Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information

Denis Kubasov, Jayanth Nayak, Christine Guillemot
IRISA / INRIA Rennes, Campus Universitaire de Beaulieu,
Avenue du Général Leclerc, 35042 RENNES Cedex - France
denis.kubasov@irisa.fr, jnayak@ee.ucr.edu, christine.guillemot@irisa.fr

***Abstract*—This paper addresses the problem of optimal minimum mean-squared error reconstruction of quantised samples in Wyner-Ziv video coding systems. Closed-form expressions of the optimal reconstructed values are derived for a Laplacian correlation model. The method is used for both single and multiple side information scenarios (the latter is also referred to as multi-hypothesis Wyner-Ziv decoding). The efficiency of the proposed optimal reconstruction method is confirmed by rate-distortion performance results, showing significant decrease of the distortion of the decoded sequence, compared to simple reconstruction methods that have been employed so far.**

I. Introduction

Distributed source coding (DSC) has emerged as an enabling technology for sensor networks. It refers to the compression of correlated signals X and Y captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf (SW) [1] and Wyner-Ziv (WZ) [2] theorems. In practice, a Wyner-Ziv (WZ) codec is obtained by adding a quantiser before the Slepian-Wolf encoder. The most widely used practical Slepian-Wolf encoders are based on the channel coding principles [3], [4]. The statistical dependence between X and Y is modelled as a virtual correlation channel, and Y is considered as a noisy version of X. The encoder computes the parity bits for X (a systematic channel code is used) and sends them to the decoder (the systematic bits are discarded). The decoder removes the 'virtual' channel noise, and thus estimates X given the received parity bits and the side information Y regarded as a noisy version of the codeword systematic bits.

Video compression has been recast into a distributed source coding framework leading to distributed video coding (DVC) systems targeting mainly low coding complexity and error resilience functionalities. Correlated samples (pixels or transform coefficients) from different frames are regarded as outputs of different sensors. A comprehensive survey of first DVC solutions can be found in [6]. The DVC architecture considered here is based on the solution proposed in [5] and further developed in [6]. The rate of the SW coder (here a turbo coder) is controlled via a feedback channel: when the residual error rate at the output of the turbo decoder exceeds a given threshold, more bits are requested from the encoder via the feedback channel. Once the index of the quantisation interval is decoded, the reconstruction is performed, taking into account the side information. In most implementations so far a simple reconstruction is employed, taking the corresponding side information sample if it falls into the quantisation interval, or one of the borders of the interval, otherwise.

In this paper, we present an optimal reconstruction approach, which exploits the actual correlation model between the source and the side information, to minimise the mean-squared error (MSE) of the reconstructed sample. Given the Laplacian correlation model, a closed-form expression of the reconstructed value is derived. This allows gains up to 1dB in rate-distortion performance at almost no cost of the additional complexity at the decoder. We also extend this approach to the decoding scenario with multiple side information. A scenario with multiple side information has been considered earlier in [7], where the two side information hypotheses are obtained as a motion compensated extrapolation of the previous and the next frames respectively. Here, one side information hypothesis is obtained by the block-based motion compensated temporal interpolation (MCTI) from [8], whereas the other is obtained by the mesh-based MCTI from [9].

The paper is organised as follows. Section II describes the architecture of the Wyner-Ziv video coding system being used. The proposed optimal reconstruction approach is presented in Section III, and extended to the multiple side information scenario in Section IV. Experimental results are presented and discussed in Section V. Finally, Section VI concludes the paper.

II. Wyner-Ziv coding framework

The codec architecture considered in this paper is based on [6] and is depicted in Fig. 1. The input sequence is structured into groups of frames (GOF), in which key frames are Intra-coded (using H.264/AVC Intra) and intermediate frames are WZ coded. The WZ data are transformed (using a 4x4 DCT) and organised in 16 subbands. For a given rate point, one has a pair of parameters: QP for the key-frames and a quantisation matrix for the WZ data. Each DCT band is

The work presented was developed within DISCOVER, a European Project (http://www.discoverdvc.org), funded under the European Commission IST FP6 program.

