this, the difference between the maximum estimation error and the minimum error is computed and normalized to the mean estimation error. The resulting quotient is compared to the threshold T_{rel}. Thus, if

$$\frac{\max\limits_{\kappa=-N_{\text{p}},\dots,N_{\text{f}}\setminus 0} \sqrt{\check{E}^{(\kappa)}} - \min\limits_{\kappa=-N_{\text{p}},\dots,N_{\text{f}}\setminus 0} \sqrt{\check{E}^{(\kappa)}}}{\frac{1}{N_{\text{p}}+N_{\text{f}}} \sum\limits_{\kappa=-N_{\text{p}},\dots,N_{\text{f}}\setminus 0} \sqrt{\check{E}^{(\kappa)}}} > T_{\text{rel}} \tag{6}$$

is fulfilled, the motion vector quality is too distinct for getting a well aligned extrapolation volume. In this case, no alignment is applied and the volume is directly taken from the sequence without revoking any motion.

2.3 Volume alignment

After estimating the motion vectors for all frames used for the extrapolation, the projection volume $\mathcal{L}$ is set up. In the center, it consists of the lost block $\mathcal{B}$. This one is enframed by surrounding, correctly received pixels from the actual frame, for paying attention to the spatial neighborhood during the extrapolation process. From the used previous and following frames, the corresponding areas are shifted according to the appropriate motion vectors and are added to the extrapolation volume. The relation between the video sequence with the lost block and the aligned extrapolation volume is illustrated in Fig. 1 for the example of one previous and one following frame. So the motion of the sequence around the lost block is compensated and the aligned extrapolation is used for the subsequent Three-Dimensional Frequency Selective Extrapolation.

2.4 Frequency Selective Extrapolation

After having performed the motion compensation for the extrapolation volume, the actual signal extrapolation can be carried out. It is based on the 3D-FSE [7] combined with the orthogonality deficiency compensation proposed in [8]. In order to extrapolate the signal from the support volume $\mathcal{A}$ into the loss area $\mathcal{B}$ the algorithm aims at generating a parametric model $g\,[m,n,p]$ for the original signal. Therefore the model is generated by approximating the known signal in volume $\mathcal{A}$. As the model is defined over the complete volume $\mathcal{L}$, it continues the signal into area $\mathcal{B}$. The model

$$g\,[m,n,p] = \sum_{\forall k \in \mathcal{K}} c_k \varphi_k\,[m,n,p] \tag{7}$$

itself is a weighted superposition of mutually orthogonal three-dimensional basis functions $\varphi_k\,[m,n,p]$. The expansion coefficients c_k control the weight of each basis function. The set $\mathcal{K}$ covers the indices of all basis functions used for the extrapolation.

The model generation works iteratively whereas in every iteration step one basis function is chosen, denoted by index u. Then, this one is added to the parametric model from the previous iteration step together with an estimate $\hat{c}_u^{(\nu)}$ for the expansion coefficient. In the ν-th iteration step, this leads to

$$g^{(\nu)}\,[m,n,p] = g^{(\nu-1)}\,[m,n,p] + \hat{c}_u^{(\nu)} \cdot \varphi_u\,[m,n,p]\,. \tag{8}$$

The initial model $g^{(0)}\,[m,n,p]$ is all zero. For all $(m,n,p) \in \mathcal{A}$ the residual approximation error can be computed according to

$$r^{(\nu)}\,[m,n,p] = r^{(\nu-1)}\,[m,n,p] - \hat{c}_u^{(\nu)} \cdot \varphi_u\,[m,n,p]\,. \tag{9}$$

As in area $\mathcal{B}$ no information about the original signal is existent, no approximation error can be computed there. The initial approximation error $r^{(0)}\,[m,n,p]$ is equal to the original signal in the volume $\mathcal{A}$.

In order to determine the basis function to use and to determine the estimate for the expansion coefficient, in every iteration step the projection coefficients $p_k^{(\nu)}$ are computed for all possible basis functions. $p_k^{(\nu)}$ results from the weighted projection of the approximation error onto the basis function $\varphi_k\,[m,n,p]$.

$$p_k^{(\nu)} = \frac{\sum\limits_{(m,n,p)\in\mathcal{L}} r^{(\nu-1)}\,[m,n,p] \cdot \varphi_k\,[m,n,p] \cdot w\,[m,n,p]}{\sum\limits_{(m,n,p)\in\mathcal{L}} w\,[m,n,p] \cdot \varphi_k^2\,[m,n,p]} \tag{10}$$

