CORRELATION NOISE MODELING FOR MULTIVIEW TRANSFORM DOMAIN WYNER-ZIV VIDEO CODING

Catarina Brites, Fernando Pereira

Instituto Superior Técnico – Instituto de Telecomunicações, Lisbon, Portugal

ABSTRACT

Multiview Wyner-Ziv (MV-WZ) video coding rate-distortion (RD) performance is highly influenced by the adopted correlation noise model (CNM). In the related literature, the statistics of the correlation noise between the original frame and the side information (SI), typically resulting from the fusion of temporally and inter-view created SIs, is modelled by a Laplacian distribution. In most cases, the Laplacian CNM parameter is estimated using an offline approach, assuming that either the SI is available at the encoder or the originals are available at the decoder which is not realistic. In this context, this paper proposes the first practical, online CNM solution for a multiview transform domain WZ (MV-TDWZ) video codec. The online estimation of the Laplacian CNM parameter is performed at the decoder based on metrics exploring both the temporal and inter-view correlations with two levels of granularity, notably transform band and transform coefficient. The results obtained show that better RD performance is achieved for the finest granularity level since the inter-view, temporal and spatial correlations are exploited with the highest adaptation.

Index Terms— Wyner-Ziv coding, multiview video, correlation noise model, online estimation, transform domain

1. INTRODUCTION

Driven by the advances in the capture and display technologies, there is a growing usage of multiview video content, which means the data to be delivered regards many (correlated) views of the same scene. In these systems, a typical predictive coding approach exploits the inter-view correlation at a joint encoder, requiring the various cameras to communicate among them. However, to keep the sensing system simple as required by many applications, the cameras should not have to communicate among them, thus preventing the usage of a predictive coding approach.

Wyner-Ziv (WZ) video coding is a different coding paradigm allowing to exploit the video statistics, partial or totally, at the decoder. Thus, the WZ video coding approach provides a significant architectural benefit to multiview video systems as the WZ based encoders do not need to jointly encode the various views and, thus, inter-camera communication is not needed; this is an important feature for applications with critical bandwidth and power consumption constraints such as visual sensor networks. According to the WZ coding principles [1], [2], the multiview WZ (MV-WZ) decoder is responsible to mostly exploit the intra-view and inter-view correlations and, thus, to achieve efficient compression. In most MV-WZ video codecs available in the literature [3], [4], [5] the side information (SI), a decoder estimate of the original WZ frame, typically results from the fusion of temporally and inter-view created SI frames. To make good use of the created SI for decoding purposes, the decoder needs to have a reliable knowledge of the statistical model characterizing the correlation noise between the original WZ frame (available at the encoder) and the fused SI frame (created at the decoder). This is a challenging task especially because the source statistics do change, not only temporally and spatially but also between views. Thus, the CNM has to take into account not only the intra-view data (as in the monoview scenario) but also the inter-view data to better adapt to the changing statistics between and within views. Most MV-WZ video coding solutions [3]-[6] rely on undesirable or unrealistic assumptions, such as the encoder having a replica of the (complex) SI or the original data being available at the decoder. The unrealistic CNM estimation approach performed at the encoder assuming that the SI is available together with the originals is known as *offline approach*. On the contrary, the realistic and practical so-called *online approach* performs the CNM estimation at the decoder without having access to the original WZ data. While the works in [7] and [8] depart from unrealistic assumptions, both target the (simpler and much less efficient) multiview *pixel domain* WZ video coding scenario, which does not represent the state-of-the-art multiview *transform domain* WZ solutions. In [7], the Laplacian distribution parameter is estimated based only on the residual between the backward and forward motion compensated frames. In [8], the Laplacian distribution parameter is estimated at the block level based on the residual between the backward and forward motion compensated frames and the residual between the left and right SI frames.

In this context, this paper proposes the first methods to online estimate the CNM for state-of-the-art multiview *transform domain* WZ coding based on both intra-view and inter-view correlation data, in this case the motion and disparity compensated residuals obtained at the decoder. Two levels of estimation granularity are proposed: transform band and transform coefficient. While both estimation granularity levels exploit the intra-view and inter-view correlations, only the finest level adapts the correlation noise statistics locally within the frame, thus allowing obtaining the best performance.