J. Nayak is now with University of California, Riverside, California 92521, USA

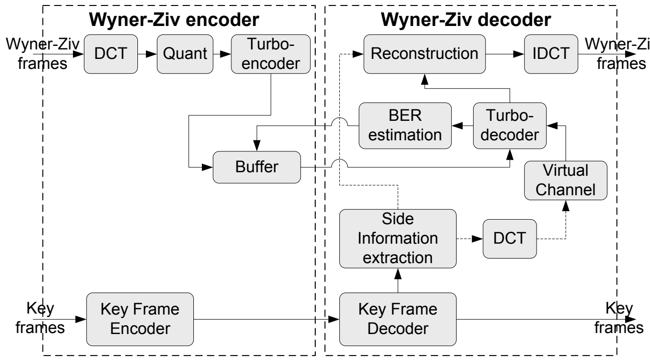


Figure 1. The architecture of the Wyner-Ziv codec

uniformly quantised over the dynamic range of coefficients in the band, using the number of levels specified in the quantisation matrix. Quantisation indices of the coefficients are then suitably binarised, and each bit-plane is fed to the turbo encoder. The systematic bits are discarded and only the parity bits of the turbo coder are stored in a buffer. Initially the encoder sends only a subset of the parity bits.

The decoder constructs the side information using MCTI of the key frames as described in [8]. A Laplacian model is adopted for the virtual correlation channel, its variance being estimated from the residue obtained by motion compensating the key frames [10]. For each bit-plane per band, upon reception of the initial subset of parity bits, the turbo decoder runs a log-MAP decoding algorithm. If the estimated bit error rate on the output of the log-MAP decoding algorithm exceeds a predefined threshold, the decoder requests more parity bits. After turbo decoding, assuming perfect decoding of the indices, a reconstruction of transform coefficients is performed, given the received quantisation index and the side information. Fig. 2 shows an abstraction of a Wyner-Ziv codec, which includes a Slepian-Wolf codec and a pair of quantisation and reconstruction modules.

III. Minimum MSE reconstruction

Here we shall concentrate on the coefficient reconstruction block. We begin with some notation: M denotes the number of quantiser levels and $z_0 < z_1 < ... < z_M$ denote the quantiser levels themselves. Since the quantiser is uniform, $z_{i+1} - z_i = \Delta, \forall i = 0, .., M-1$, where Δ is the quantisation step size. A straightforward approach [11] to reconstruct the source coefficient x using a side information value y is as follows:

$$\hat{x} = \begin{cases} z_i, & y < z_i \\ y, & y \in [z_i, z_{i+1}) \\ z_{i+1}, & y \geq z_{i+1} \end{cases}, \quad (1)$$

where $\hat{x}$ denotes the reconstructed value, and i denotes the quantisation index of x. Another reconstruction approach was proposed recently in [12], which makes use of the average statistical distribution of transform coefficients.

However, both these approaches are suboptimal. One can instead compute $\hat{x}$ as the expectation $E[x|x \in [z_i, z_{i+1}), y]$ of the random variable x given the quantisation interval $[z_i, z_{i+1})$

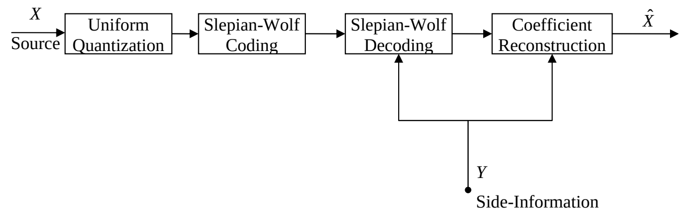


Figure 2. Abstraction of the Wyner-Ziv codec used in DVC

and the side information value y [5]:

$$\hat{x}_{\mathrm{opt}} = E[x|x \in [z_i, z_{i+1}), y] = \frac{\int_{z_i}^{z_{i+1}} x f_{X|y}(x) dx}{\int_{z_i}^{z_{i+1}} f_{X|y}(x) dx}, \quad (2)$$

where $f_{X|y}(x)$ is the conditional p.d.f. of X given y.

