CONFIDENCE FEATURES EXTRACTION FOR WYNER-ZIV VIDEO DECODING

Ralph Hänsel, and Erika Müller

Institute of Communications Engineering, University of Rostock
Richard-Wagner-Straße 31, 18119 Rostock, Germany
phone: + (49) 381 498 7305, fax: + (49) 381 498 7302, email: {ralph.haensel,erika.mueller}@uni-rostock.de
web: www.int.uni-rostock.de

ABSTRACT

Distributed video coding (DVC) is well suited for low complexity encoding and error robust transmission. The feedback channel is the main handicap of state-of-the-art DVC systems. We previously proposed error locating coding (ELC) in conjunction with image inpainting as solution. This paper is focused on ELC with confidence information. It is shown that the proposed confidence feature improves the ELC performance and adapts well to the data rate.

1. INTRODUCTION

Video compression is an important topic in modern multimedia communication systems. The conventional video codecs (e.g. MPEG-4, H.264/AVC) are well suited for broadcast video transmission. Hence, the encoding is done once with a very high computational complexity, whereas every client can decode the video with low complexity. That behavior is not well suited for video encoding on mobile devices, where the encoder is very limited in computing power and energy.

Distributed Video Coding (DVC) gives the ability to design video codecs with low encoder complexity. The complex motion estimation part is shifted from the encoder to the decoder. Thus, the temporal correlation is exploited only at the decoder. The theories of D. Slepian and J. Wolf [12] as well as A.D. Wyner and J. Ziv [14] proof that distributed coding can reach the performance of conventional coding. Further applications for DVC are given in [8].

The main handicap for state-of-the-art DVC [1, 4] is the feedback channel [3]. It is necessary for rate control, to guaranty high rate distortion (RD) performance. If no feedback channel is available, state-of-the-art DVC codecs will have very low RD performance. Data transmitted to the decoder will not increase the decoding quality, until it reaches a critical limit. The Slepian-Wolf (SW) decoder only decodes a valid quantisation symbol beyond the critical rate.

SW decoding can be seen as forward error correction (FEC, turbo code). FEC only works beyond the critical rate. In contrast, Error Locating Coding (ELC) adapts to the data rate. Depending on the rate it is more or less fine grained or reliable. Thus ELC is working beneath the critical rate without a mandatory feedback channel. In [5] we proposed ELC to mark unreliable regions in a frame, which were filled in by inpainting, obtaining the reconstructed frame. In this paper we will focus on an improved error locating and confidence feature extraction.

An overview of the proposed DVC codec is given in the next section, followed by a review of the related work in section 3. The proposed error locating methods and confidence features are described in section 5. The corresponding results and conclusion are given in section 6 and 7.

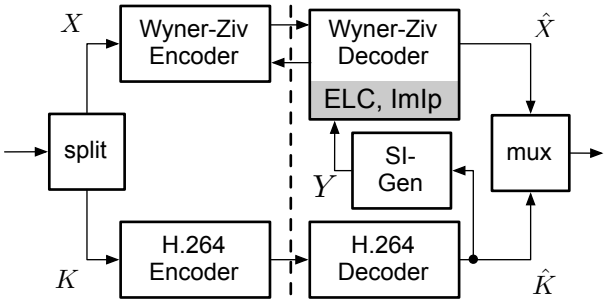


Figure 1: The proposed DVC architecture

2. SYSTEM

The input sequence is split up into key frames K and Wyner-Ziv (WZ) frames X at the encoder (fig. 1). Whereas the key frames are encoded by a conventional H.264 intra encoder, the WZ frames are encoded by the WZ encoder. The pixel domain WZ encoder follows the DVC principles and has very low complexity. It quantizes every pixel and does SW encoding (based on systematic turbo code [9]). The decoded key frames are used for temporal interpolation (BiMESS,[2]) at the decoder, gaining the side information Y. The side information is corrected by the WZ decoder to get the decoded WZ frame $\hat{X}$. On the one hand, the decoder can request parity bits from the encoder until the critical rate for successful decoding is reached. On the other hand, the proposed error locating and inpainting can be applied to generate the reconstructed frame, whereas no feedback channel is needed.