The weighting function

$$w\,[m,n,p] = \begin{cases} \rho\,[m,n,p] & , \quad \forall\,(m,n,p) \in \mathcal{A} \\ 0 & , \quad \forall\,(m,n,p) \in \mathcal{B} \end{cases} \tag{11}$$

is used to exclude area $\mathcal{B}$ from the projection process and to control the influence each pixel has on the extrapolation process depending on its position by means of $\rho\,[m,n,p]$. According to [7], the weighting function used is generated by an isotropic model centered in the mid of the lost block. This leads to

$$\rho\,[m,n,p] = \hat{\rho}^{\sqrt{\left(m-\frac{M-1}{2}\right)^2+\left(n-\frac{N-1}{2}\right)^2+\left(p-\frac{P-1}{2}\right)^2}} \tag{12}$$

with constant $\hat{\rho}$ between 0 and 1. So the pixels get less influence on the model generation with an increasing distance to the center of the lost block. After having computed all projection coefficients, the basis function to be added to the model is determined. The decision is made for the basis function that minimizes the distance between the residual error and the projection onto the basis function.

After having chosen one basis function the corresponding expansion coefficient has to be estimated. Although the basis functions are orthogonal with respect to the whole extrapolation volume $\mathcal{L}$, they are not orthogonal when evaluated with respect to the support volume $\mathcal{A}$. Due to this, the projection does not only lead to the real portion a basis function has on the approximation error but incorporates portions from other basis functions as well. In order to estimate the expansion coefficient $\hat{c}_u^{(\nu)}$ from the projection coefficient the fast orthogonality deficiency compensation proposed in [8] is used, resulting in

$$\hat{c}_u^{(\nu)} = \gamma \cdot p_u^{(\nu)}. \tag{13}$$

The orthogonality deficiency compensation factor γ is constant and typically is in the range of 0 to 1.

The iteration steps are repeated until the maximum number of iterations is reached. Finally the pixels corresponding to the loss area $\mathcal{B}$ are cut out of $g[m,n,p]$ and are used to conceal the lost block.

3. SIMULATION SETUP AND RESULTS

In order to demonstrate the extrapolation quality of the MC-FSE, the concealment of block losses in the CIF sequences "City", "Foreman", and "Vimto" is evaluated. Therefore, for all regarded sequences, in the frames number 17, 47, 77, 107, and 137 blocks of size 16×16 pixels are cut out according to the loss pattern showed on the left side of Fig. 5. Then, for the luminance component these blocks are extrapolated and the extrapolation results are compared to the original blocks in terms of PSNR.

The basis functions used for the extrapolation are the functions of the three-dimensional discrete Fourier transform. According to [7], this set of basis functions is especially suited for concealment of errors in natural images, since flat as well as noise like areas and edges can be reconstructed well. Additionally, an efficient implementation operating in the Fourier domain is possible [8]. For the actual extrapolation, $N_\mathrm{p} = 2$ previous and $N_\mathrm{f} = 2$ following frames are used. The support area in the actual frame is a band of 16 pixels width around the lost block. This results in an extrapolation volume with an overall size of $48 \times 48 \times 5$ pixels. The decision area $\mathcal{M}$ for motion estimation is of 4 pixels width and the search range is $d_\mathrm{max} = 16$ pixels in each direction at fullpel accuracy. As the efficient implementation operates in the Fourier domain, the extrapolation volume has to be transformed into this domain. Therefore, a FFT of size $64 \times 64 \times 16$ is used. As mentioned before, the weighting function is generated by an isotropic model. Regarding the variable $\hat{\rho}$ of the isotropic model, a value of 0.8 has lead to good extrapolation results. For controlling the orthogonality deficiency compensation, the factor γ is chosen to 0.6 as this value is a good tradeoff between compensation quality and the number of iterations needed for achieving the maximum PSNR. The two thresholds for evaluating the motion estimation quality are chosen to $T_\mathrm{rel} = 3$ and $T_\mathrm{abs} = 100$. Fortunately, these two thresholds as well as the values given before are not very critical and can be varied in a relatively wide range without affecting the extrapolation result very much.