This reconstructed value $\hat{x}_{\mathrm{opt}}$ will correspond to the minimum mean-squared error estimate of the source x. Given the Laplacian model of the residue between the source DCT band X and the side information DCT band Y, we derive a closed-form expression for this optimal reconstruction value $\hat{x}_{\mathrm{opt}}$ to avoid the numerical computation of integrals in (2). The Laplacian model implies that:

$$f_{X|y}(x) = \frac{\alpha}{2} e^{-\alpha|x-y|}, \quad (3)$$

where α is the model parameter related to the variance σ^2 of the Laplacian distribution as $\sigma^2 = \frac{2}{\alpha^2}$.

Now, given the expression (3) for $f_{X|y}(x)$, the numerator $N = \frac{\alpha}{2} \int_{z_i}^{z_{i+1}} x e^{-\alpha|x-y|} dx$ and the denominator $D = \frac{\alpha}{2} \int_{z_i}^{z_{i+1}} e^{-\alpha|x-y|} dx$ of (2) are found analytically for $y \notin [z_i, z_{i+1})$:

$$\begin{aligned} N &= \frac{(1 \pm \alpha z_i) e^{\pm\alpha(y-z_i)}}{2\alpha} \\ &\quad - \frac{(1 \pm \alpha z_{i+1}) e^{\pm\alpha(y-z_{i+1})}}{2\alpha}, \end{aligned} \quad (4)$$

$$D = \pm \frac{e^{\pm\alpha(y-z_i)} - e^{\pm\alpha(y-z_{i+1})}}{2}, \quad (5)$$

where $+$ and $-$ correspond to cases $y < z_i$ and $y \geq z_{i+1}$ respectively. For $y \in [z_i, z_{i+1})$ it can be easily seen that:

$$N = \int_{z_i}^{z_{i+1}} x e^{-\alpha|x-y|} dx = \int_{z_i}^{y} x e^{-\alpha(y-x)} dx + \int_{y}^{z_{i+1}} x e^{-\alpha(x-y)} dx. \quad (6)$$

where each of two integrals at the right hand is given by (4) with either $+$ or $-$. The denominator D can be represented in a similar way. Substituting (4) and (5) into (2) and using the decomposition rule (6) we finally get:

$$\hat{x}_{\mathrm{opt}} = \begin{cases} z_i + \frac{1}{\alpha} + \frac{\Delta}{1-e^{\alpha\Delta}}, & y < z_i \\ y + \frac{(\gamma+\frac{1}{\alpha})e^{-\alpha\gamma} - (\delta+\frac{1}{\alpha})e^{-\alpha\delta}}{2-(e^{-\alpha\gamma}+e^{-\alpha\delta})}, & y \in [z_i, z_{i+1}) \\ z_{i+1} - \frac{1}{\alpha} - \frac{\Delta}{1-e^{\alpha\Delta}}, & y \geq z_{i+1} \end{cases}, \quad (7)$$

where $\gamma = y - z_i$ and $\delta = z_{i+1} - y$.

Compared to (1), (7) shifts the reconstruction levels towards the center of the quantisation interval. As expected, when $\alpha = 0$, that is when y conveys no information about x, $\hat{x}_{\mathrm{opt}} = \frac{z_i + z_{i+1}}{2}$. On the other hand, when $\alpha \to \infty$, $\hat{x}_{\mathrm{opt}}$ approaches the $\hat{x}$ in (1).

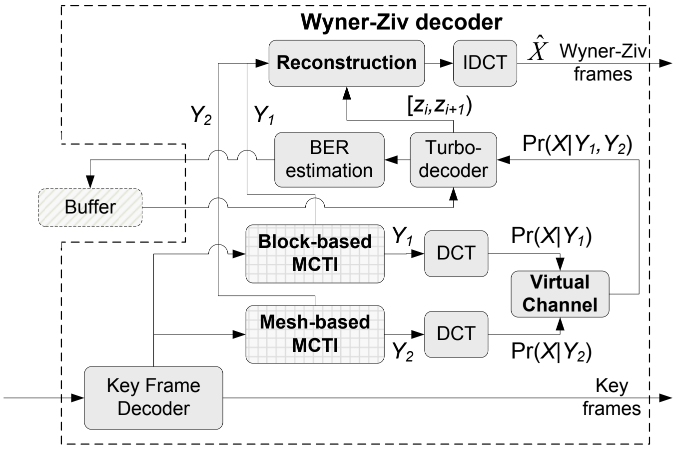


Figure 3. Decoding with multiple side information

IV. Scenario with multiple side information

In this section we describe a decoding scenario with multiple side information which is inspired by [7]. In our case we have two side information hypotheses, which are obtained with different motion-compensated temporal interpolation (MCTI) methods, namely the block-based MCTI from [8] and the mesh-based MCTI from [9].