Typically, a DVC codec uses a DCT transform to improve the RD performance. But transform domain DVC has two drawbacks: at first the encoder complexity is increased and secondly, the rate control and thus the feedback channel problem is more complex. The rate control has to operate on every band (DCT). Therefore, we propose a pixel domain DVC codec.

The major improvements in the proposed system are the error locating coding (ELC) and the image inpainting (ImIp) (see fig. 2). During the decoding process the SW decoder is observed by the error locating module, which aims to detect error-prone pixels. This information is used in an image inpainting process, which improves the quality of the side information Y and the quality of the reconstructed WZ frame $\hat{X}$. In [5] we were focused on the ELC and ImIp. But pixel are only classified in erroneous and non erroneous pixel. In this paper we propose ELC algorithms which give an additionally confidence information. No extra data is transmitted for ELC, because it only observes the SW decoding process.

 ISSN 2076-1465

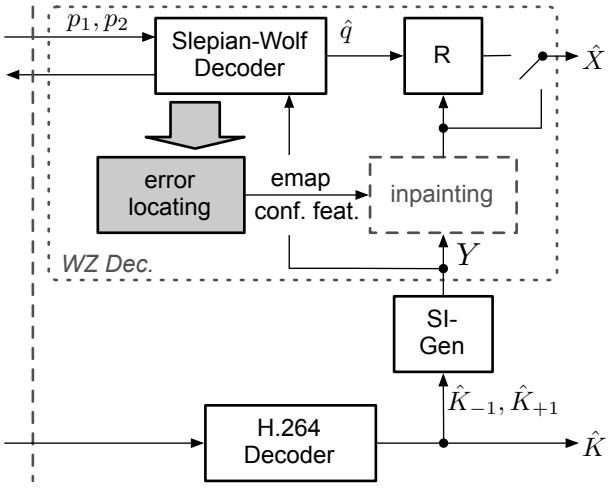


Figure 2: DVC Decoder consisting of SW Decoder and the proposed error location with confidence feature extraction

3. RELATED WORK

3.1 Error Locating in Distributed Source Coding

Error location in conjunction with distributed source coding (DSC) was proposed in [7]. Here, the DSC decoder locates the tampering in a watermarking scenario. A two state channel model (tampered / non tampered) is applied, where the likelihood of each state is estimated on a block basis. The underlying SW coder is implemented by an LDPC code, which is as powerful as a turbo code.

3.2 Stopping criteria for Turbo Codes

In [11] several stopping criteria for turbo codes are evaluated in conjunction with DVC. The entropy coding in DVC is done by the SW coder, which is implemented by e.g. LDPC or turbo code. For a high RD performance of the codec, it is important to know, whether SW decoding was successful or not. Based on the evaluated stopping criteria the decoder can decide whether to request more data from the encoder or to stop decoding. The modified stopping criteria are the basis for the proposed confidence feature. Therefore, the criteria are reviewed briefly.

The sign-change ratio (SCR) criterion [10] counts the sign-changes in the extrinsic information L_{2e}. If there are only some sign-changes, the decoding is mentioned to be error free and is stopped.

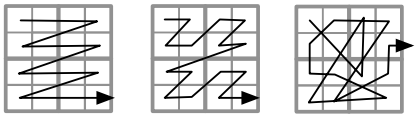
The hard-decision-aided (HDA) criterion [10] counts the sign-changes in the a posteriori information L_2 (hard output) between subsequent iterations and also stops decoding if the number of sign-changes is low.

The sign-difference ratio (SDR) [13] does not compare the sign in subsequent iteration, but compare the sign of the a priori L_{2a} and extrinsic information L_{2e}. Also the decoding is stopped if only a few sign-changes occur.

Account of the magnitude of the a posteriori information L_2 is taken by the mean-estimate (ME) criterion [15]. If the mean of the magnitude of the a posteriori information $|L_2|$ is greater than a threshold, the decoding is considered successful.

