Error Locating for plausible Wyner-Ziv Video Coding using Turbo Codes

Ralph Hänsel, Erika Müller

Institute of Communications Engineering, University of Rostock
Richard-Wagner-Straße 31, 18119 Rostock, Germany
{ralph.haensel, erika.mueller}@uni-rostock.de

***Abstract*—Distributed video coding (DVC, Wyner-Ziv coding) gained a lot of interest in this research decade. The major application of DVC is low complexity video encoding, which is well investigated in the literature. Generally, a distributed video coding system uses a feedback channel for rate control, otherwise the data send cannot be used for improved reconstruction quality in most cases. The feedback channel is the biggest handicap of DVC for most real world scenarios.**

In this paper, we are studying image reconstruction techniques, to benefit from all data received, also in cases where the conventional Slepian-Wolf (SW) decoding fails. Therefore, we propose error locating coding (ELC) and subsequent image inpainting (ImIp) as new Wyner-Ziv decoding method. ELC and inpainting do not rely on a feedback channel, hence enabling feedback channel free decoding. Simulation results show, that the proposed approaches improve the PSNR by 1dB. Furthermore, no transmission overhead is introduced due to the ELC.

***Index Terms*—Slepian-Wolf Coding, Wyner-Ziv Coding, image inpainting, error locating coding**

I. Introduction

Distributed video coding (DVC) became more and more important in the recent years. It is based on the theories of D. Slepian, J. Wolf [1] and A. D. Wyner, J. Ziv [2]. In contrast to conventional video coding systems (MPEG-4, H.264/AVC), distributed video coding gives the ability to develop low complexity encoders. Error-robust transmission and multiview video compression are further application fields.

The known DVC systems, in the literature, use a feedback channel. But the feedback channel is a problem in practical applications e.g. storage, unidirectional streaming or real-time decoding. Typically, a feedback channel is necessary to allocate the correct bit rate for successful Slepian-Wolf decoding, hence reaching high RD performance. Furthermore bit rate estimation at the encoder [8] can replace the feedback channel, but will not ensure successful decoding. In this paper we propose Wyner-Ziv decoding methods for non feedback channel operation, based on error locating and inpainting.

The basic setup of a low encoder complexity codec (fig.1) as used in e.g. [3], [4] and [5], consists of five major components. The key frame encoder processes every second frame using

MMSP'09, October 5-7, 2009, Rio de Janeiro, Brazil.
978-1-4244-4464-9/09/$25.00

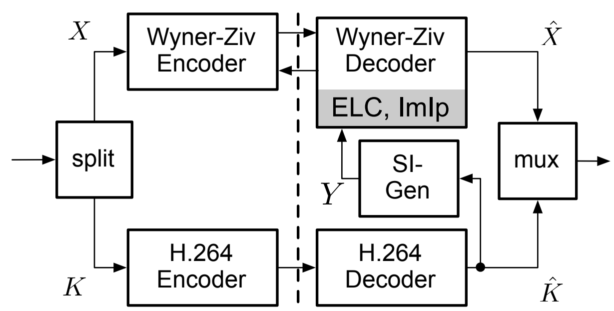


Fig. 1. The proposed DVC architecture

a H.264/AVC intra or JPEG2000 encoder. The key frames (K) are decoded by the key frame decoder and stored in the output sequence. Furthermore, the decoded key frames are used for temporal interpolation (SI-Gen, BiMESS,[6]) to generate an estimate (Y) of Wyner-Ziv frame (X). The pixel-domain Wyner-Ziv encoder performs quantisation and Slepian-Wolf (SW) encoding. The parity bits generated by the SW encoder are stored and send to the decoder on request. They are used to correct some errors in the side information (Y), obtaining the reconstructed WZ frame ($\hat{X}$).