To achieve its purposes, this paper is organized as follows: Section 2 presents a brief overview of the MV-TDWZ video codec adopted in this paper. Section 3 motivates the adoption of the Laplacian CNM used and proposes the novel online CNM parameter estimation methods. Section 4 presents and discusses the experimental results and, finally, Section 5 closes the paper with final remarks and future work directions.

2. THE MULTIVIEW TDWZ VIDEO CODEC

The MV-TDWZ video codec architecture adopted in this paper is illustrated in Fig. 1 and is based on the monoview transform domain WZ video codec presented in [9]. As common in the literature, a 3-camera setup has been adopted. The MV-TDWZ video codec works as follows:

- The (left and right) lateral views and the central view key frames are H.264/AVC Intra coded, a common practice in the related literature, e.g. [3], [5]. At the central view WZ encoder, the WZ frames are encoded with a 4×4 integer DCT transform, a uniform quantizer and a bitplane based turbo encoder [9]; the number of WZ frames in between two key frames (+1) defines the group of pictures (GOP) size. At the decoder, the key frames are first H.264/AVC Intra decoded and the temporal SI, SI-MCFI, is then

This work was supported by FCT project MUVIS with reference PTDC/EEA-TEL1119804/2010.

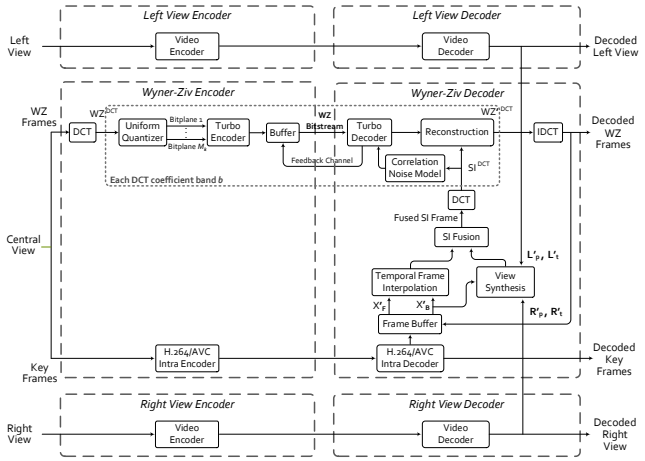

Fig. 1. The 3-camera setup MV-TDWZ video codec architecture.

created through the popular motion compensated frame interpolation (MCFI) technique [10] using backward/past (X'_B) and forward/future (X'_F) decoded frames.

- The central view decoder also receives from the left (resp. right) view decoder, two (decoded) reference frames for time instants p and t, L'_p and L'_t (resp. R'_p and R'_t), which are used by the view synthesis (VS) module to generate the inter-view SI frame, SI-VS; while t corresponds to the time instant of the current WZ frame to be decoded, p is associated to the X'_B frame time instant. Basically, the inter-view SI creation technique starts by compensating the illumination variations between the current WZ frame and L'_p (resp. R'_p). Then, block-wise disparity search, constrained to the epipolar geometry, is performed between the reference frames L'_t and R'_t. 3D reconstruction is then applied to the block correspondence pairs (obtained through disparity search), filling the corresponding block locations in the central view frame (SI-VS frame). Since the inter-view SI creation technique accounts for the scene geometry, it allows creating high quality inter-view SI.
- The SI-MCFI and SI-VS frames are then fused to create a high quality SI frame. Basically, the SI fusion approach performs pixel-based foreground detection between the central view reference frames used to create the SI-MCFI (X'_B and X'_F), to identify pixel intensities which have significantly changed in time. In an analogous way, foreground detection is also performed between L'_t (resp. R'_t) and L'_p (resp. R'_p). The resulting foreground detection is then projected onto the central view, according to the disparity field obtained from the SI-VS frame creation process, thus allowing to identify the pixel intensities which have significantly changed between views. The temporal and inter-view foreground detection results are then merged and used to drive the fused SI frame creation process.
- The fused SI is 4×4 DCT transformed and converted to soft probabilities (needed for turbo decoding) with the help of the derived CNM. The CNM follows a Laplacian distribution (as usual in the MV-WZ video coding literature [5]) whose parameter is estimated as proposed in Section 3. The turbo decoder uses the soft probabilities to decode the source, in this case each DCT band bitplane. Finally, all DCT bands are reconstructed [11] and the inverse DCT is carried out to obtain the reconstructed WZ frame.