The decoder using multiple side information is shown in Fig. 3. The two side information hypotheses Y_1 and Y_2 are both considered correlated with the source X (the parameter α of the Laplacian correlation model may be different), giving the two conditional probabilities $\Pr(X|Y_1)$ and $\Pr(X|Y_2)$. As no a priori advantage is given to either hypothesis, we assign $\Pr(X|Y_1, Y_2) = \frac{1}{2}\left(\Pr(X|Y_1) + \Pr(X|Y_2)\right)$ to initialise the virtual correlation channel.

Another module requiring adaptation for multiple side information, is the reconstruction module. Optimal minimum MSE reconstruction for this scenario is also given by (2) with $f_{X|Y_1,Y_2}(x)$ instead of $f_{X|Y}(x)$. We consider here $f_{X|Y_1,Y_2}(x) = \frac{1}{2}\left(f_{X|Y_1}(x) + f_{X|Y_2}(x)\right)$, which coincides with our calculation of the conditional probability $\Pr(X|Y_1, Y_2)$:

$$\hat{x}_{\mathrm{opt,MH}} = E[x|x \in [z_i, z_{i+1}), y_1, y_2] = \tag{8}$$
$$= \frac{\sum\limits_{k=1}^{2} \int\limits_{z_i}^{z_{i+1}} x f_{X|y_k}(x) dx}{\sum\limits_{k=1}^{2} \int\limits_{z_i}^{z_{i+1}} f_{X|y_k}(x) dx} = \frac{\sum\limits_{k=1}^{2} \int\limits_{z_i}^{z_{i+1}} \frac{\alpha_k}{2} x e^{-\alpha_k |x - y_k|} dx}{\sum\limits_{k=1}^{2} \int\limits_{z_i}^{z_{i+1}} \frac{\alpha_k}{2} e^{-\alpha_k |x - y_k|} dx}.$$

To simplify the calculation, the interval $[z_i, z_{i+1})$ can be divided into s non-overlapping subintervals by points $z_i = q_0 < q_1 < ... < q_s = z_{i+1}$ so that each difference $(x - y_k), \forall k$, keeps its sign constant over each subinterval $[q_j, q_{j+1}), \forall j$. Thus we obtain the following formula (9) for $\hat{x}_{\mathrm{opt,MH}}$:

$$\hat{x}_{\mathrm{opt,MH}} = \frac{\sum_{k=1}^{2} \sum_{j=0}^{s-1} \left[\frac{\alpha_k}{2} \int_{q_j}^{q_{j+1}} x e^{-\alpha_k |x - y_k|} dx \right]}{\sum_{k=1}^{2} \sum_{j=0}^{s-1} \left[\frac{\alpha_k}{2} \int_{q_j}^{q_{j+1}} e^{-\alpha_k |x - y_k|} dx \right]}, \tag{9}$$

where both expressions in square brackets are directly given by (4) and (5).