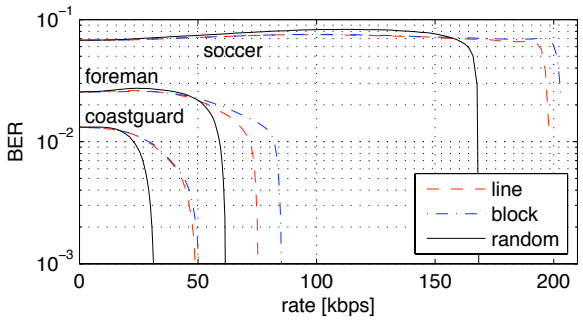
The fifth criterion is used in the DISCOVER codec [6]. If the number of symbols, with the a posteriori information $|L_2|$ is below a threshold, decoding is stopped. Additionally a CRC is applied.

In our proposed error location and confidence estimation algorithm we use the mentioned stopping criteria in a modified form to estimate, whether a bit is correct decoded or not. Therefore, we are aiming to get a confidence information for each bit and not only for the decoded whole frame.



(a) line (b) block (c) random

Figure 3: Pixel scan order



(a) BER vs. data rate

Sequence	line	block	random
foreman	0,204 bpp 77,5 kbps	0,230 bpp 87,6 kbps	0,167 bpp 63,5 kbps
coastguard	0,134 bpp 50,9 kbps	0,138 bpp 52,5 kbps	0,086 bpp 32,6 kbps
soccer	0,536 bpp 203,6 kbps	0,551 bpp 209,5 kbps	0,456 bpp 173,4 kbps

(b) critical data rate

Figure 4: Bit Error Rate (BER) vs. Data Rate, Wyner Ziv Frames - block, line and random pixel scan order (coastguard, foreman, soccer, QCIF, 30fps)

4. PIXEL SCAN ORDER IN DVC

In a pixel-domain Wyner-Ziv encoder the 2D frame needs to be transformed into an one-dimensional vector for SW encoding. The pixel are typically scanned line by line (fig. 3(a)). In [5] block by block (8×8) pixel scan order (fig. 3(b)) was proposed for improved error locating capabilities. Furthermore, random pixel scan order (fig. 3(c)) is also possible (same performance as a second TC interleaver).

The bit error rate (BER) vs. data rate of the first bit plane is shown in figure 4(a). One can see, that the BER is not significantly decreased if the data rate is increased. At least when the critical rate (fig. 4(b)) is reached the BER will drop to zero. Furthermore, if random pixel scan order is used, successful decoding is possible at the lowest rate. The random scan of pixel will scatter block errors, which produces a good error pattern for turbo code based SW decoding. Therefore, we will apply the random pixel scan order in this paper. The pixel order is fixed and don't need to be transmitted.

5. PROPOSED ERROR MAP AND CONFIDENCE FEATURE ESTIMATION

The stopping criteria reviewed in section 3 give information, whether the decoding was successful or not. Successful decoding means, that the bit plane u_k of the original frame X is equal to the bit plane $\hat{u}_k$ of the decoded frame $\hat{X}$. If the reviewed criteria can give information whether the whole sequence is decoded successfully, it can also give information about successful decoding of each single bit $\hat{u}_k$.

The key idea of the proposed algorithm is to generate an estimated error map $emap_{est}$ in the first step, by comparing the SW decoder output bits with the input bits. In a second, step a modified stopping criterion (confidence feature) is estimated to separate the predicted error map into reliable and non reliable segments. This two step approach improves the error locating performance.

The proposed error map prediction and confidence feature estimation does not rely on a transmission overhead, because it only uses information available from the turbo decoder (SW decoder). Furthermore, the WZ encoder is left untouched, preserving the low encoder complexity.

5.1 Error Map Estimation

The estimated error map $emap_{est}(k)$ is generated by comparing the hard SW decoder output $\hat{u}_k$ with the hard decoder input $y_k^{(b)}$ (eq. 1). The hard decoder input is the corresponding bit plane b of the side information Y. For performance evaluation the real error map $emap_{real}(k)$ defined in equation 2 is also calculated.

