THE VISNET II DVC CODEC: ARCHITECTURE, TOOLS AND PERFORMANCE

João Ascenso[1]*, Catarina Brites*[1]*, Frédéric Dufaux*[2]*, Anil Fernando*[3]
Touradj Ebrahimi[2]*, Fernando Pereira*[1] *and Stefano Tubaro*[4]

[1] Instituto Superior Técnico - Instituto de Telecomunicações, 1049-001 Lisboa, Portugal
email: joao.ascenso@lx.it.pt, catarina.brites@lx.it.pt, fp@lx.it.pt

[2] Ecole Polytechnique Fédérale de Lausanne (EPFL) - Multimedia Signal Processing Group, 1015 Lausanne, Switzerland
email: frederic.dufaux@epfl.ch, touradj.ebrahimi@epfl.ch

[3] University of Surrey – I-Lab Multimedia Group, Guildford, UK GU2 7XH
email: W.Fernando@surrey.ac.uk

[4] Politecnico di Milano - Dipartimento di Elettronica e Informazione, 20133 Milano, Italy
email: stefano.tubaro@polimi.it

ABSTRACT

This paper introduces the VISNET II DVC codec. This codec achieves very high RD performance thanks to the efficient combination of many state-of-the-art coding tools into a fully practical video codec. Experimental results show that the proposed DVC codec consistently outperforms H.264/AVC Intra. For sequences with coherent motion, it even surpasses H.264/AVC zero-motion. Finally, it is also always better than the DISCOVER DVC codec. Therefore, it is expected that the proposed high performing DVC codec will be used by other researchers in the field as a reference to benchmark their results.

1. INTRODUCTION

With the wide deployment of mobile and wireless networks, a growing number of emerging applications, such as low-power sensor networks, video surveillance cameras and mobile communications, rely on an up-link model rather than the typical down-link communication model. Typically, these applications are characterized by many senders transmitting data to a central receiver. In this context, light encoding or a flexible distribution of the codec complexity, robustness to packet losses, high compression efficiency and low latency/delay are important requirements.

To address the needs of these up-link applications, the usual predictive video coding paradigm has been revisited based on Information Theory theorems from the 70s. The Slepian-Wolf (SW) theorem [1] establishes lower bounds on the achievable rates for the lossless coding of two or more correlated sources. More specifically, considering two statistically dependent random signals X and Y, it is well-known that the lower bound for the rate is given by the joint entropy $H(X,Y)$ when these two signals are jointly encoded (as in conventional predictive coding). Conversely, when these two signals are independently encoded but jointly decoded (distributed coding), the SW theorem states that the minimum rate is still $H(X,Y)$ with a residual error probability which tends towards 0 for long sequences. Later, Wyner and Ziv (WZ) have extended the SW theorem and showed that the result holds for the lossy coding case under the assumptions that the sources are jointly Gaussian and a mean square error distortion measure is used [2]. Subsequently, it was shown that this result remains valid as long as the difference between X and Y is Gaussian.

Video coding schemes based on these theorems are referred to as Distributed Video Coding (DVC) solutions. Since the new coding paradigm is based on a statistical framework and does not rely on joint encoding, DVC architectures may provide several functional benefits which are rather important for many emerging applications: i) flexible allocation of the global video codec complexity; ii) improved error resilience; iii) codec independent scalability; and iv) exploitation of multiview correlation.

Based on these theoretical results, practical implementations of DVC have been proposed since 2002. The PRISM (Power-efficient, Robust, hIgh compression Syndrome-based Multimedia coding) [3] solution works at the block level and performs motion estimation at the decoder. Based on the amount of temporal correlation, estimated using a zero-motion block difference, each block can either be conventionally (intra) coded, skipped or coded using distributed coding principles. Another DVC architecture working at frame level has been proposed in [4]; this DVC solution includes a feedback channel which allows performing decoder rate control based on the available correlation.

In this paper, the DVC codec developed within the European Network of Excellence VISNET II project [5] is described. This codec is based on the early architecture in [4] and integrates numerous advanced tools [6]-[14], either developed within the VISNET II network or proposed in the literature. It is the outcome of an intensive collaboration, resulting in a complete DVC system with state-of-the-art Rate-Distortion (RD) performance.