Table I
Average decoding time per frame with single side information

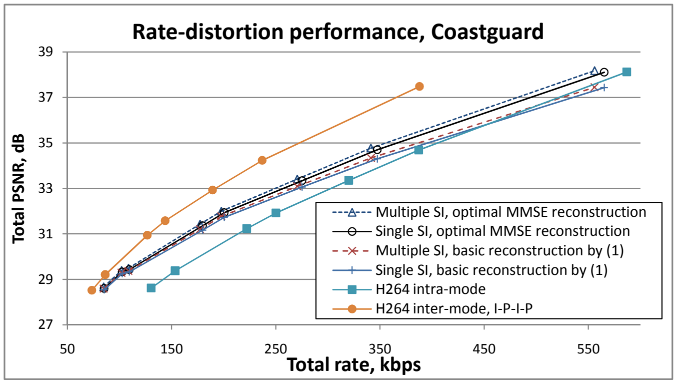
Sequence	RD #	Simple Recons.	MMSE Recons.	Δt, sec
Coastguard	1	9.94s	9.95s	0.01
	2	11.44s	11.42s	-0.02
	3	13.48s	13.50s	0.02
	4	23.05s	23.07s	0.02
	5	24.98s	25.06s	0.08
	6	35.26s	35.31s	0.05
	7	44.75s	44.76s	0.01
	8	73.98s	74.00s	0.02
Foreman	1	13.69s	13.49s	-0.19
	2	15.64s	15.65s	0.01
	3	18.32s	18.36s	0.04
	4	30.32s	30.32s	0.00
	5	34.02s	34.06s	0.03
	6	45.95s	46.02s	0.07
	7	56.16s	56.18s	0.02
	8	86.45s	86.22s	-0.24
Soccer	1	18.74s	18.73s	0.00
	2	21.13s	21.14s	0.01
	3	25.11s	24.73s	-0.38
	4	40.89s	40.66s	-0.23
	5	43.86s	43.88s	0.02
	6	57.02s	56.96s	-0.07
	7	70.74s	70.78s	0.04
	8	107.61s	107.51s	-0.09
			Median:	**0.01**

V. Experimental results

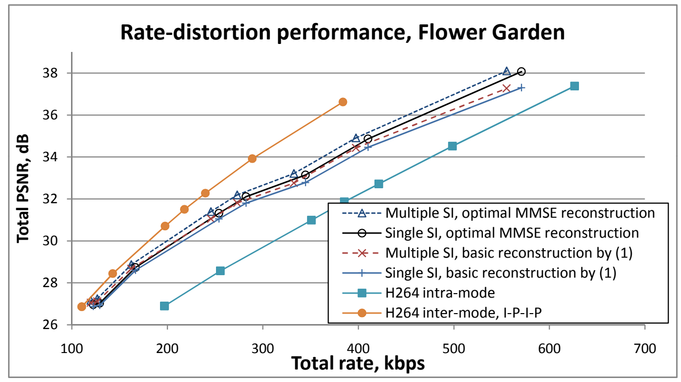
The rate-distortion performance of the proposed optimal reconstruction method has been assessed on several sequences at QCIF/15 Hz, with a GOF size of 2. For each rate point, a value of the quantisation parameter (QP) has been chosen for H.264 Intra-coding the key frames. To each value of the QP parameter corresponds a quantisation matrix for the different subbands resulting from the DCT transform of the Wyner-Ziv frames. These quantisation matrices are given in [5], and have been defined so that key-frames and Wyner-Ziv frames have comparable average quality.

Fig. 4 shows the average PSNR of the Luminance component for both key and Wyner-Ziv frames versus the total bit-rate. We compare our method with the straightforward reconstruction approach (1). For the multiple side information scenario, y in (1) is the average of the two side information hypotheses y_1 and y_2. As the reconstruction does not affect the bit-rate, only the PSNR is improved when using the optimal reconstruction. It can be observed for both single and multiple side information scenarios that the proposed method gains up to 1dB in PSNR compared to the straightforward reconstruction method. Note that using multiple side information improves the rate-distortion performance by up to 0.3dB compared to the case when only one side information is used. We have also plotted the performance of a conventional codec (H.264) both in intra and inter modes for comparison. As for the Wyner-Ziv codec, only the Luminance component was taken into account.

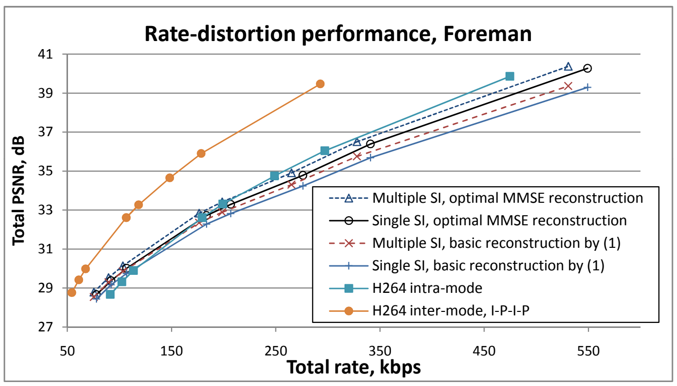
The additional complexity incurred at the decoder is negligible in comparison to the complexity of other decoder modules, due to the simplified calculation approach, proposed in Sections III and IV. As it can be observed from Table I, using the optimal MMSE reconstruction increases the decoding time