In such a system the feedback channel is mandatory for rate allocation, because parity bits send, are only useful for the decoder if its number reaches a critical limit. Beneath this limit SW decoding fails and thus the quality of reconstructed WZ frames is low. In this paper we develop error concealment methods, which are used for decoding beneath the critical limit. Our algorithms do not rely on a feedback channel. We propose to introduce error locating coding (ELC) and image inpainting (ImIp) as new decoding methods in the WZ decoder (fig. 1). ELC works beneath the critical limit (for successful SW decoding). Thus, the inpainting generates a subjective high quality reconstructed image based on the error map ($emap$) obtained by ELC. We studied the use of inpainting for plausible decoding on a fixed error map in [7]. In this paper, we propose to use a non fixed error map (generated by ELC) and thus allows a feedback channel free operation. The decoder can decide whether to use SW or ELC decoding. Therefore, the feedback channel is still included in figure 1.

Error concealment techniques (e.g. inpainting) typically depend on the known error positions ($emap$), therefore some error locating techniques are reviewed in sec. II-B. Error concealment is already used in distributed video coding (sec. II-A), furthermore image inpainting is well studied in the literature (sec. II-C). Our proposed system is based on different error location techniques, which are: convolutional code based ELC, SI-generation based locating and a mixed method (sec. III). The proposed scheme increases the subjective (HVS) and the objective image quality (PSNR) (sec. IV) in case of successful and unsuccessful SW decoding. Furthermore, conclusions are given at the end of this paper (sec. V).

II. Related Work

A. Error Concealment in DVC

Error concealment is used in conventional [9] and distributed video coding schemes [10] to conceal distortion due to transmission errors. Typically, the location of an error is known by data from the transport layer. Subsequent error concealment exploits spatial and temporal correlation. For spatial concealment edge adaptive filters are used in [10]. These filters produce sharp edges and thus a high quality image for the human visual system (HVS). Hence, the subjective impression of the reconstructed image is more important than an objective metric (e.g. PSNR, SAD) for concealment techniques. Furthermore, no feedback is needed when using concealment techniques, which is the major benefit. So far, error concealment was only applied in case of transmission errors. In contrast, this paper is focused on error concealment for decoding.

B. Error Locating Coding

Channel coding is widely used in modern communication systems for forward error correction (FEC). It is also an essential part of DVC systems, because the SW Codec is typically based on a Turbo or a LDPC Code (FEC). Furthermore, channel coding techniques are used for error detection (EDC, e.g. rate control in DVC). The decoder can flexibly decide which mode to use. For example a Hamming code has the ability to work as FEC and EDC code.

Error Locating Coding (ELC) is the third class of channel coding methods. ELC is not as powerful as FEC thus it can not correct errors, but it indicates the location of an error. Special ELC are proposed in [11] and [12]. An ELC works well on burst error channels. Typically, the transmitted sequence is divided in segments and each segment is checked for errors. A lower code rate R_c (higher data rate) leads to smaller segments and thus finer error location. If one have binary symbols and an exact error location (1 bit per segment), ELC and FEC converge.

Error locating coding, with channel codes also used for SW coding, was proposed in [13],[14]. A LDPC code was applied to detect (in a watermarking scenario) the location of the introduced tampering. A two state channel model (tampered / non tampered) was used, where the likelihood of each state is estimated on a block basis.

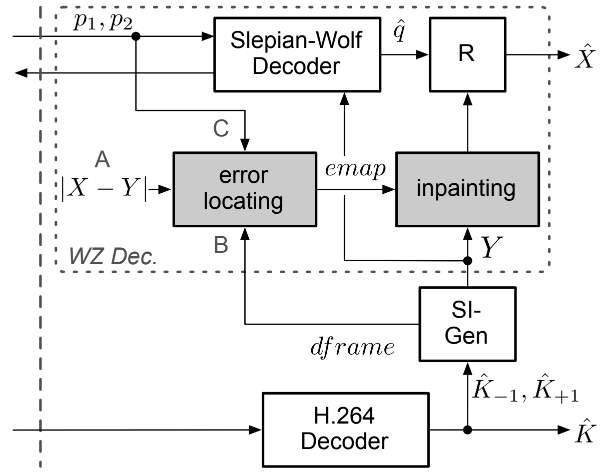


Fig. 2. The proposed Wyner-Ziv Decoder

Up to now, there is no ELC using highly punctured turbo codes (TC). These codes are only used for SW coding in most DVC Codecs. ELC by turbo codes is the focus of this paper. We propose to use the same channel codes (TC) for high performance SW coding and fine-grained ELC.

