

Performance evaluation of two state of the art DVC codecs

N. Anantrasirichai, D. Agrafiotis
Dept. of Electrical & Electronic Engineering
University of Bristol
Bristol, UK
{N. Anantrasirichai} {D. Agrafiotis} @bristol.ac.uk

M. Ouaret, F. Dufaux, T. Ebrahimi
Multimedia Signal Processing Group
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{mourad.ouaret} {frederic.dufaux} {touradj.ebrahimi} @epfl.ch

Abstract— The performance of existing DVC codecs is still lacking relative to that of H.264 and work is being carried out in order to close this gap. The authors of this paper have been and still are involved in the development of two DVC codecs respectively, the performance of which is compared herein. The aim is to identify strengths and weaknesses of the two codecs that can be exploited / addressed in order to improve the achieved performance relative to H.264.

Keywords-distributed video coding

I. INTRODUCTION

Distributed video coding (DVC) is currently seen as a promising approach to low power, low complexity encoding which, in theory at least [1][2], could offer similar performance to that of typical hybrid video codecs. A common DVC scenario involves splitting of the video frames in two categories, KEY frames and Wyner-Ziv (WZ) frames [3][4], which then undergo conventional (usually intra) and Wyner-Ziv coding respectively. The WZ data, which may undergo transformation, are quantized and fed to a channel coder in a bit-plane by bit-plane fashion. At the decoder, the decoded KEY frames are used for creating an estimate of the WZ frames, the side information (SI). This SI is seen as the systematic part of the channel coder output, i.e. a noisy version of the original WZ frame. The received parity bits are used to correct the errors present in the SI. The number of bits spent for the WZ data largely depends on the quality of the SI and the accuracy of the of the noise estimation.

In this paper, we evaluate and compare the current performance of two DVC codecs that have been developed by the EU funded DISCOVER project (2005-2007) and the UK EPSRC funded MEDIEVaL project (2007-2010) respectively. The comparison will potentially reveal advantages and characteristics of the two schemes, which could then be exploited for the design of a more efficient DVC codec.

The MEDIEVaL project has proposed a novel spatially interleaved DVC, referred to herein as SPI-DVC. We employ hybrid KEY/WZ frames via spatial interleaving of blocks of pixels. This allows efficient SI generation through block-based error concealment and potentially more accurate estimation of the local correlation noise. We have extended our work from an early extrapolation based (P type) only to one that supports bi-directional prediction for creating the side information. We

have added a more sophisticated spatiotemporal concealment (SI prediction) method, and have introduced the use of a Gray code and of a diversity (multi-hypothesis) scheme to produce more reliable results [5]-[7].

The DISCOVER codec is based on Transform Domain Wyner-Ziv (TDWZ) coding, following the architecture initially introduced in [3] and [4]. WZ coding is performed at the frame level and in the transform-domain, on DCT-like coefficients, using turbo codes. The SI is obtained by Motion Compensated Temporal Interpolation (MCTI). A feedback channel is used for decoder-side rate control. The DISCOVER codec is publicly available here [8].

The rest of this paper is organized as follows: Section 2 describes the SPI-DVC codec. Section 3 presents the DISCOVER codec. Results and comparisons are shown in Section 4 followed by conclusions in Section 5.

II. SPI-DVC

The framework of the SPI-DVC codec is illustrated in Figure 1 and described in this section.

A. Spatio-Temporal Interleaving

The first step involves splitting of the current input frame into KEY and WZ groups in a similar fashion to the dispersed type of flexible macroblock ordering specified in H.264. The interleaving block size is fixed for the entire input sequence and can range from 16x16 to 4x4 pixels. If the GOP length is more than 2, the KEY and WZ groups alternate relative to the previous frame so as to avoid creating potentially annoying regions of different subjective quality. The KEY groups of two consecutive frames are combined to avoid any significant performance loss, especially in the case where the complex spatial prediction modes are used, relative to full frame intra KEY coding. The same procedure is applied to the respective WZ groups in order for the frame length of the input to the turbo encoder to be adequate for good performance.