3. ONLINE CORRELATION NOISE MODELING

Before presenting the novel CNM parameter estimation methods aforementioned, a brief motivation for the Laplacian CNM adopted here is provided in the following.

3.1. Laplacian CNM in MV-WZ Video Coding

As mentioned before, the Laplacian distribution as in (1) is widely used to model the correlation noise in the (multiview and monoview) WZ video coding literature, e.g. [5], [12].

$$f[WZ(\boldsymbol{p}) - SI(\boldsymbol{p})] = \frac{\alpha}{2} e^{-\alpha|WZ(\boldsymbol{p})-SI(\boldsymbol{p})|} \quad (1)$$

In (1), f[.] stands for the probability density function, $\boldsymbol{p}$=(x, y) is the position to be evaluated within the WZ and (fused) SI frames and α is the Laplacian distribution parameter defined by

$$\alpha = \sqrt{2/\sigma^2} \quad (2)$$

where σ^2 is the variance of the residual between the original WZ and (fused) SI data. Fig. 2 depicts the actual histogram of the (WZ – fused SI) residual for the *Outdoor* full sequence at 256×192 resolution, 15 Hz, GOP 2. A Laplacian distribution with curve fitting determined α value equal to 0.37 is also plotted in Fig. 2.

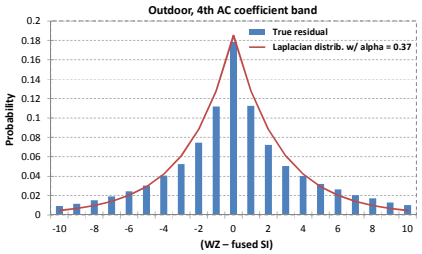

Fig. 2. Residual histogram for the *Outdoor* sequence, 256×192 resolution, 15 Hz, GOP 2.

The hypothesis that the Laplacian distribution fits the (WZ – fused SI) histogram was validated by using the chi-square goodness-of-fit test; a chi-square value of 0.08 was obtained for the *Outdoor* sequence in Fig. 2. For a typical significance level of 5%, this value implies accepting the hypothesis that (WZ – fused SI) follows a Laplacian distribution; the same conclusion was obtained for all the video sequences and test conditions evaluated in this paper.

Since the SI quality varies along time and space, the choice of the elementary unit (or granularity level) at which the α parameter is estimated is an important decision to achieve a good MV-WZ video coding RD performance. Pursuing a coarse to fine strategy to estimate the α parameter allows to evaluate the impact of the CNM granularity level in the RD performance as the CNM increasingly exploits the changing statistics between and within frames.

Two online Laplacian distribution parameter estimation methods, working at the decoder at two different granularity levels, are proposed here for the MV-TDWZ video codec: the DCT band level and the DCT coefficient level. In the DCT band level approach, one α value is estimated for each DCT band within a frame. As these α values are updated for each WZ frame and, thus, vary along the video sequence, the temporal variation of the correlation noise statistics is considered. In the coefficient level approach, one α value is estimated for each DCT coefficient within each WZ frame, leading to a finer adaptation of the CNM, both spatially (within a frame) and temporally. These two Laplacian distribution parameter estimation methods make use of the X'_B, X'_F, L'_t and R'_t frames along with the motion and disparity vectors obtained during the SI-MCFI and SI-VS creation processes. To the best of the authors' knowledge, there are no algorithms available in the literature to realistically and practically online estimate the CNM parameters for MV-TDWZ video coding.