C. Image inpainting

Image inpainting is a technique to fill unspecified areas in an image [15],[16]. This includes e.g. the removal of text, scratches or real objects. Furthermore, it is able to reconstruct an image from incomplete data. One algorithm for image inpainting [16] solves a partial differential equation (PDE) to obtain the inpainted image. This produces sharp edges and thus a high quality image for the human visual system (HVS).

The application of image inpainting to DVC was proposed in [7]. It was shown, that inpainting has the ability to exploit spatial correlation in a pixel-domain DVC codec. Our proposed method includes the WZ encoding of a subsampled frame, where the missing pixels were reconstructed by image inpainting. This leads to a plausible decoding with high subjective quality.

III. Proposed Wyner-Ziv Decoder

The proposed pixel-domain Wyner-Ziv decoder includes an error locating (ELC) and image inpainting (ImIp) algorithm besides the conventional Slepian-Wolf decoder and frame reconstruction (R) (fig. 2). The aim is to avoid the feedback channel. ELC and inpainting do not require feedback channel data and will improve the quality of the reconstructed frame $\hat{X}$. Furthermore, WZ decoding is performed below the critical rate (for successful SW decoding). The proposed approach is the first Wyner-Ziv decoder which can benefit from all transmitted data, without mandatory feedback channel. But it is meaningful to use a feedback channel to switch the SW decoder to the next bit plane for improved quality (see sec. IV-F).

Error locating coding (ELC) is used to generate an error map ($emap$). The granularity depends on the data rate transmitted.

ELC works well in case of block errors, for this reason a block by block pixel scan order is applied. Pixels in an erroneous segment are replaced by inpainting, based on the error map. Subsequently the inpainted side information is used for reconstruction. The partly decoded quantisation symbols $\hat{q}$ guide the reconstruction process. This improves the quality of the reconstructed frame, also if the data rate is below the critical rate for SW decoding.

The proposed decoding method follows the idea of our plausible [7] and flexible [17] decoding, which gives the decoder more flexibility to handle the received data.

The error map ($emap$) is achieved by four different methods. The first one uses an error map oracle (A). The second error location algorithm (B) is based on the $dframe$ (difference between motion compensated key frames) provided by the side information generation (BiMESS) and thus is applicable in a real-world scenario. Algorithm C generates the error map based on the parity bits received and algorithm D combines the error maps of B and C.

A. Error Map Oracle

The error map oracle generates the error map by thresholding the magnitude of the residual frame (difference between original frame X and side information Y) as shown in eq. 1, where th is the threshold. We found good performance at a value of $th = 10$. This method is used as reference and is no real-world solution, because the residual frame is not available at the decoder side.

$$emap_A(x,y) = \begin{cases} 1 & : \quad |X(x,y) - Y(x,y)| > th \\ 0 & : \quad else \end{cases} \tag{1}$$

B. dframe Error Locating (EL)

The *dframe* EL is similar to the oracle method, but uses the $dframe$ [18] provided by the side information generation (BiMESS). The $dframe$ is the difference between the motion compensated neighboring key frames ($\hat{K}_{-1}, \hat{K}_{+1}$, eq. 2). The error map is generated by thresholding the magnitude of the dframe (eq. 3). In contrast to the error map oracle (A), it is only based on data available at the decoder.

$$\begin{aligned} dframe(x,y) &= \hat{K}_{-1}(x-dx, y-dy) \\ &\quad -\hat{K}_{+1}(x+dx, y+dy) \end{aligned} \tag{2}$$

$$emap_B(x,y) = \begin{cases} 1 & : \quad |dframe(x,y)| > 2 \cdot th \\ 0 & : \quad else \end{cases} \tag{3}$$

C. Error Locating Code (ELC)

Parallel concatenated convolutional codes (turbo codes) are used for SW encoding/decoding in our setup. The same codes are also used for error locating, so there is no transmission overhead. Typically, an error correction code (FEC) can also be used for error detection (EDC). For example, a bit plane of the side information $Y^{(b)}$ can be re-encoded at the decoder with the same channel code used at the encoder. The comparison of the parity bits indicates, whether there is an error in the side information or not. Error location is done by dividing the bit plane into segments and apply an error check on each segment. Thus the approximate location of the error is detected.

