

Multi-hypothesis Transform Domain Wyner-Ziv Video Coding including Optical Flow

Xin Huang^{1#}, Lars Lau Rakêt^{2*}, Huynh Van Luong[#], Mads Nielsen^{*}, François Lauze^{*}, Søren Forchhammer[#]

[#]*DTU Fotonik, Technical University of Denmark,*

Building 343, Lyngby 2800, Denmark

¹xhua@fotonik.dtu.dk

^{*}*Department of Computer Science, University of Copenhagen,*

Universitetsparken 1, Copenhagen 2100, Denmark

²larslau@diku.dk

Abstract—Transform Domain Wyner-Ziv (TDWZ) video coding is an efficient Distributed Video coding solution providing new features such as low complexity encoding, by mainly exploiting the source statistics at the decoder based on the availability of decoder side information. The accuracy of the decoder side information has a major impact on the performance of TDWZ. In this paper, a novel multi-hypothesis based TDWZ video coding is presented to exploit the redundancy between multiple side information and the source information. The decoder used optical flow for side information calculation. Compared with the best available single estimation mode TDWZ, the proposed multi-hypothesis based TDWZ achieves robustly better Rate-Distortion (RD) performance and the overall improvement is up to 0.6 dB at high bitrate and up to 2 dB compared with the DISCOVER TDWZ video codec.

I. INTRODUCTION

Distributed Video Coding (DVC) [1] provides a video coding paradigm which fully or partly exploits the temporal redundancy of video at the decoder, instead of at the encoder as in predictive video coding, thereby shifting computational requirements from encoder to decoder. This may be of interest when communicating video from mobile devices. Further the use of distributed source coding also provides flexibility on the decoder side. DVC is based on two major information theoretic results: the Slepian-Wolf [2] and Wyner-Ziv [3] theorems. The Slepian-Wolf theorem proves that, two statistically dependent discrete random sequences (X, Y) which are independently and identically distributed (i.i.d.) may be independently encoded but jointly decoded, at the same rate as for joint encoding and decoding. The Wyner-Ziv theorem extends the Slepian-Wolf theorem to lossy source coding of X based on side information Y at the decoder. This suggests that a novel video coding system, which encodes individual frames independently, but decodes them jointly, may achieve low complexity encoding with similar coding efficiency as conventional hybrid predictive video coding, notably if X and Y are jointly Gaussian and a mean-square error distortion measure is considered.

Transform Domain Wyner-Ziv (TDWZ) video coding [4] is one efficient approach to DVC. Its coding efficiency is highly dependent on the accuracy of side information at the decoder. Most TDWZ video codecs are based on a single side

information estimation mode. For example, in [5]-[9], there is one soft-input estimate available at the Wyner-Ziv decoder, obtained from side information frame generation and noise modeling. Although the quality of side information frames and the accuracy of the noise model have been improved in [6][8][9], the coding efficiency of the single estimation mode TDWZ trails that of conventional video coding solutions, such as H.264/AVC, most notable in high motion sequences. Related work about multiple side information based TDWZ have been proposed in [10][11]. In [10], two different frame interpolation methods are employed, but the Wyner-Ziv decoder is only dealing with the average of two estimates for decoding and reconstruction. In [11], a set of weighted multiple soft-inputs are firstly developed in TDWZ, which is based on one frame interpolation and one frame extrapolation technique. However, the contribution brought by frame extrapolation is rather limited to the multiple soft inputs and the reconstruction in [11] was only based on one side information. Calculations of multiple side information and multi-hypothesis may also be considered in a distributed computing environment.

In order to enhance performance and reduce the RD gap between TDWZ and conventional video coding, which is especially pronounced in high motion sequences, a novel multi-hypothesis based TDWZ decoder is introduced in this paper. The multiple side information is generated by both block based and optical flow based side information generation techniques. The intuition is that optical flow based frame interpolation can generate different side information and compensate the estimation weakness in block based methods. An additional contribution of this paper is that the multiple soft-inputs for decoding and reconstruction are based on a weighted joint distribution in contrast to [11]. In this way, the proposed multi-hypothesis based TDWZ decoder will not only reduce the required bitrate for decoding but also improve the quality of reconstructed frames.

The rest of this paper is organized as follows: Section II briefly describes the state-of-the-art TDWZ video coding with single side information estimation. In Section III, the proposed Wyner-Ziv decoder is introduced. Finally, the performance results are presented in Section IV.