3.2 Correlation Noise Estimation at DCT Band Level

To online model the correlation noise at the DCT band level, the

following method is proposed to estimate one α parameter per each DCT band of each WZ frame:

1. *Temporal residual frame generation:* First, the residual frame N_{time} between the motion compensated versions of the X'_B and X'_F frames is computed as:

$$N_{time} = 0.5\left(X'_F(x + dx_F, y + dy_F) - X'_B(x + dx_B, y + dy_B)\right) \quad (3)$$

where $X'_B(x + dx_B, y + dy_B)$ and $X'_F(x + dx_F, y + dy_F)$ represent the X'_B and X'_F motion compensated frames, respectively, and (x, y) corresponds to the pixel location in the N_{time} frame. In (3), (dx_B, dy_B) and (dx_F, dy_F) represent the motion vectors for the X'_B and X'_F frames, respectively. The N_{time} frame corresponds to the matching success of the estimated motion field between X'_B and X'_F and intends to express how much two (motion compensated) blocks in the references frames are similar.

2. *Inter-view residual frame generation:* In a similar way, the residual frame N_{view} between the disparity compensated versions of L'_t and R'_t frames (time aligned with the WZ frame being decoded) is computed as:

$$N_{view} = L'_t\left(x + dx_{L,t}, y + dy_{L,t}\right) - R'_t\left(x + dx_{R,t}, y + dy_{R,t}\right) \quad (4)$$

where $L'_t(x + dx_{L,t}, y + dy_{L,t})$ and $R'_t(x + dx_{R,t}, y + dy_{R,t})$ represent the L'_t and R'_t disparity compensated frames, respectively, and $(dx_{L,t}, dy_{L,t})$ and $(dx_{R,t}, dy_{R,t})$ represent the disparity vectors for the L'_t and R'_t frames, respectively. The N_{view} frame corresponds to the matching success of the estimated disparity field between L'_t and R'_t and expresses how much two corresponding (disparity compensated) blocks in the lateral views frames are similar.

3. *Temporal and inter-view residual frames fusion:* The two residual frames previously computed are merged into a single residual frame N_{fused} according to the SI fusion process. In other words, if the fused SI pixel is obtained from the SI-MCFI (resp. SI-VS) frame, the co-located N_{fused} frame pixel is obtained from the N_{time} (resp. N_{view}) frame; this way, both inter-view and temporal correlation data are exploited in the CNM process. The N_{fused} frame intends to express how much two (motion/disparity compensated) blocks in the respective references frames are similar; therefore, the N_{fused} frame can be taken as an estimate of the true residual frame computed between the original WZ frame and the (fused) SI.

4. *N_{fused} frame DCT transform:* A 4×4 DCT is applied over N_{fused} to obtain frame $|T_{fused}|$, whose elements are the absolute values of the corresponding elements in the DCT coefficients frame.

5. *$|T_{fused}|$ frame DCT band b variance computation:* The variance of each $|T_{fused}|$ frame DCT band b, $\hat{\sigma}_b^2$, is computed.

6. *DCT band b α parameter estimation:* The α parameter for each DCT band b, $\hat{\alpha}_b$, is estimated as

$$\hat{\alpha}_b = \sqrt{2/\hat{\sigma}_b^2}. \quad (5)$$

3.3. Correlation Noise Estimation at DCT Coefficient Level

To further adapt to the (fused) SI changing statistics, the Laplacian distribution parameter may be online estimated at the DCT coefficient level, i.e. each DCT coefficient of each DCT band will have a specific α parameter associated to it. Here Steps 1 to 5 are equal to the steps described in Section 3.2.

6. *DCT coefficient classification:* Each $|T_{fused}|$ frame DCT coefficient (of each DCT band) is classified as *inlier* or *outlier* depending whether the squared distance $D_b(u, v)$ between it and the corresponding $|T_{fused}|$ frame DCT band average value μ_b,

$$D_b(u, v) = (|\mathrm{T}_b|(u, v) - \mu_b)^2 \quad (6)$$

is lower or higher that the respective $|T_{fused}|$ frame DCT band variance (obtained in Step 5), respectively. This classification intends to identify and cluster coefficients belonging to regions which share similar 'difficulty' in the SI creation process; this way, it is possible to better adapt the CNM to the SI statistics.