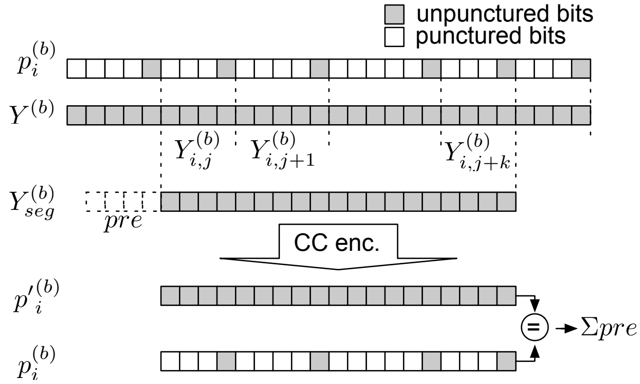


Fig. 3. Segment-based error check for error locating (ELC)

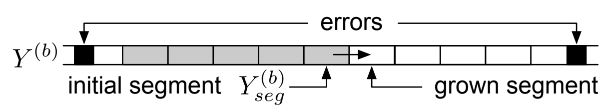


Fig. 4. Segment growing between erroneous bits

The bit plane of the side information is split up into micro segments $Y_{i,j}^{(b)}$ for error localisation. The size and location of the micro segments is determined by the puncturing scheme of the corresponding parity bits $p_i^{(b)}$, as shown in figure 3.

In the next step, a segment of side information bits $Y_{seg}^{(b)}$ is re-encoded by the convolutional encoder (CC enc.). A minimum segment contains 7 micro segments. The obtained parity bits $p'^{(b)}_i$ are compared to the received parity bits $p_i^{(b)}$. Furthermore, a preamble pre is added (fig. 3), because prior accurate information bits are unknown. The preamble contains m bits (m - memory length of the CC encoder) and thus $pre \in [0 \ldots 2^m - 1]$. A segment is marked erroneous only if the parity bits are unequal for all possible preambles. See equation 4, where Σpre is the number of right re-encoded preambles.

Segment growing is applied, if an initial error free segment is found (fig. 4). Micro segments are added subsequently at the end of the segment. Thus, the segment is grown between the error positions. The number of micro segments in a segment is equal to the number of unpunctured parity bits in the segment. The more parity bits in a segment the more robust is the error detection code. Thus, it is important to build large segments.

The granularity (size of micro segments) of the error location is defined by the puncturing scheme, because each micro segment contains one parity bit. Thus, the more parity bits send by the encoder the finer the puncturing scheme, the smaller the micro segments and the more accurate the error locating is. Hence the performance of the ELC scales with the amount of parity bits send by the encoder.

The described error location algorithm generates an error map ($emap_{Ci}$) for both parity bit chains ($p_1^{(b)}, p_2^{(b)}$). The maps are combined in a way, that only pixel are marked as erroneous, if both parity chains indicates an error (eq. 5).

$$emap_{Ci}(x,y) = \begin{cases} 1 & : \ \Sigma pre = 0 \\ 0 & : \ else \end{cases} \quad (4)$$

$$emap_C = emap_{C1} \wedge emap_{C2} \quad (5)$$

D. Fusion of dframe and ELC

The fourth method of our error map generation combines the dframe EL (B) and ELC (C) to get a fine-grained error map. Only erroneous pixels in both error maps ($emap_B, emap_C$) are marked as errors in the merged map ($emap_D$, eq. 6).

$$emap_D = emap_B \wedge emap_C \quad (6)$$

IV. Simulation Results

Simulations are based on the setup described in section III. The sequences foreman, coastguard and soccer were used (QCIF, 30fps). Furthermore, a GOP size of 2 was used (KWKW, one Wyner-Ziv frame between key frames). The key frames are encoded and decoded by a H.264/AVC intra coder. The quantisation parameter (QP) was adjusted, that all key frames have an average quality of 30 dB. Furthermore, all Wyner-Ziv frames are encoded at a constant data rate. The feedback channel is used to switch to the next bit plane, but not for rate allocation.