II. TRANSFORM DOMAIN WYNER-ZIV VIDEO CODING WITH SINGLE SIDE INFORMATION

In a TDWZ video codec with single side information estimation mode, the frame sequence is split into key frames and so-called Wyner-Ziv frames. Key frames are intra coded using conventional video coding techniques such as H.264/AVC intra coding. The Wyner-Ziv frames X are transformed, quantized and decomposed into bit planes. Each bit plane is fed to a rate-compatible LDPC Accumulate (LDPCA) encoder [12] from most significant bit plane (MSB) to least significant bit plane (LSB). Corresponding error correcting information is stored in a buffer. The amount of bits to be transmitted depends on the requests made by the decoder through a feedback channel. The original Wyner-Ziv frame available at the encoder is estimated at the decoder side by using already decoded frames as references. An adaptive weighted Overlapped Block Motion Compensation (OBMC) based frame interpolation scheme [6] is employed to generate an estimated side information frame Y . With the obtained estimation, soft-input information within each bit plane is estimated using a cross-band based noise model [13]. Thereafter the LDPCA decoder starts to decode the various bit planes, ordered from MSB to LSB, to correct the bit errors. After all the bit planes are successfully decoded, the Wyner-Ziv frame can be decoded through combined de-quantization and reconstruction [10] followed by an inverse transform.

In single side information estimation mode TDWZ video coding, the certainty of soft-input information plays a key role in terms of overall RD performance. The soft-input Pr is defined as a conditional probability of each bit b_i being equal to 0 or 1, i.e. $Pr = P(b_i | y, f_{xy}, b^-)$, where y denotes the corresponding estimated side information value in transform domain for bit b_i , f_{xy} is an estimated probability density function obtained from the adopted noise model and b^- is the information from the previously decoded bit planes. The quality of the reconstructed frame is highly dependent on the accuracy of the estimated noise distribution f_{xy} [10]. The TDWZ video codec with single side information estimation mode presented in this section is considered as the best available TDWZ codec [13]. As seen in Section IV, this gives better performance than the DISCOVER video codec due to better frame interpolation [6] and noise model [13]. For a detailed description, we refer to [13].

III. MULTI-HYPOTHESIS BASED WYNER-ZIV DECODING

As described in Section II, the essential aspects to improve the coding efficiency of TDWZ are the certainty of the soft-input information fed into the LDPCA decoder and the accuracy of the noise distribution for frame reconstruction. To address these issues, a multi-hypothesis based Wyner-Ziv decoding is proposed. The Wyner-Ziv encoder is not changed, as the basic idea is to generate $M (>1)$ different side information frames Y_k , $k \in \{1, M\}$, at the decoder for each Wyner-Ziv frame. Each side information frame is considered as an observation of the original Wyner-Ziv frame X with a different amount of noise. Processing multiple side

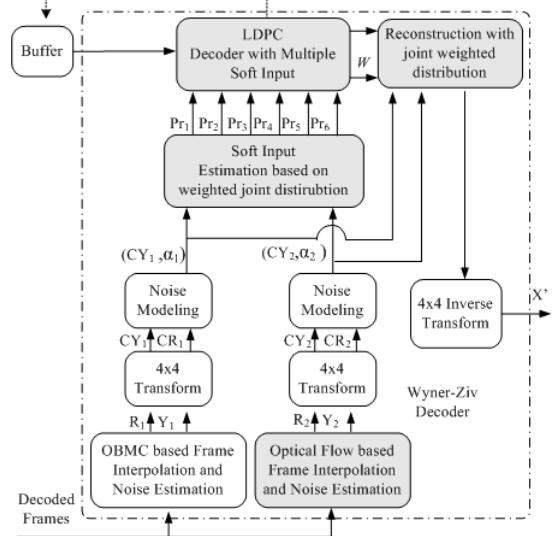


Fig. 1. Architecture of multi-hypothesis TDWZ video codec based on two frame interpolation schemes

information frames may reduce the bitrate required and improve the quality of reconstructed frame.

The architecture of the proposed Wyner-Ziv video decoder with an example of two side information generation schemes ($M=2$) is presented in Fig. 1. In principle, there can be any number of competitive side information generation schemes at the proposed Wyner-Ziv decoder. The two different interpolation methods shown in Fig. 1 are the OBMC based frame interpolation introduced in [6] and a novel optical flow based frame interpolation method described in Section III-A. Each side information generation scheme not only creates an estimation of Wyner-Ziv frame, Y_k , but also an estimated noise residue frame R_k . R_k is used to express the correlation noise between the Wyner-Ziv frame X and its estimated side information frame Y_k . (We refer to Section III-A for more details.)

R_k is used to calculate the parameter of the noise distribution $f_{x|y_k}$ outlined. The estimated noise residue R_k and side information frame Y_k undergo the same 4x4 block DCT. Taking its corresponding transform coefficients as inputs for a coefficient level noise model [13], the noise distribution $f_{x|y_k}$ between the estimated side information frame Y_k and Wyner-Ziv frame X in transform domain is modeled by Laplace distributions. Each transform coefficient in a given band b_l is assigned with an estimated Laplace distribution parameter $\alpha_k^{b_l}(m, n)$, where (m, n) are the coordinates of the corresponding DCT coefficient. The value of the Laplacian parameter expresses the reliability of the corresponding estimated side information frame.