As shown in figure 4(a), the BER is not decreased before the critical rate is reached. So the number of errors in the decoder input $y_k^{(b)}$ (fig. 4(a),data rate 0 kbps) and output $\hat{u}_k$ is nearly the same. But the erroneous bits are in different positions. Therefore, the predicted error map is generated by comparing the hard decoder input and output.

$$emap_{est}(k) = \hat{u}_k \neq y_k^{(b)} \quad (1)$$

$$emap_{real}(k) = u_k \neq y_k^{(b)} \quad (2)$$

5.2 Confidence Feature Estimation

The estimated confidence feature is used to separate the estimated error map into true positives (tP, estimated and real error position) and false positive (fP, estimate error but no real error). The confidence feature generation is based on the modified stopping criteria. In this section we present several confidence features, which are evaluated in section 6.

At first, we focus on the sign-change stopping criteria. The SCR, HDA and SDR criteria are used to generate a sign-change vector for each iteration $sc_{criterion}^{(i)}(k)$. A maximum number of 10 iteration is used (eq. 3-5).

$$sc_{SCR}^{(i)}(k) = \text{sgn}\left\{L_{2e}^{(i)}(u_k)\right\} \neq \text{sgn}\left\{L_{2e}^{(i+1)}(u_k)\right\} \quad (3)$$

$$sc_{HDA}^{(i)}(k) = \text{sgn}\left\{L_{2}^{(i)}(u_k)\right\} \neq \text{sgn}\left\{L_{2}^{(i+1)}(u_k)\right\} \quad (4)$$

$$sc_{SDR}^{(i)}(k) = \text{sgn}\left\{L_{2a}^{(i)}(u_k)\right\} \neq \text{sgn}\left\{L_{2e}^{(i)}(u_k)\right\} \quad (5)$$

Where $L_{2a}^{(i)}(u_k)$, $L_{2e}^{(i)}(u_k)$ and $L_{2}^{(i)}(u_k)$ are the a priori, extrinsic and a posteriori information of the bit u_k in the i-th iteration.

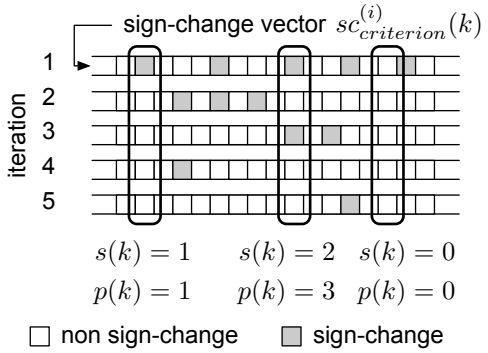


Figure 5: Calculation of the sum feature $s_{criterion}(k)$ and position feature $p_{criterion}(k)$ based on the sign-change vector $sc^{(i)}(k)$

Based on the sign-change vector $sc_{criterion}^{(i)}(k)$ two confidence features for each criterion are extracted. On the one hand the sum of sign-changes $s_{criterion}(k)$ (eq. 6) is calculated for a bit u_k over all iterations. On the other hand the iteration of the last sign-change $p_{criterion}(k)$ (position, eq. 7) is extracted (fig. 5). Corresponding to the stopping criteria fewer sign-changes or only sign-changes in first iteration indicate lower error probability and thus a reliable error map.

$$s_{criterion}(k) = \left|\left\{i; 1 \leq i \leq 10, sc_{criterion}^{(i)}(k) = true\right\}\right| \quad (6)$$

$$p_{criterion}(k) = \max_i \left\{i; sc_{criterion}^{(i)}(k) = true\right\} \quad (7)$$

Based on the ME criterion, the $me(k)$ feature (eq. 8) is calculated. It separates the error map into $n = 11$ segments, corresponding to the absolute value of the a posteriori information after the last iteration $|L_2(u_k)|$. The greater the absolute value of the a posteriori information, the higher the probability that the decoded bit is equal to the original one $\hat{u}_k = u_k$ and thus the estimated error map is right.