7. *DCT coefficient α parameter estimation:* The (u, v) DCT coefficient α parameter is estimated as

$$\hat{\alpha}_b(u, v) = \begin{cases} \hat{\alpha}_b & , D_b(u, v) \le \hat{\sigma}_b^2 \\ \sqrt{2/D_b(u, v)} & , D_b(u, v) > \hat{\sigma}_b^2. \end{cases} \quad (7)$$

For the inlier DCT coefficients, the α parameter is given by the respective DCT band α parameter (estimated from Step 6 in Section 3.2); inlier DCT coefficients typically belong to regions well interpolated and, thus, the DCT band level $\hat{\alpha}_b$ parameter is a reliable estimation. For the outlier coefficients, the α parameter is estimated from (2) where $D_b(u, v)$ is used instead of σ_b^2; outlier DCT coefficients typically correspond to regions where the residual error is high, which means that the SI creation process (temporal or inter-view) failed for those regions. Thus, the DCT band level $\hat{\alpha}_b$ parameter estimate is not the best approach in this case as it corresponds to an average value for all DCT coefficients within band b. By using $D_b(u, v)$ instead of $\hat{\sigma}_b^2$ to estimate the DCT coefficient α parameter, the turbo decoder runs with less confidence for the DCT coefficients belonging to blocks where the MCFI and the inter-view SI creation algorithms essentially failed.

4. EXPERIMENTAL RESULTS

To evaluate the RD performance impact as the CNM better exploits the changing statistics between and within frames, four multiview video sequences at 15Hz with 256×192 spatial resolution are considered: *Breakdancers*, *Ballet* [13], *Book Arrival* and *Outdoor* [14]; while the first two are non-rectified sequences, the last two are rectified. *Breakdancers* and *Ballet* correspond to high and low motion content, respectively, where cameras are arranged along a 1D arc. *Book Arrival* and *Outdoor* are typical indoor and outdoor video surveillance sequences, respectively, where cameras are arranged along a straight line with parallel optical axis (1D parallel). Besides representing some content variety, these sequences also exhibit different camera arrangements, which is important to obtain representative and meaningful performance results. The first three views of each sequence are used and all frames of each sequence are considered which means 100 frames for each view of each sequence. The test conditions for the DCT, quantizer, temporal frame interpolation, turbo codec and reconstruction modules are the same as in [9]. The key frames as well as the lateral views frames are H.264/AVC Intra coded (Main profile) [15] with a quantization parameter (QP) depending on the RD point. All experiments are conducted only for the luminance component, as usual in WZ video coding, and GOP sizes of 2, 4 and 8 are considered. Bitrate (resp. PSNR) includes both the luminance rate (resp. PSNR) for WZ and key frames of the central view (WZ) to be coded since the left and right views PSNR and rates are always the same; this allows better assessing the RD performance of the WZ central view coding.

Table 1 shows the obtained RD performance expressed in terms of Bjontegaard metrics for the four multiview sequences for GOP sizes of 2, 4 and 8. For each GOP size, the 1st row shows the Bjontegaard difference (both bitrate saving, BD-Rate, and PSNR gain, BD-PSNR) between the online DCT coefficient and the DCT band levels; the 2nd row shows the Bjontegaard difference between

Table 1. DCT band and coefficient levels Bjontegaard bitrate savings and PSNR gains at 256×192 resolution, 15 Hz, GOPs of 2, 4 and 8.