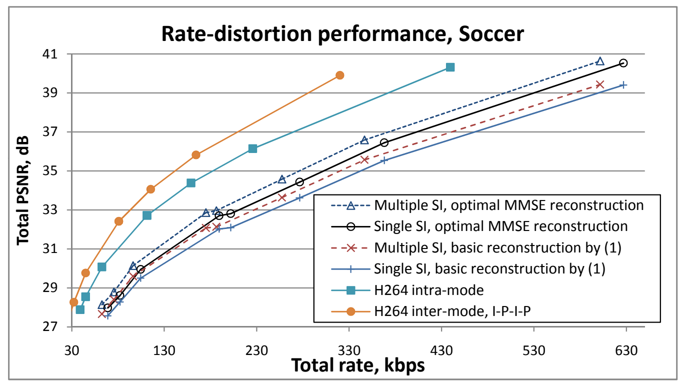
(a) Coastguard sequence, QCIF/15 Hz, 149 frames



(b) Flower Garden sequence, QCIF/15 Hz, 125 frames



(c) Foreman sequence, QCIF/15 Hz, 149 frames



(d) Soccer sequence, QCIF/15 Hz, 149 frames

Figure 4. Rate-distortion performance with optimal MMSE reconstruction.

with single side information only by 0.01 seconds per frame (median value), which constitutes only 0.01%-0.1% of the total decoding time per frame. Here a median is taken instead of an average to better handle outliers of the time measurement resulting from casual operating system activity.

VI. Conclusion

In this paper, a novel optimal reconstruction approach has been presented aimed at minimising mean-squared error (MSE) of reconstructed samples (pixels or DCT coefficients). Its performance has been assessed for a transform domain Wyner-Ziv video codec with both single and multiple side information scenarios, and significant distortion decrease (more than 1dB higher PSNR) was ascertained.

Acknowledgements

We are thankful to the IST and the DISCOVER software development team for having provided the DVC codec software.

References

[1] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. on Information Theory* , Vol. 19, No. 4, 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, January 1976.

[3] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes", *IEEE Comm. Letters*, Vol. 5, pp. 417-419, Oct. 2001.

[4] Da-ke He, En-hui Yang, "On the duality between Slepian-Wolf coding and channel coding", *Proc. of the 2006 IEEE Intl. Symp. Inform. Theory*, Seattle, WA, U.S.A., 2006.

[5] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video", *Proc. SPIE Conference on Visual Communication and Image Processing*, Jan. 2004.

[6] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, "Distributed video coding", *Proc. IEEE, Special issue on advances in video coding and delivery*, vol. 93, No. 1, pp. 71-83, Jan. 2005.

[7] K. Misra, S. Karande, H. Radha, "Multi-hypothesis distributed video coding using LDPC codes", *43rd Annual Allerton Conference on Communication, Control, And Computing*, Monticello, USA, September 2005.

[8] J. Ascenso, C. Brites and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak Republic, July 2005.

[9] D. Kubasov, C. Guillemot, "Mesh-based motion-compensated interpolation for side information extraction in distributed video coding", *Int. Conf. on Image Processing (ICIP)*, Atlanta, USA, October 2006.

[10] C. Brites, J. Ascenso, F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding", *Int. Conf. on Image Processing (ICIP)*, Atlanta, USA, October 2006.

[11] J. Ascenso, C. Brites, F. Pereira, "Motion compensated refinement for low complexity pixel based distributed video coding", *Int. Conf. on Advanced Video and Signal-Based Surveillance*, Como, Italy, September 2005

[12] Y. Vatis, S. Klomp, J. Ostermann, "Enhanced reconstruction of the quantised transform coefficients for Wyner-Ziv coding", *Int. Conf. on Multimedia & Expo (ICME), Multimedia Coding and Processing Track*, Beijing, China, July 2007.