The aim of the proposed approach is to improved the quality of the reconstructed frame below the critical rate for SW decoding, without feedback channel. It achieves plausible decoding results, with high quality for the human visual system. Therefore, the results are evaluated based on the PSNR metric and also some visual results are exemplarily presented, because PSNR is not well correlated with the human visual system (HVS).

The initial side information Y (fig. 5) is estimated based on the lossy coded neighboring key frames. Block artifacts and double edges are the most annoying distortions for the HVS.

A. Error Map Oracle

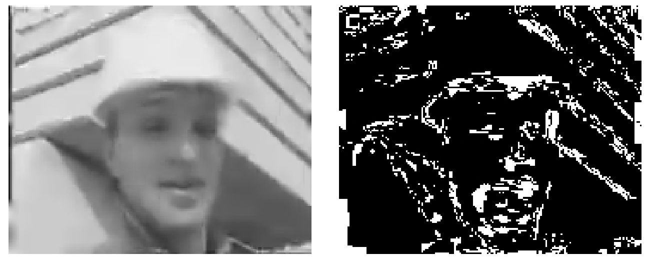
Figure 6(b) shows the error map used for inpainting by the error map oracle. The double edges and some block artifacts are removed in the inpainted side information Y_{oracle} (fig. 6). Therefore, the quality is improved in contrast to the original side information in PSNR and to the HVS. This algorithm can not be used in a practical setup, because the error map is not available at the decoder. It shows the performance of the image inpainting.

B. dframe Error Locating (EL)

The dframe EL works like the oracle algorithm, but it only depends on data available at the decoder. The error map and inpainted side information is shown in figure 7. The quality to the HVS and the PSNR is still improved. This algorithm does not need extra transmission capacity and rely on data available at the decoder. It visually improves the reconstruction quality at non-transmission overhead.

(a) Y, initial side information (b) Original Frame

Fig. 5. Initial side information Y, foreman, frame 2, $PSNR = 26,34$ dB



(a) Y_{oracle} (b) $emap_A$

Fig. 6. (a) Inpainted side information, foreman, frame 2, $PSNR = 26,52$ dB, (b) error map

(a) Y_{dframe} (b) $emap_B$

Fig. 7. (a) Inpainted side information, foreman, frame 2, $PSNR = 26,59$ dB, (b) error map

C. Error Locating Coding (ELC)

The figures 8 and 9 show the error maps generated by segment growing in the first bit plane (MSB - most significant bit). Error map $emap_{C1}$ has a blocky structure because of the pixel scan order, whereas $emap_{C2}$ has a random structure because of the random interleaver in the turbo code. The error locating coding is weak, thus it is not used directly for image inpainting.

Furthermore, the granularity of the error location becomes finer, if the data rate increases. Thus, it scales with the data rate transmitted by the encoder and do not need a feedback channel, which is the major advantage. The performance of the ELC is increased at higher data rates. Furthermore, the ELC uses the same parity bits the SW decoder does. Therefore, no transmission overhead is introduced in the system. The proposed algorithms are only implemented in the Wyner-Ziv decoder, thus the low complexity of the Wyner-Ziv encoder is kept.

(a) $emap_{C1}$ (b) $emap_{C2}$

Fig. 8. Error maps generated by ELC, $rate = 29,00\,\mathrm{kbps}$

(a) $emap_{C1}$ (b) $emap_{C2}$

Fig. 9. Error maps generated by ELC, $rate = 69,60\,\mathrm{kbps}$

D. Fusion of dframe and ELC

Fusion EL combines the error maps of the dframe EL and ELC method. Figures 10 and 11 show the combined error maps and the inpainted side information at different data rates (below critical rate for SW decoding, $81,21\,\mathrm{kbps}$, MSB, foreman, frame 2). In both cases the quality is improved compared to the initial side information and inpainted side information (Y_{oracle}, Y_{dframe}). If more data rate is available, the quality of the inpainted image is increased, because ELC scales with data rate. This behavior is plausible because more data generally leads to higher reconstruction quality.