In Fig. 4, for the mentioned sequences the obtainable PSNR is shown with respect to the number of iterations used for the model generation. For illustrating the gain of MC-FSE over 3D-FSE the graph also shows the PSNR for the 3D-FSE that performs the extrapolation on the non-aligned extrapolation volume. Except for the motion compensation, all other parameters are chosen in the same way for 3D-FSE as for MC-FSE. Thus, the orthogonality deficiency compensation from [8] is also applied to the 3D-FSE, although originally not used in [7]. Obviously, by applying an explicit

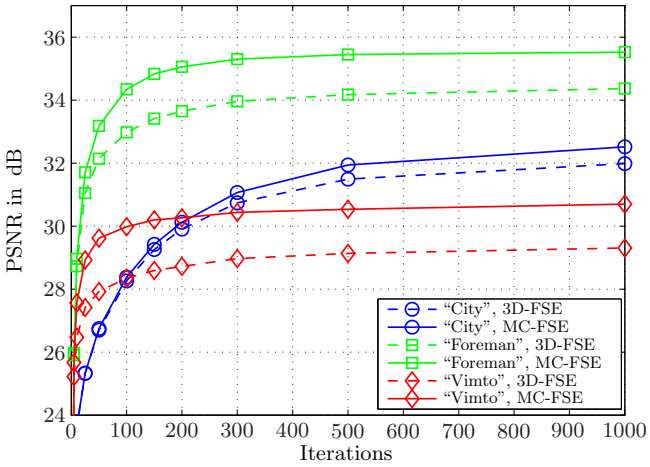

Figure 4: PSNR over iterations for 3D-FSE and MC-FSE. Block of size 16×16 pixels, $N_\mathrm{p} = 2$, $N_\mathrm{f} = 2$, 16 pixels support area width, 4 pixels wide $\mathcal{M}$, FFT $64 \times 64 \times 16$, $\hat{\rho} = 0.8$, $\gamma = 0.6$, $d_\mathrm{max} = 16$, $T_\mathrm{rel} = 3$, $T_\mathrm{abs} = 100$

	"City"	"Foreman"	"Vimto"
TR	25.63 dB	27.60 dB	22.79 dB
EBMA [5]	22.53 dB	30.10 dB	28.14 dB
DMVE [6]	27.94 dB	33.04 dB	28.21 dB
3D-FSE [7]	31.99 dB	34.37 dB	29.31 dB
MC-FSE	32.52 dB	35.53 dB	30.70 dB

Table 1: Comparison of different concealment algorithms

motion compensation and alignment of the extrapolation volume prior to the Frequency Selective Extrapolation the extrapolation quality can be increased significantly. For the considered sequences a maximum increment in PSNR of up to 1.4 dB is possible. Furthermore, most of the maximum gain already is existent at low numbers of iterations. Although the difference in PSNR depends on the sequence, for all tested sequences the explicit motion compensation leads to better extrapolation results.

Additionally, the MC-FSE is compared to several existent concealment techniques. As, according to [9], in general temporal concealment algorithms are superior to spatial ones, the comparison is carried out to temporal methods only. The simplest temporal method for concealing a block loss is the Temporal Replacement (TR). Thereby, the lost block is replaced by the block from the previous frame that has the same position. A more sophisticated approach is the Extended Boundary Matching Algorithm (EBMA) from [5] that replaces the lost block with a block from the previous frame that minimizes the boundary error between the correctly received neighboring blocks and the candidate block. Another method for concealing lost blocks is the Decoder Motion-Vector Estimation (DMVE) from [6] that already served as basis for the motion estimation in Section 2.1. The search range for EBMA and DMVE is set to $d_\mathrm{max} = 16$ pixels as well. In Table 1 the obtainable PSNR is listed for the mentioned reference algorithms, the 3D-FSE [7], and the proposed Motion Compensated Frequency Selective Extrapolation (MC-FSE). For 3D-FSE and MC-FSE the number of previous and following frames is chosen to $N_\mathrm{p} = 2, N_\mathrm{f} = 2$. In addition, for 3D-FSE and MC-FSE Table 2 shows the concealment quality when different numbers of previous and following frames are used for the model generation. Apparently, due to making use of the spatial information as well, the 3D-FSE and the MC-FSE outperform the pure temporal concealment techniques. But, as already shown in Fig. 4,