B. A Gray Code

At the WZ encoder, the quantized symbols are converted into binary data. Subsequently, before extracting into bit-planes, the binary codes are converted to Gray codes by XORing binary values with their logical shift-right values. At

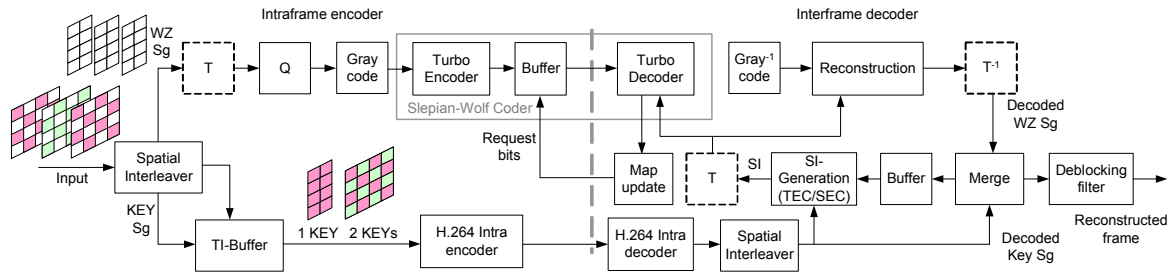


Figure 1. The SPI-DVC codec

the decoder, the decoded Gray codes are converted back to binary values in a similar fashion. The Gray code improves the codec's performance at the cost of a small increase in complexity.

C. Side Information Generation

The generation of the SI is equivalent to an error concealment process for missing blocks (WZ blocks) in the presence of their 4-neighbours (KEY blocks). We employ temporal and spatial error concealment methods (TEC/SEC) controlled by a mode selection algorithm as proposed in [9] for generating the SI. For TEC, each missing WZ block is divided into four sub-blocks. Each sub-block is concealed using two motion vectors coming from neighboring KEY blocks – where motion estimation has previously been performed – and motion vectors generated using an external boundary matching error (EBME) process (Figure 2), fused via a cosine weighted overlapping step. The SEC module uses bordering KEY pixels to conceal the WZ blocks through bilinear or directional interpolation depending on the directional entropy of neighboring edges. The mode selection algorithm switches between TEC and SEC based on the levels of motion compensated activity and spatial activity in the neighborhood of the processed block. If forward reference frames are available, bi-directional motion estimation is employed so that the replacement (SI) block can additionally result from averaging a forward and backward replacement block. The multiple prediction blocks resulting from this process form the basis for the multi-hypothesis SI coding.

D. Multi-hypothesis Decoding

The WZ decoder uses the correlation noise to calculate bit probabilities. We estimate the correlation noise on a block basis using the 4-neighbouring KEY blocks of each WZ block. We model the noise as a Laplacian distribution with a specific variance that changes from block to block. The resulting distribution should follow closely that of the difference between the SI block and the transmitted WZ block as it employs actual received pixels in the vicinity of the processed block, as opposed to frame based interpolated values. As multiple SI data are available, multiple bit probabilities can lead to a more precise log-likelihood ratio (LLR) [10]. Moreover, having multiple SI allows usage of the correct SI to compensate for the errors appearing in others. We define the SI as SI_i , $i \in \{1, \dots, N\}$, with increasing i denoting the ranking of the SI, from best to worst, according to the MSE of the predicted

KEY group associated with each SI group. After the decoded bit-planes are converted back to a quantisation symbol \tilde{Q} , the pixel Y is reconstructed by firstly using the best SI (SI_1) based on its quantisation bins Q_{SI_1} . Next, the pixels of SI_2 with $Q_{SI_2} = \tilde{Q}$ are used to replace the clipped pixels where $Q_{SI_1} \neq \tilde{Q}$. The other SI is sequentially applied to the clipped pixels as follows.

$$Y_p = \begin{cases} SI_p & Q_{SI_p} = \tilde{Q} \text{ and } Q_{SI_{p-1}} \neq \tilde{Q} \\ Y_{p-1} & \text{otherwise} \end{cases} \quad (1)$$

where Y_p is the pixel value after the p^{th} SI has been received, $p \in \{2, \dots, N\}$. Thus the final refined version of the reconstructed WZ data ($Y = Y_N$) can exhibit better image quality.