E. Wyner-Ziv frame RD Performance

The Rate-Distortion (RD) performance of the Wyner-Ziv frames is shown in figures 12 and 13, whereas the key frames are coded at an average quality of $30\,\mathrm{dB}$. The proposed decoding method (dframe EL) increases the quality of the WZ frames by $1\,\mathrm{dB}$ for the sequence foreman and soccer. But there is also a loss for coastguard sequence.

The reconstruction quality is decreased for the sequence coastguard. Thus, an adaptive selection of the best decoding method is needed. A reason for this behavior might be, that there is less spatial correlation and higher frequencies in the sequence coastguard than in the other used sequences.

The RD performance for all sequences stagnates for high data rates. This is a common behavior of DVC Codecs. The relevant data rate range for QCIF resolution video is below $200\,\mathrm{kbps}$. In this region our proposed approach shows good performance compared to conventional pixel-domain Wyner-Ziv video decoding.

The error map oracle method outperforms the dframe EL in all cases. Futhermore there is a gap of $1\,\mathrm{dB}$ between the oracle and dframe EL method. Therefore, additionally improved quality is the scope of our ongoing research.

(a) Y_{fusion} (b) $emap_D$

Fig. 10. (a) inpainted side information, foreman, frame 2, $PSNR = 26,75\,\mathrm{dB}$, (b) error map, $rate = 29,00\,\mathrm{kbps}$

(a) Y_{fusion} (b) $emap_D$

Fig. 11. (a) inpainted side information, foreman, frame 2, $PSNR = 27,02\,\mathrm{dB}$, (b) error map, $rate = 69,60\,\mathrm{kbps}$

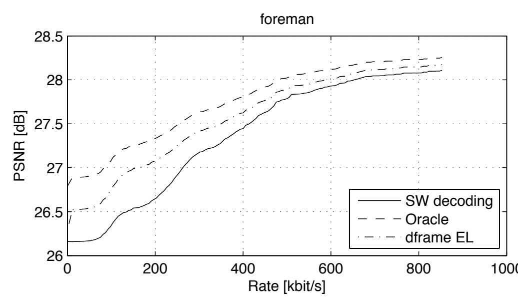


Fig. 12. Overall RD performance foreman

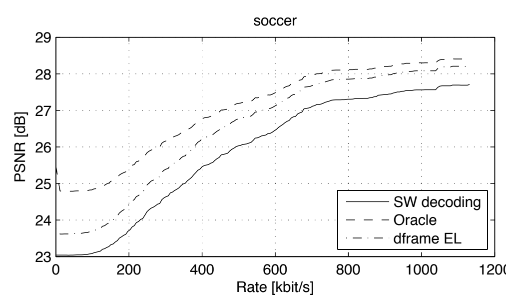
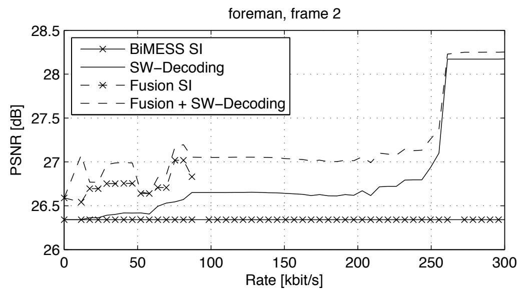


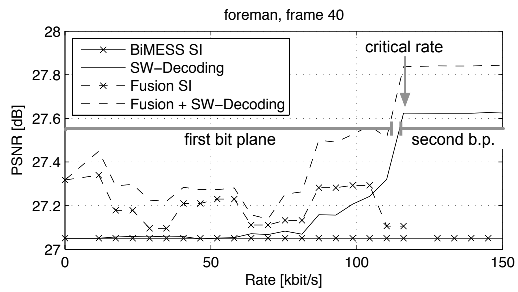
Fig. 13. Overall RD performance soccer

F. Performance of Fusion

Figure 14 shows the RD performance of the conventional BiMESS side information generation and SW decoding. Furthermore, the performance of the proposed fusion (D) side information and subsequent SW decoding is shown. The quality