		Breakdancers		Ballet		Outdoor		Book Arrival	
		BD-Rate [%]	BD-PSNR [dB]	BD-Rate [%]	BD-PSNR [dB]	BD-Rate [%]	BD-PSNR [dB]	BD-Rate [%]	BD-PSNR [dB]
GOP 2	Online: Coeff. vs. Band	-4.94	0.28	-5.06	0.33	-0.22	0.01	-1.58	0.10
	Offline: Coeff. vs. Band	-26.97	1.37	-22.77	1.33	-6.85	0.46	-12.00	0.72
	Band: Online vs. Offline	0.03	0.00	2.84	-0.19	4.78	-0.34	2.47	-0.16
	Coeff.: Online vs. Offline	18.38	-1.15	16.34	-1.16	11.50	-0.80	11.58	-0.79
GOP 4	Online: Coeff. vs Band	-8.11	0.44	-10.77	0.64	-5.19	0.35	-4.07	0.26
	Offline: Coeff. vs Band	-45.30	2.09	-36.30	1.94	-10.22	0.69	-20.00	1.17
	Band: Online vs. Offline	0.04	0.00	3.63	-0.23	9.03	-0.68	3.49	-0.24
	Coeff.: Online vs. Offline	25.62	-1.71	21.59	-1.65	14.62	-1.05	15.99	-1.20
GOP 8	Online: Coeff. vs Band	-9.60	0.51	-14.35	0.81	-5.19	0.39	-5.71	0.35
	Offline: Coeff. vs Band	-57.84	2.52	-47.01	2.32	-11.60	0.78	-25.10	1.40
	Band: Online vs. Offline	0.07	-0.01	2.92	-0.18	10.18	-0.74	4.33	-0.27
	Coeff.: Online vs. Offline	31.49	-2.03	25.67	-1.84	14.49	-1.12	19.15	-1.35

the offline DCT coefficient and the DCT band levels. The online approach refers to the methods proposed in this paper while the offline approach (coefficient and DCT band levels) corresponds to the widely used methods proposed in [12]. The offline approach performance provides insights on the optimal CNM parameter estimation performance that can be achieved when the correlation noise is obtained using both the original WZ data and (fused) SI; due to space limitations, the reader is referred to [12] for details on the offline approach. For each GOP size, the last two rows show the Bjontegaard difference when the online approach is used instead of the offline one, for the coarser (DCT band) and the finer (coefficient) granularity levels, respectively: this is, in fact, the penalty of using a realistic CNM approach. In Table 1, the "minus" sign in 'BD-Rate' column means that there is a bitrate saving while "minus" sign in 'BD-PSNR' column means that there is a PSNR drop.

As explained in Section 1, the finer the estimation granularity level, the better the CNM should adapt to the changing statistics between and within frames. This behavior is confirmed in Table 1 (1st and 2nd rows of each GOP size) where there is a bitrate decrease as the online (as well as offline) estimation granularity level gets finer (band to coefficient); for the online α estimation methods, bitrate savings of 5.1%, 10.8% and 14.4% for GOP of 2, 4 and 8, respectively (at maximum), are observed. The band-to-coefficient bitrate savings for the offline approach are higher because in this case the correlation noise is obtained using both the original WZ data and the (fused) SI and, thus, a more accurate CNM can be obtained. The compression gains are especially higher under critical conditions, such as larger GOP sizes and complex cameras arrangements, where poorer SI quality is typically created and thus a finer granularity estimation level is essential to better adapt the CNM to the changing statistics.

As expected, the (realistic) online CNM parameter estimation methods proposed in Section 3 have a compression efficiency loss when compared with the corresponding (unrealistic) offline ones. This behavior is explained by the fact that with the offline approach unrealistically the decoder uses original WZ frames while with the realistic online techniques the decoder only uses an estimation of the true residual variance given the motion and the disparity compensated residuals. The bitrate losses are up to 4.8%, 9.0% and 10.2% for GOP sizes of 2, 4 and 8, respectively, for the DCT band level, and up to 18.4%, 25.6% and 31.5% for GOP sizes of 2, 4 and 8, respectively, for the DCT coefficient level.