2. VISNET II CODEC ARCHITECTURE AND TOOLS

This section provides a description of the VISNET II DVC codec architecture and tools illustrated in Figure 1.

© EURASIP, 2010 ISSN 2076-1465

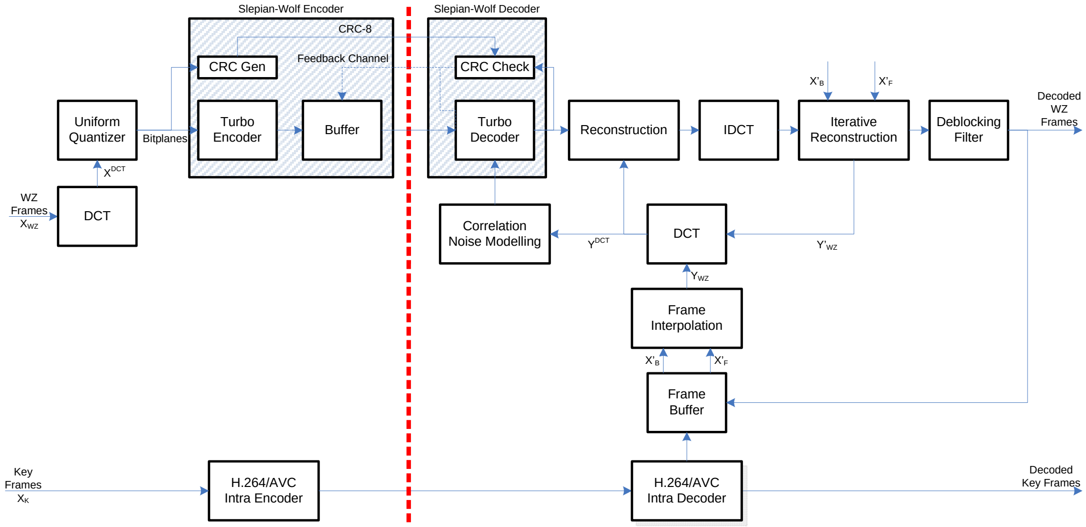

Figure 1 – Proposed VISNET II DVC architecture

The main objective of this codec is to reach the best possible RD performance. For this purpose, advanced tools, either developed within the VISNET II project or proposed in the literature by other research groups, have been adopted for most of the modules described hereafter. The main contribution of this paper consists in the effective combination of these tools in order to develop an efficient DVC codec. In addition, the VISNET II DVC codec represents a significant advance over the state-of-the-art, since it integrates novel tools such as iterative reconstruction and deblocking filter.

Finally, it is important to stress that the VISNET II DVC codec is a fully practical video codec. For instance, no original frames are used at the decoder to create the side information, to estimate the bitplane error probability or to estimate the correlation noise model parameters.

2.1 Encoder

First, the video sequence is divided into WZ frames and key frames. Key frames are typically periodically inserted with a certain GOP size and are coded using a H.264/AVC Intra codec. An adaptive GOP size selection process may also be used; in this case, the key frames are inserted depending on the amount of temporal correlation in the video sequence [6]. Most results available in the literature use a GOP size of 2 which means that odd and even frames are key frames and WZ frames, respectively.

For the WZ frames, the following operations are carried out at the encoder side:

- **Transform** - For WZ frames, an integer 4×4 block-based DCT is applied. The DCT coefficients of the entire WZ frame are then grouped together, according to their position within the 4×4 blocks, forming DCT coefficients bands.
- **Quantization** - Next, each DCT coefficients band b_k is uniformly quantized with 2^{M_k} levels (where the number of levels 2^{M_k} depends on the band b_k). The resulting quantized symbols are then split into bitplanes, i.e. for a given band, the bits of the same significance are grouped together in a bitplane array.
- **Turbo Encoding** - These bitplanes are then independently encoded. More precisely, the turbo encoding procedure for the band b_k starts with the most significant bitplane array. A pre-interleaver is also applied [7] for improved RD performance. The parity information generated by the turbo encoder for each bitplane is then stored in the buffer and sent in chunks/packets upon decoder request, through the feedback channel.
- **CRC** - The encoder also calculates an 8 bit Cyclic Redundancy Check (CRC) hash for each bitplane and sends it to the decoder [8]. This will help the decoder to detect any remaining residual errors left by the stopping criterion computed at the turbo decoder.