(a) foreman, frame 2, critical rate $81\,\mathrm{kbps}$ (first bit plane)



(b) foreman, frame 40, critical rate $116\,\mathrm{kbps}$ (first bit plane)

Fig. 14. Side information and reconstruction quality for BiMESS-based and fusion-based SW decoding

of the BiMESS side information (SI) is worse than the fusion side information. Slepian-Wolf decoding is used without a feedback channel until the critical rate for the first bit plane. Below this rate, the reconstruction quality of the fusion method (ELC, dframe) is faster rising than the reconstruction quality of BiMESS and SW decoding. If SW decoding of the first bit plane was successful, the exact error positions are known. Therefore, ELC and inpainting cannot increase the quality additionally. Therefore, the feedback channel can be used to switch the SW decoder to the second bit plane. The quality is increased if the second bit plane is successfully decoded (SW). Thus, feedback channel free operation is meaningful only below the critical rate. Beyond the critical rate the reconstruction quality remains constant. But there is a quality offset between fusion and the SW decoding with feedback channel. So it depends on the data rate whether a feedback channel is meaningful or not.

G. Computational Complexity

The Error Map Oracle and dframe Error Locating has very low complexity. Furthermore the computational complexity of the error location process (segment growing) is as high as the complexity of standard SW decoding. The complexity of image inpainting high, but it is less complex than the SW decoding.

V. Conclusion

Wyner-Ziv decoding below the critical data rate, for Slepian-Wolf decoding, was faced in this paper. Thus a new Wyner-Ziv decoder was porposed, which additionally include turbo code based error locating and image inpainting. The novel components do not rely on a feedback channel.

Error Locating Coding (ELC) was done by the proposed segment growing algorithm. It was shown, that the proposed approach increases the subjective and objective quality (PSNR) by $1\,\mathrm{dB}$. The proposed ELC algorithm works with the same parity bits the SW decoder does. Therefore, no transmission overhead is introduced. Furthermore, it does not need a feedback channel because it scales with the data rate, which is the major advantage.

Further activities will focus on the optimization of the ELC and improved image inpainting.

Acknowledgement

I would like to thank my colleagues and students for some inspiring discussions and helpful editorial remarks.

References

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[3] A. Aaron and B. Girod, "Wyner-ziv video coding with low encoder complexity," in *Proc. Picture Coding Symposium (PCS)*, SF/USA, December 15-17 2004.

[4] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

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

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

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

[8] C. Brites and F. Pereira, "Encoder rate control for transform domain wyner-ziv video coding," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2007.

[9] J. Seiler and A. Kaup, "Adaptive joint spatio-temporal error concealment for video communication," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, 2008.

[10] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi, "Hybrid spatial and temporal error concealment for distributed video coding," in *Proc. Int. Conf. on Multimedia and Expo (ICME)*, 2008.

[11] M. Kitakami and J. Sano, "Code design and decoding methods for burst error locating codes," in *Proc. Pacific Rim Int. Symposium an Dependable Computing (PRDC)*, 2005.

[12] A. Said, "Error burst detection with high-rate convolutional codes," in *Proc. IEEE Int. Symposium on Information Theory*, 1995.

[13] Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication and tampering localization using distributed source coding," in *IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Crete, Greece, October 2007.

[14] Y.-C. Lin, D. Varodayan, T. Fink, E. Bellers, and B. Girod, "Localization of tampering in contrast and brightness adjusted images using distributed source coding and expectation maximization," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2008.

[15] M. Bertalmao, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," *Proc. SIGGRAPH*, pp. 417–424, 2000.

[16] D. Tschumperlé, "Pde's based regularization of multivalued images and applications," Ph.D. dissertation, Ecole Doctorale Sciences et Technologies de l'Information et de la Communication, 2002.

[17] R. Hänsel and E. Müller, "Simplifying the rate control scheme for distributed video coding by flexible slepian-wolf decoding," in *Proc.Pacific Rim Symposium on Image and Video Technology (PSIVT)*, 2009.

[18] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel based wyner-ziv video coding," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2006.