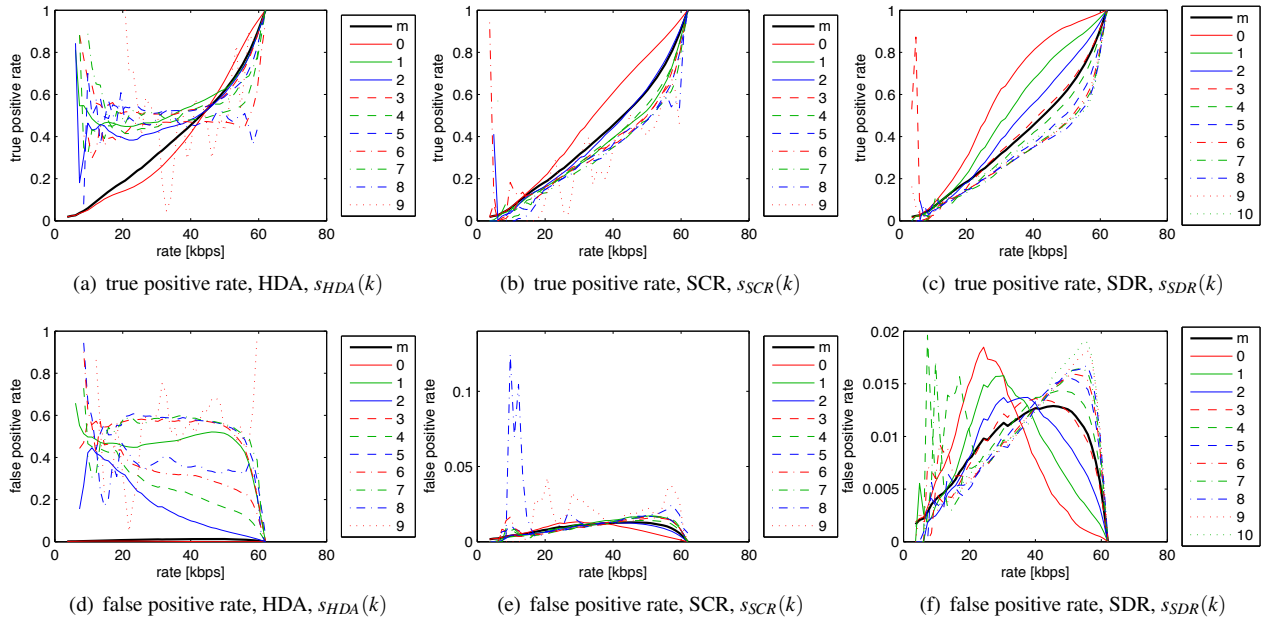
$$me(k) = \{r; L_{border,r} \leq |L_2(u_k)| \leq L_{border,r+1}\} \quad (8)$$

$$L_{border,r} = a \sum_{j=1}^{r} b^j \qquad b = 1.5$$

$$a = \frac{\max\{|L_2(u_k)|\}}{\sum_{j=1}^{n} b^j}$$

We used a smaller interval size for small L-values $|L_2(u_k)|$ than for bigger ones, because we want to have a more fine-grain confidence feature in unreliable regions and coarse-grain feature in reliable regions.

The proposed confidence features are the sum sign-change features ($s_{HDA}(k)$,$s_{SCR}(k)$,$s_{SDR}(k)$), the position features ($p_{HDA}(k)$,$p_{SCR}(k)$,$p_{SDR}(k)$) and the ME feature ($me(k)$). The features divide the error map into 10 (HDA, SCR) or 11 (ME, SDR) segments. The error map in every segment is more or less accurate, thus the feature indicates whether the error map is accurate or not. The significance of each feature is analysed in the simulation results (sec. 6).



(a) true positive rate, HDA, $s_{HDA}(k)$ (b) true positive rate, SCR, $s_{SCR}(k)$ (c) true positive rate, SDR, $s_{SDR}(k)$

(d) false positive rate, HDA, $s_{HDA}(k)$ (e) false positive rate, SCR, $s_{SCR}(k)$ (f) false positive rate, SDR, $s_{SDR}(k)$

Figure 6: HDA, SCR and SDR sum feature $s_{criterion}(k)$, foreman, QCIF, 30fps (colored lines $\rightarrow$ feature values, eq. 6)

6. SIMULATION RESULTS

The simulation results are based on the QCIF sequences coastguard, foreman and soccer at 30 fps and a KWK GOP structure.