Conversely, the key frames are encoded with the efficient H.264/AVC Intra mode in main profile with Context-adaptive Binary Arithmetic Coding (CABAC) active and all the spatial Intra prediction modes (4×4 and 8×8) enabled, thus achieving a good coding performance.

2.2 Decoder

For the WZ frames, the following operations are carried out at the decoder side:

- **Frame Interpolation** - The decoder creates the side information for each WZ coded frame with a motion compensated frame interpolation framework, using the previous and next temporally closer reference frames to generate an estimate of the WZ frame. The side information for each WZ frame corresponds to an estimation of the original WZ frame. With a better quality of this estimation, the turbo decoder has to correct fewer ‘errors’ and the bitrate necessary for successful decoding (i.e. to achieve a small error probability) decreases. In the proposed codec, hierarchical motion estimation [6] and spatial motion smoothing [9] are applied in order to improve RD performance.
- **Side Information Transformed** - A block-based 4×4 DCT is then carried out over the side information in order to obtain the DCT coefficients which are an estimate of the WZ frame DCT coefficients.

- **Correlation Noise Modeling** - The residual statistics between corresponding WZ frame DCT coefficients and the side information DCT coefficients is modeled by a Laplacian distribution. The Laplacian parameter is estimated online and at the coefficient granularity level [10].
- **Turbo Decoding** - Once the DCT transformed side information and the residual statistics for a given DCT coefficients band b_k are known, the decoded quantized symbol stream associated to the DCT band b_k can be obtained through the turbo decoding procedure. The turbo decoder receives from the encoder successive chunks of parity bits following the requests made through the feedback channel. After successfully turbo decoding the most significant bitplane array of the b_k band, the turbo decoder proceeds in an analogous way with the remaining M_{k-1} bitplanes associated to the same band. Once all the bitplanes of the DCT coefficients band b_k are successfully turbo decoded, the turbo decoder starts decoding the next b_{k+1} band. This procedure is repeated until all the DCT coefficients bands, for which WZ bits are transmitted, are turbo decoded.
- **Request Stopping Criterion** - To decide whether more parity bits are needed for the successful decoding of a certain bitplane, the decoder uses a simple request stopping criterion. The stopping criterion estimates the current bitplane error probability P_e for a given DCT band based on the a posteriori probabilities ratio [11]. If P_e is higher than 10^{-3} the decoder requests for more parity bits from the encoder via feedback channel; otherwise, the bitplane turbo decoding task is considered successful.
- **CRC Checking** - Because some residual errors remain when the request stopping criterion is fulfilled and these errors may have a rather negative subjective impact on the decoded frame quality, a CRC checksum is transmitted to help the decoder detect and correct the remaining errors in each bitplane. Since this CRC is combined with the developed request stopping criterion, it does not have to be very strong in order to guarantee a vanishing error probability (≈ 0) for each decoded bitplane. As a consequence, a CRC-8 checksum for each bitplane was found to be strong enough for this purpose which only adds minimal extra rate (8 bits) for each decoded bitplane (in other words, 1536 source bits for QCIF sequences). Thus, if the decoded bitplane has the same CRC checksum that the encoder CRC checksum, the decoding is declared to be successful and the decoding of another band/bitplane can start; otherwise, more parity bits are requested and the turbo decoding process starts again.
- **Bin Forming** - After turbo decoding the M_k bitplanes associated to the DCT band b_k, the bitplanes are grouped together to form the decoded quantized symbol stream associated to the b_k band. This procedure is performed over all the DCT coefficients bands for which WZ bits are transmitted. The DCT coefficients bands for which no WZ bits were transmitted are replaced by the corresponding DCT bands from the DCT side information.
- **Reconstruction** - The reconstruction corresponds to the inverse of the quantization but exploits the side information DCT coefficients and all turbo decoded symbol streams (quantization bins/intervals) [12]. The reconstruction is performed iteratively for each refinement of the side information.
- **Inverse Transform** - After, a block-based 4×4 IDCT is performed and the reconstructed pixel domain WZ frame is obtained.
- **Iterative Reconstruction** - Next, iterative reconstruction is performed. It relies on the partially decoded WZ frame which becomes available in the WZ decoder, i.e. the decoded result after the correction of some errors in the side information. This partially decoded frame has higher quality than the side information and thus it can be exploited to generate side information again but with improved quality [13]. This procedure is based on the refinement of the motion vectors and the reference frame selection (backward, forward and bidirectional predictions are allowed) and exploits the obtained partially decoded frame. With the improved side information the reconstruction can be performed again obtaining a higher quality frame.
- **Deblocking Filter** - To improve both subjective and objective qualities of the WZ decoded frames, a decoder side adaptive deblocking filter is proposed [14]. More specifically, the filter is inserted inside the SI motion estimation loop as an In-Loop Deblocking Filter (ILDF), i.e. the frame generated by the filter is used as reference in the side information generation process (for GOP sizes greater than 2). This technique is based on the well known H.264/AVC deblocking filter where the calculations of the boundary strength and filter parameters were adapted to the DVC context.