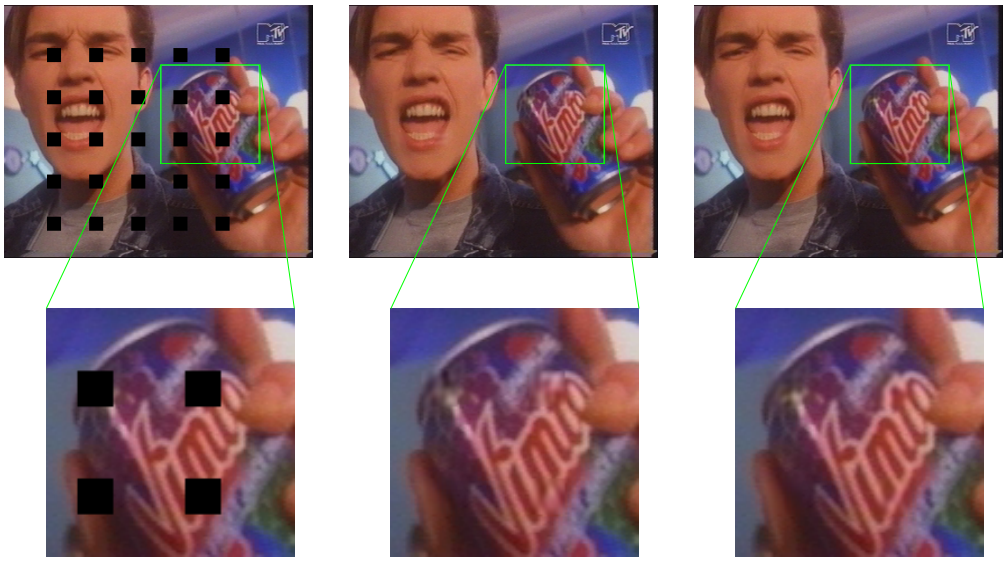

Figure 5: Visual extrapolation quality for frame 11 of CIF sequence "Vimto". Left: Error pattern. Mid: 3D-FSE. Right: MC-FSE

	3D-FSE [7]	MC-FSE
$N_\mathrm{p} = 1, N_\mathrm{f} = 0$	29.21 dB	29.35 dB
$N_\mathrm{p} = 2, N_\mathrm{f} = 0$	28.80 dB	30.01 dB
$N_\mathrm{p} = 1, N_\mathrm{f} = 1$	30.63 dB	31.08 dB
$N_\mathrm{p} = 2, N_\mathrm{f} = 2$	29.31 dB	30.70 dB

Table 2: Comparison of different numbers of frames used for concealment for sequence"Vimto"

the explicit motion compensation of the extrapolation volume leads to better extrapolation results for the proposed algorithm.

Besides the objective evaluation of the extrapolation quality in terms of PSNR, the subjective visual extrapolation quality is important. Although the 3D-FSE is able to conceal the losses without almost any visible artifacts, in some rare cases, the original signal cannot be reconstructed sufficiently. As already indicated by the PSNR, the MC-FSE is superior to the 3D-FSE and additionally is able to conceal these critic losses better. Fig. 5 is used to illustrate such a case. The middle image is concealed with 3D-FSE and shows some minor artifacts, e. g. at the top of the "t" or the bottom of the "m". On the right side the same image is concealed with MC-FSE. With this new technique, the block losses can be concealed very good, and no artifacts are visible any more.

4. CONCLUSION

The proposed algorithm is an enhancement to the already existent Three-Dimensional Frequency Selective Extrapolation. By explicitly compensating the motion of a sequence prior to the actual extrapolation, the potential of this powerful algorithm can better be tapped. With that, a very high subjective as well as objective extrapolation quality can be achieved. Demonstrated for concealment of isolated block losses, the proposed Motion Compensated Frequency Selective Extrapolation leads to an almost perfect reconstruction of the signal in the lost areas. Nevertheless, further work will focus on performing the motion estimation with fractional-pel accuracy and on reducing the complexity of the algorithm by selecting several basis functions in an iteration step. In addition to that, an implementation into a state-of-the-art video decoder is planned in order to evaluate the performance of the proposed algorithm in combination with more complex loss patterns in a realistic scenario. Although the MC-FSE is introduced for concealment of block losses it can as well be used for other signal extrapolation tasks as e. g. the prediction in video coding.

REFERENCES

[1] T. Stockhammer and M. H. Hannuksela, "H.264/AVC video for wireless transmission," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 6–13, Aug. 2005.

[2] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–977, May 1998.

[3] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 470–477, April 1995.

[4] X. Li and M. T. Orchard, "Novel sequential error-concealment techniques using orientation adaptive interpolation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 857–864, Oct. 2002.

[5] 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)*, April 1993, pp. V417–V420.

[6] J. Zhang, J. F. Arnold, and M. F. 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.

[7] K. Meisinger and A. Kaup, "Spatiotemporal selective extrapolation for 3-D signals and its applications in video communications," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2348–2360, Sept. 2007.

[8] J. Seiler and A. Kaup, "Fast orthogonality deficiency compensation for improved frequency selective image extrapolation," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, March 2008.

[9] A. S. Bopardikar, O. I. Hillestad, and A. Perkis, "Temporal concealment of packet-loss related distortions in video based on structural alignment," in *Proceedings of the Eurescom summit 2005*, Heidelberg, Germany, April 2005.