For completeness purposes, Fig. 3 illustrates the RD performance obtained with the adopted MV-TDWZ video codec when the CNM parameter is estimated using both online and offline approaches as well as the performance attained with H.264/AVC Intra, which provides here a predictive video coding benchmark with an encoding complexity similar to the MV-TDWZ video codec in the sense that no motion estimation is performed at the encoder. There is no comparison with alternative CNM solutions in the literature as this is the first solution for MV-TDWZ. Comparing with the H.264/AVC Intra coding performance, the MV-TDWZ RD performance using online modeling is similar to or above the H.264/AVC Intra curve for all test sequences, independently of the granularity level used to estimate the α parameter.

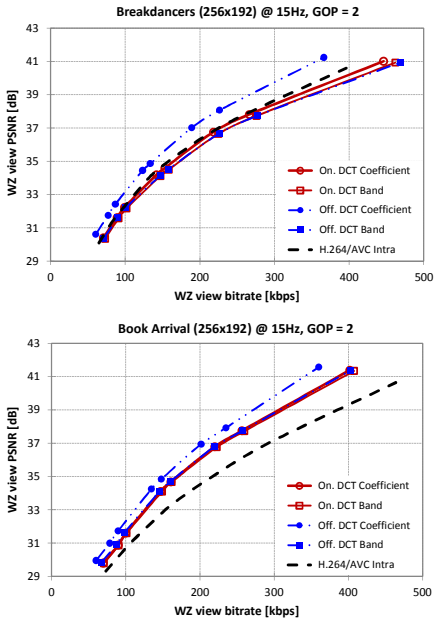

Fig. 3. RD performance for the *Breakdancers* and *Book Arrival* sequences, 256×192, 15 Hz, GOP 2.

5. FINAL REMARKS

This paper proposes the first realistic, online correlation noise modeling solution for MV-TDWZ video coding considering two granularity levels. Experimental results show that better RD performance is achieved for the finest granularity level since the inter-view, temporal and spatial correlations are exploited at the highest adaptation capability. Future work will consider combining the proposed online α estimation methods with SI block structural similarity information to improve the MV-TDWZ video codec RD performance.

6. REFERENCES

[1] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions 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 Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[3] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Iterative multiview side information for enhanced reconstruction in distributed video coding," *EURASIP Journal on Image and Video Processing*, *Special Issue on Distributed Video Coding*, vol. 2009, Article ID 591915, 17 pages. doi:10.1155/2009/591915.

[4] T. Maugey, W. Miled, and B. Pesquet-Popescu, "Dense disparity estimation in a multi-view distributed video coding system," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, pp. 2001–2004, April 2009.

[5] F. Dufaux, "Support vector machine based fusion for multi-view distributed video coding," *International Conference on Digital Signal Processing*, Corfu, Greece, July 2011.

[6] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multi-view video coding," *SPIE Visual Communications and Image Processing*, San Jose, CA, USA, January 2006.

[7] Y. Li, S. Ma, D. Zhao, and W. Gao, "Modeling correlation noise statistics at decoder for multi-view distributed video coding," *IEEE International Symposium on Circuits and Systems*, Taipei, Taiwan, pp. 2597–2600, May 2009.

[8] K. Xu and S. Fang, "Accurate correlation noise modeling for pixel domain distributed multiview video coding," *International Symposium on Intelligent Signal Processing and Communication Systems*, Chengdu, China, December 2010.

[9] C. Brites, J. Ascenso, J. Q. Pedro, and 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.

[10] J. Ascenso, C. Brites, and F. Pereira, "Content adaptive Wyner-Ziv video coding driven by motion activity," *IEEE International Conference on Image Processing*, Atlanta, GA, USA, pp. 605–608, October 2006.

[11] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner-Ziv video coding with multiple side information," *IEEE International Workshop on Multimedia Signal Processing*, Chania, Crete, Greece, pp. 183–186, October 2007.

[12] C. Brites and 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.

[13] [Online].http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/

[14] "HHI Test Material for 3D Video," ISO/IEC JTC1/SC29/WG11, MPEG08/M15413, Archamps, France, May 2008.

[15] Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding, ISO/IEC International Standard 14496-10:2003