3. PERFORMANCE ASSESSMENT

In order to assess the RD performance of the proposed VISNET II DVC codec, simulation results are reported in this section. Experiments are carried out with four test sequences Hall Monitor, Coastguard, Foreman and Soccer, with QCIF resolution and 15 fps, adopting the same test conditions as described in detail in [15].

Figure 2 shows the RD performance in comparison with two H.264/AVC variants with low encoding complexity: H.264/AVC Intra and H.264/AVC zero-motion. The former is one of the most efficient Intra coding solutions available: While no temporal correlation is exploited, it is important to note that H.264/AVC Intra exploits quite efficiently the spatial correlation with several Intra prediction modes. Conversely, H.264/AVC zero-motion exploits the temporal redundancy in a IB…BI structure, but without performing motion estimation (i.e. all motion vectors are zero). In this way, better performance than Intra coding is achieved, as temporal redundancy is partly exploited. However, it requires far less complexity than full motion compensated Inter coding since no (encoder) motion search is performed.

From these results, it is possible to observe that the VISNET II DVC RD performance is consistently better than the H.264/AVC Intra RD performance with the exception of content with highly complex motion such as the Soccer sequence. For simple content, such as the Hall Monitor video surveillance sequence, DVC gains over H.264/AVC Intra can go up to 5 dB for GOP size 8.

The VISNET II DVC RD performance is typically worse than the H.264/AVC zero-motion RD performance. However, for sequences with regular global motion, like the Coastguard video surveillance sequence, the VISNET II DVC codec performs better than H.264/AVC zero-motion due to the inability of the latter to efficiently exploit the temporal redundancy.

The best VISNET II DVC RD performance is typically reached for GOP size 2, showing the difficulty to generate effective side information at the decoder when the key frames are farther apart. However, for simple sequences such as Hall Monitor, the RD performance increases with the GOP size in the range of values tested.

Figure 3 shows the RD performance in comparison with the DISCOVER DVC codec [16] which is one the best performing DVC codecs available in the literature. It can be observed that the VISNET II DVC RD performance is consistently better than the DISCOVER DVC RD performance for all sequences and bitrates. The gains are more substantial for high motion sequences and for longer GOP sizes. These gains are mainly associated to the improvements in the side information creation process, the iterative reconstruction process and the deblocking filter.

In terms of complexity, a thorough analysis of a DVC codec which shares a similar architecture as the proposed VISNET II DVC codec is presented in [17]. It is shown that WZ frames encoding complexity is about 1/6 of the average H.264/AVC Intra or H.264/AVC zero-motion encoding complexity. Conversely, the DVC decoding complexity is always much higher than H.264/AVC Intra or H.264/AVC zero-motion decoding complexity.

4. CONCLUSIONS

In this paper, the VISNET II DVC codec, which is the result of an intensive collaborative work, is presented. This codec integrates multiple advanced coding tools to build a powerful DVC system with state-of-the-art RD performance. The RD performance assessment has shown that the proposed VISNET II DVC codec consistently achieves better RD performance when compared to the H.264/AVC Intra codec. Moreover, for sequences with regular global motion, the VISNET II DVC codec even performs better than H.264/AVC zero-motion. Finally, it is also consistently better than the DISCOVER DVC codec which is one the best performing DVC codecs in the literature.