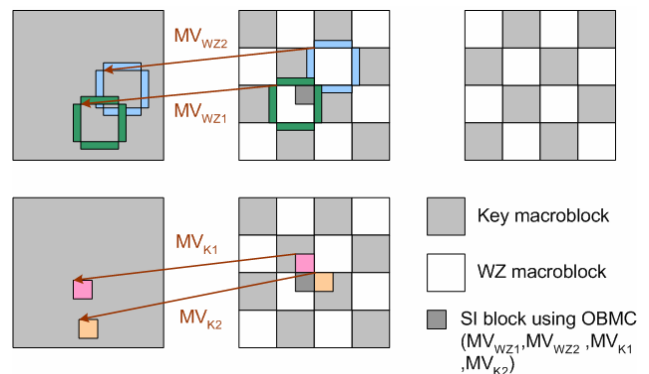


Figure 2. The TEC process

III. DISCOVER DVC

The DISCOVER DVC architecture is illustrated in Figure 3. A detailed description and thorough performance evaluation is given in [11]. The encoder first divides the input video sequence into key frames Y_k and WZ frames X_k . The frequency of the key frames is determined by the Group of Picture (GOP) size. Key frames Y_k are encoded using conventional H.264 Intra coding. Conversely, WZ frames X_k are transformed using the same 4x4 DCT-like transform as in H.264. The quantized transform coefficients are then grouped into bands and split into bitplanes which are turbo encoded. Moreover, for each quantized bitplane a Cyclic Redundancy Check (CRC) is computed and transmitted to the decoder.

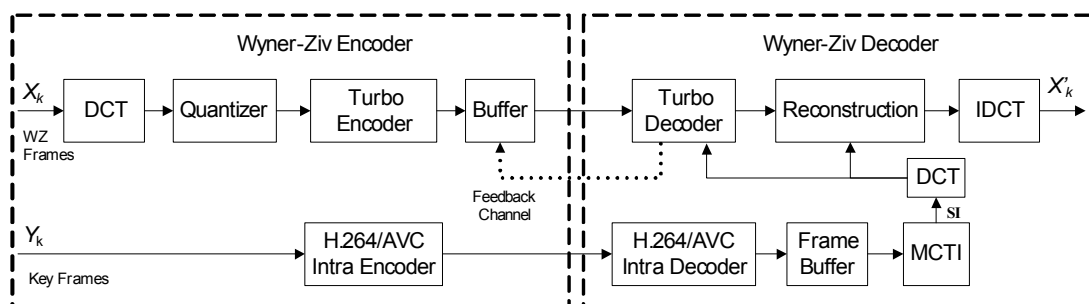


Figure 3: DISCOVER DVC architecture

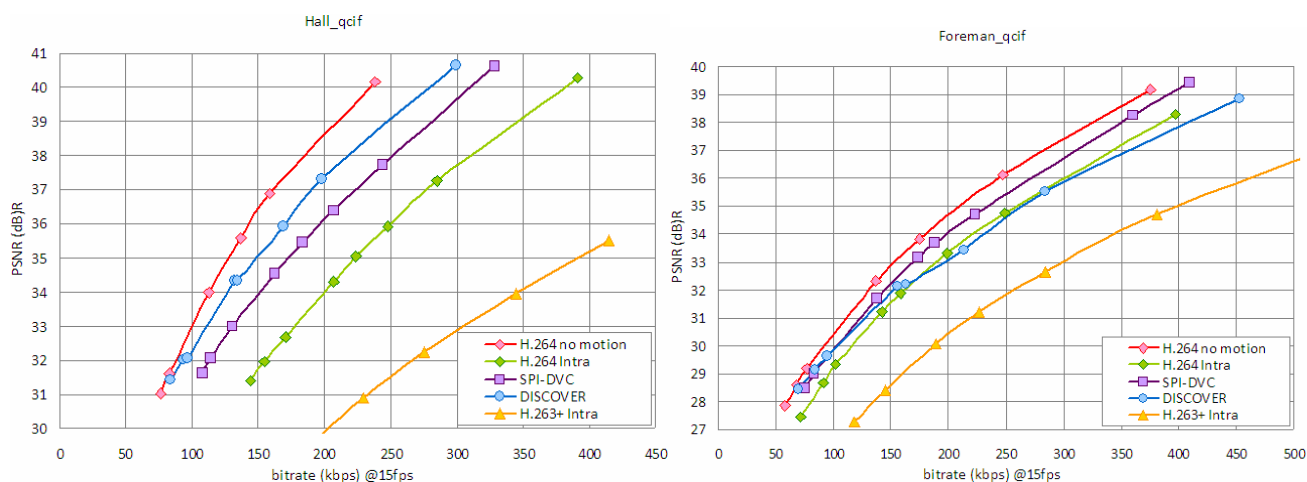


Figure 4. Rate-distortion performance.