At first the sum feature $s_{criterion}(k)$ for the modified HDA, SCR and SDR criteria are evaluated. In figure 6(a-c) the true positive rate $P[emap_{est} = 1|emap_{real} = 1]$ for the foreman sequence is shown. The black line (m) indicates the true positive rate without applying any feature, whereas the other colored lines depend on the feature value (see legend). For the HDA criterion (fig. 6(a)) there is a big difference in the true positive rate depending on the feature value. But the performance (higher true positive rate) is not increased with an higher data rate. The SCR criterion (fig. 6(b)) shows only small capability to separate reliable error map segments and non reliable, because the true positive rate is nearly independent of the feature. At least, the SDR criterion (fig. 6(c)) shows good properties, because the sum feature can separate reliable and non reliable error map segments. Furthermore, the performance of this feature is increased with an increased data rate.

On the other hand the false positive rate $P[emap_{est} = 1|emap_{real} = 0]$ (error estimate but no real error, fig. 6(d-f)) is also important. The SDR sum feature also shows the best performance, because the error map is divided into segments with high difference in the false positive rate, depending on the feature. Furthermore, the feature values $0 \ldots 2$ shows high true positive rate and low false positive rate for medium to high data rates. This is a very good property. For low data rate the SDR sum feature and the estimated error map is not accurate. Therefore, the error map estimation needs to be improved before an accurate feature can be selected. In conclusion a low value of the feature indicates a good error map and thus high confidence.

In contrast to the sum feature $s_{criterion}(k)$ the performance of position feature $p_{criterion}(k)$ is slightly worse. From a computational complexity perspective sum and position feature are equal, therefore we propose to use the SDR sum feature.

The ME feature is only dependent on the absolute value of the a posteriori information $L_2(u_k)$. A higher feature value corresponds to a higher L-value and thus to a more reliable decoding and error map. The true positive, false positive and bit error rate (BER) are shown in figure 7. The true positive rate depends on the feature value, but for a medium data rate all the curves intersect. Therefore, there is no advantage of this feature for medium data rates (40 kbps, foreman). Furthermore, the feature is well suited to detect segments with high BER (fig. 7). That is outstanding for the evaluated features, because non of the other features is well suited to separate low and high BER regions.

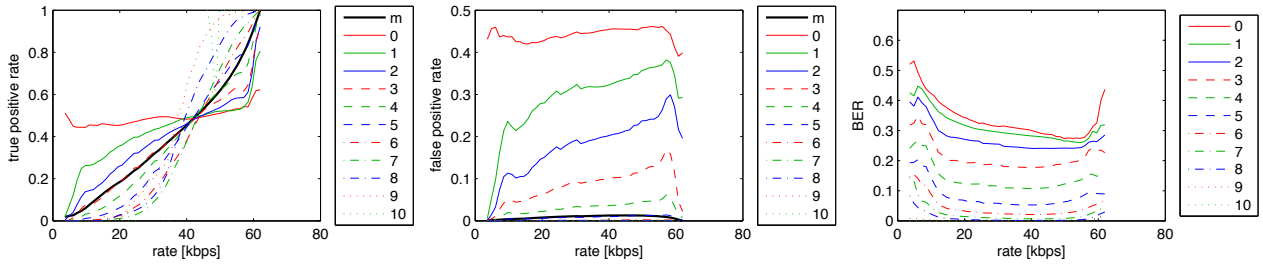
Figure 8 shows the true positive rate for the sequences coastguard and soccer. The SDR sum feature and ME feature shows the same behavior as for the foreman sequence. Thus, the major properties of the features are independent of the video sequence.

At least, the pixel scan order has also impact on the features presented here. The random pixel scan order shows the best performance of the confidence estimate features. Thus we recommend it for ELC and SW coding. Furthermore, the SW coding performance do not suffer from additional ELC coding (no transmission overhead, no negative constrains).