5. ACKNOWLEDGEMENTS

This work was partially supported by the European Network of Excellence VISNET II (http://www.visnet-noe.org) funded under the European Commission IST 6th Framework Program (IST Contract 1-038398). The authors would like to acknowledge the contributions of M. Badem, R. Martins, M. Ouaret, J. Pedro, R. Weerakkody and S. Ye.

REFERENCES

[1] J. 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. 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, January 1976.

[3] R. Puri, A. Majumdar, and K. Ramchandran, “PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder”, *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436-2448, October 2007.

[4] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, “Distributed Video Coding”, *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71-83, January 2005.

[5] http://www.visnet-noe.org.

[6] J. Ascenso, C. Brites, F. Pereira, “Content Adaptive Wyner-Ziv Video Coding driven by Motion Activity”, in *Proc. IEEE International Conference on Image Processing*, Atlanta, USA, October 2007.

[7] M. Dalai, R. Leonardi, F. Pereira, “Improving Turbo Codec Integration in Pixel-Domain Distributed Video Coding”, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[8] D. Kubasov, K. Lajnef, C. Guillemot, “A Hybrid Encoder/Decoder Rate Control for Wyner-Ziv Video Coding with a Feedback Channel”, in *Proc. IEEE International Workshop on Multimedia Signal Processing*, Chania, Greece, October, 2007.

[9] J. Ascenso, C. Brites, F. Pereira, “Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding”, in *Proc. 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.

[10] C. Brites, F. Pereira, “Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner-Ziv Video Coding”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp.1177-1190, September 2008.

[11] M. Tagliasacchi, J. Pedro, F. Pereira, S. Tubaro, “An Efficient Request Stopping Method at the Turbo Decoder in Distributed Video Coding”, in *Proc. European Conference on Signal Processing*, Poznan, Poland, September 2007.

[12] D. Kubasov, J. Nayak, C. Guillemot, “Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information”, in *Proc. IEEE International Workshop on Multimedia Signal Processing*, Chania, Greece, October, 2007.

[13] S. Ye, M. Ouaret, F. Dufaux and T. Ebrahimi, “Improved Side Information Generation for Distributed Video Coding by Exploiting Spatial and Temporal Correlations”, *EURASIP Journal on Image and Video Processing*, vol. 2009, 15 pages, 2009.

[14] R. Martins, C. Brites, J. Ascenso, F. Pereira, “Adaptive Deblocking Filter for Transform Domain Wyner-Ziv Video Coding”, *IET Signal Processing*, vol. 3, no. 6, pp. 315-328, December 2009.

[15] http://www.img.lx.it.pt/~discover/home.html.

[16] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, “The DISCOVER Codec: Architecture, Techniques and Evaluation”, in *Proc. of Picture Coding Symposium*, Lisboa, Portugal, November 2007.

[17] C. Brites, J. Ascenso, J. Pedro, F. Pereira, “Evaluating a Feedback Channel based Transform Domain Wyner-Ziv Video Codec”, *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 269-297, April 2008.