At the decoder side, SI is generated by MCTI of the decoded key frames [12]. MCTI includes block-based bi-directional motion estimation, motion interpolation and spatial motion smoothing. A virtual channel is used to model the correlation between the DCT coefficients of the original and SI frames. The SI is exploited by the turbo decoder. The latter requests WZ parity bits by means of a feedback channel until successful decoding, which is reached when the error probability of the decoded bitplane falls below 10^{-3} and its CRC matches the one received from the encoder.

IV. RESULTS AND DISCUSSION

Two QCIF test sequences (Foreman, Hall) were used for evaluating the performance of the two DVC codecs, as well as those of H.264 Intra, H.264 inter with zero motion and H.263+ Intra. The test conditions for the DISCOVER codec can be found here [13].

The results indicate (Figure 4) that SPI-DVC can offer a performance advantage with high motion sequences, such as Foreman. This perhaps is expected given the blockwise generation of the side information and the fact that noise estimation exploits the locally available KEY data. With sequences of low motion (hall) it seems that the frame based approach does better. This disparity in performance can be

traced in a number of parameters including the added overhead of intra coding in SPI-DVC due to spatial interleaving and the reduced frame length of the turbo encoder. The performance of SPI-DVC should be investigated further with regards to the following:

- The coding penalty associated with intra coding a spatiotemporally interleaved version of the sequence, and especially when the interleaving step is small due to the QCIF resolution used
- The noise estimation and how it performs when motion is constant or zero
- The concealment algorithm for the SI generation and how it performs when motion is low or uniform

It is positive to note that the DVC performance in both cases is getting closer to that of H.264 inter with zero motion. The performance of SPI-DVC exceeds that of H.264 Intra in all cases.

V. CONCLUSIONS

This paper has examined two DVC codecs and has evaluated their performance relative to each other and relative to H.264 Intra and Inter with zero motion. The evaluation

reveals the potential advantage of SPI-DVC when motion is high, but also indicates that there is still room for improvement especially with more static sequences, where the DISCOVER codec performs better.

REFERENCES

- [1] D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," IEEE Transactions on Information Theory, vol. 19, no. 4, July 1973.
- [2] A. D. Wyner and J. Ziv, "The Rate Distortion Function for Source Coding with Side Information at the Decoder," IEEE Transactions on Information Theory, vol. 22, no. 1, January 1976.
- [3] A. Aaron, R. Zang, and B. Girod, "Wyner-Ziv Coding of Motion Video," in ASILOMAR Conf. on Signals and Systems, Nov. 2002.
- [4] C. Brites, J. Ascenso, F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance", in Proc ICASSP, May 14-19, 2006.
- [5] D. Agrafiotis, P. Ferré, D. R. Bull, "Hybrid key/Wyner-Ziv frames with flexible macroblock ordering for improved low delay distributed video coding", VCIP 2007, 2007, California, USA.
- [6] N. Anantrasirichai, D. Agrafiotis, D. R. Bull, "A Concealment based Approach to Distributed Video Coding", ICIP2008, San Diego, USA.
- [7] N. Anantrasirichai, D. Agrafiotis and D. R. Bull, "Enhanced Spatially Interleaved DVC Using Diversity and Selective Feedback," ICASSP2009.
- [8] http://www.discoverdvc.org/cont_Codec.html
- [9] D. Agrafiotis, D. R. Bull, N. Canagarajah, "Enhanced error concealment with mode selection", IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 8, pp. 960-973, August 2006.
- [10] K. Misra, S. Karande, H. Radha, "Multi-Hypothesis Distributed Video Coding Using LDPC Codes," Proc. Allerton Conference on Communication, Control, And Computing, 2005.
- [11] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation", Picture Coding Symposium 2007, Lisbon, Portugal.
- [12] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, July 2005.
- [13] http://www.img.lx.it.pt/~discover/test_conditions.html