7. CONCLUSIONS

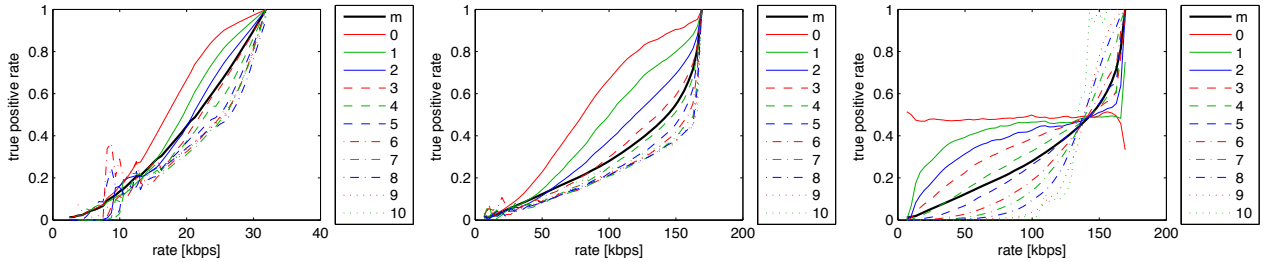
An error map and confidence feature estimation for turbo code based DVC were proposed in this paper. It was shown that the proposed SDR sum confidence feature gives the best performance for true positive detection. Furthermore, the ME feature is best suited for BER classification. In conjunction with image inpainting the proposed algorithms will help to eliminate the feedback channel in future DVC systems.

Further work will include the research on a modified image inpainting algorithm, which can handle confidence infor-



(a) true positive rate (b) false positive rate (c) bit error rate

Figure 7: ME feature $me(k)$, foreman, QCIF, 30fps (colored lines $\rightarrow$ feature values, eq. 6)



(a) coastguard, SDR, $s_{SDR}(k)$ (b) soccer, SDR, $s_{SDR}(k)$ (c) soccer, ME, $me(k)$

Figure 8: true positive rate, SDR sum $s_{SDR}(k)$ and ME $me(k)$ feature, QCIF, 30fps (colored lines $\rightarrow$ feature values, eq. 6)

mation to improve the quality of the reconstructed frames.

REFERENCES

[1] 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.

[2] J. Ascenso, C. Brites, and F. Pereira. Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding. In *Proc. EURASIP*, Slovak Republic, July 2005.

[3] C. Brites, J. Ascenso, and F. Pereira. Feedback Channel in Pixel Domain Wyner-Ziv Video Coding: Myths and Realities. In *Proc. European Signal Processing Conference (EUSIPCO)*, Florence/Italy, September 2006.

[4] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. Distributed Video Coding. *Proc. of the IEEE*, 93(1):71–83, Jan. 2005.

[5] R. Hänsel and E. Müller. Error Locating for plausible Wyner-Ziv Video Decoding using Turbo Codes. In *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Rio de Janeiro, October 2009.

[6] D. Kubasov, K. Lajnef, and C. Guillemot. A Hybrid Encoder/Decoder Rate Control for a Wyner-Ziv Video Codec with a Feedback Channel. In *Proc. IEEE Multimedia Signal Processing Workshop (MMSP)*, Chania, Crete, October 2007.

[7] 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)*, San Diego, October 2008.

[8] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp. Distributed Video Coding: Selecting the Most Promising Application Scenarios. *Signal Processing: Image Communication*, 23:339–352, 2008.

[9] D. Rowitch and L. Milstein. On the performance of hybrid fec/arq systems using rate compatible punctured turbo (rcpt) codes. *IEEE Transactions on Communications*, 48(6):948–959, June 2000.

[10] R. Y. Shao, S. Lin, and M. P. C. Fossorier. Two Simple Stopping Criteria for Turbo Decoding. *IEEE Transactions on Communications*, 47(8):1117–1120, August 1999.

[11] J. Ŝkorupa, J. Slowack, S. Mys, P. Lambert, R. V. de Walle, and C. Grecos. Stopping Criterions for Turbo Coding in a Wyner-Ziv Video Codec. In *Proc. Picture Coding Symposium (PCS)*, Chicago, May 2009.

[12] D. Slepian and J. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Trans. on Information Theory*, 19(4):471–480, July 1973.

[13] Y. Wu, B. D. Woerner, and W. J. Ebel. A Simple Stopping Criterion for Turbo Decoding. *IEEE Communications Letters*, 4(8):258–260, August 2000.

[14] A. D. Wyner and J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Trans. on Information Theory*, 22(1):1–10, Jan 1976.

[15] F. Zhai and I. J. Fair. Techniques for Early Stopping and Error Detection in Turbo Decoding. *IEEE Transactions on Communications*, 51(10):1617–1623, October 2003.