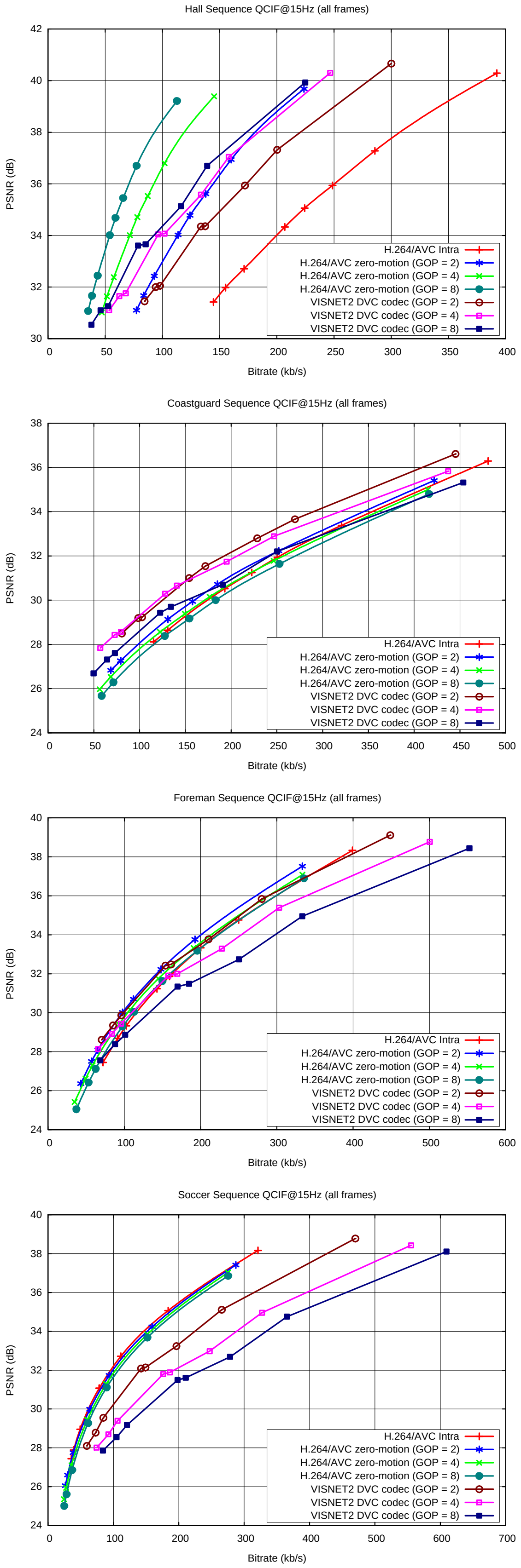

Figure 2 – RD performance comparison with H.264/AVC Intra and H.264/AVC zero-motion

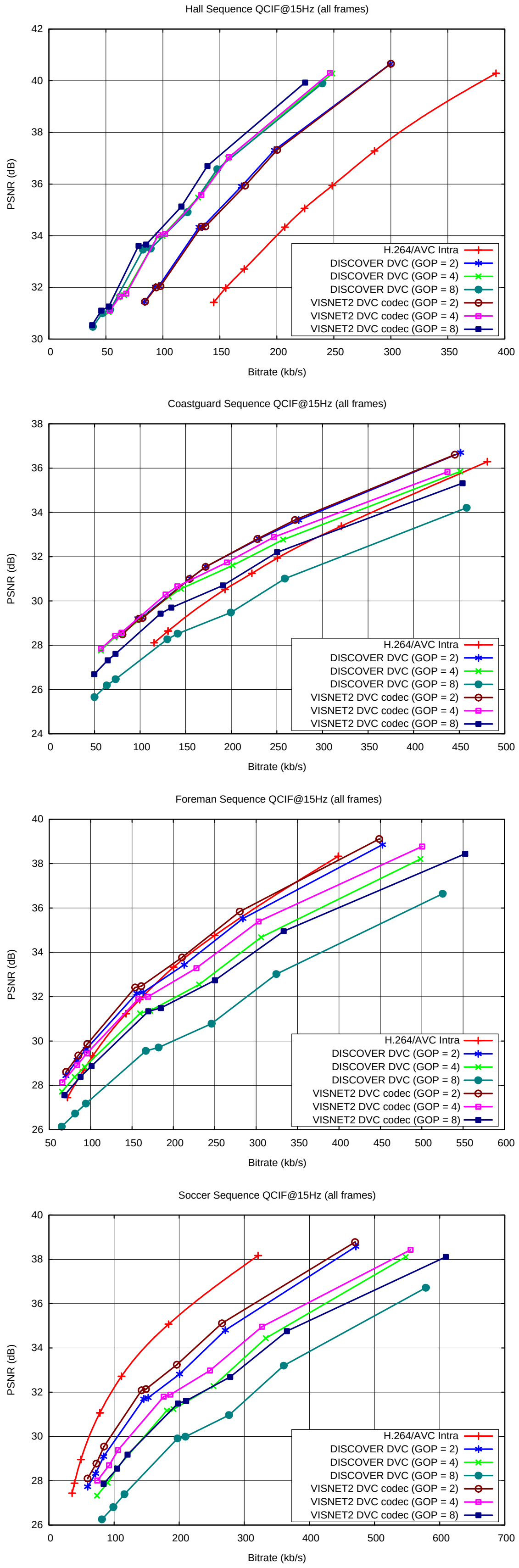

Figure 3 – RD performance comparison with the DISCOVER DVC codec