With Laplacian parameters based on different calculations of Y_k , multiple soft-inputs are calculated based on a proposed weighted joint distribution (described in Section III-B). All the hypotheses of soft-input are fed into a modified LDPCA decoder. The soft-input that converges first is chosen as the best candidate soft input for LDPCA decoding. Subsequently, the corresponding weighted joint distribution is given to the

reconstruction module for minimum mean-square error reconstruction [10]. More details about the soft-input calculation and the reconstruction based on the joint weighted distribution are described in Section III-B.

A. Optical Flow based Frame Interpolation

The goal of optical flow estimation is to determine the apparent motion in a given video sequence, and the optical flow between two video frames I_0 and I_1 is the displacement field v between these, i.e. in a suitable sense

$$I_1(\mathbf{z} + v(\mathbf{z})) = I_0(\mathbf{z}). \quad (1)$$

where $\mathbf{z} \in \mathbb{R}^2$ denotes the image coordinates.

In recent years a large number of techniques have been proposed to determine optical flow, and the accuracy of optical flow algorithms has improved tremendously [14]. We use an optical flow algorithm that determines the optical flow by minimizing the energy functional OF given by

$$OF(v) = \lambda \int \|I_1(\mathbf{z} + v(\mathbf{z})) - I_0(\mathbf{z})\|_2 d\mathbf{z} + \int \|\nabla v(\mathbf{z})\|_2 d\mathbf{z}. \quad (2)$$

This functional consists of a robust L^1 norm of the data fidelity term (1) and a total variation (TV) regularization of the flow v . The parameter λ determines the tradeoff between data fidelity and regularity of the estimated optical flow, and has been set to 10 in all experiments. The minimizing flow is calculated using a highly efficient method, relying on tools from non-smooth convex analysis in a coarse-to-fine pyramid setup, which was introduced in [15], and further improved in [16]. Here we use the very efficient GPU implementation described in [17], which allows for the inclusion of higher order terms in the frames I_0 and I_1 , e.g. gradient information either alone or in conjunction to the luminance information. This can improve flow quality under difficult lighting conditions [18], but since the lighting is fairly constant in the test sequences, gradient information has not been included in the evaluation in Section IV. In general this method for calculating optical flow should be more robust under high motion than block based methods, as the estimation procedure is done in a spatial continuous setting and considers the frames on a number of scales, to align structures of different size. Furthermore it should handle luminance patterns undergoing strong deformation better, as neighborhood relations are less rigid when using TV regularization compared to the constraints imposed by blocks. On the other hand block based methods will typically perform better in low motion sequences. This is because the optical flow algorithm is fine-tuned for complex motion, so it will be more disposed to falsely interpreting small changes in the video frames as motion.

For the interpolation, the forward flow v_f is calculated between two consecutive decoded key frames and the backward flow v_b is calculated between the same frames in reverse temporal order. For each of the flows an estimate of the intermediate frame can be produced as the one transporting the brightness patterns half way along the flow lines between the two key frames. This has the unfortunate

consequence that the interpolated frames will be blank in regions undergoing disocclusion, since the optical flow will map away from these. To avoid this problem, one can utilize the principle that first following a forward flow line, and then taking the backward flow should bring one back to the starting point

$$v_b(\mathbf{z} + v_f(\mathbf{z})) + v_f(\mathbf{z}) = 0. \quad (3)$$

This identity only holds as long as we stay away from occlusion, but ignoring this, one can translate the coordinate system by v_b for the forward flow and v_f for the backward flow, which gives that the intermediate frame interpolations Y_f and Y_b can be calculated as

$$Y_f(\mathbf{z}) = I_0(\mathbf{z} + 1/2v_b(\mathbf{z})) \quad \text{and} \quad Y_b(\mathbf{z}) = I_1(\mathbf{z} + 1/2v_f(\mathbf{z})) \quad (4)$$

in every pixel point \mathbf{z} , where the key frame evaluation points are rounded to nearest pixel. If a region of points in I_0 is undergoing disocclusion along the forward flow lines to Y_f , the equation (3) will not hold, but using (4) will then just assign the disoccluded region in Y_f with similar (in terms of value) nearby values in I_0 , because when looking along the backward flow lines, the disocclusion will be an occlusion, and this occluded region will be mapped to a nearby region in I_0 that has similar values, since we are minimizing the data fidelity term (1). This means that we automatically have an implicit inpainting of holes in this scheme. Finally, the interpolated frame Y will be the average of Y_f and Y_b

$$Y(\mathbf{z}) = \frac{1}{2}(Y_f(\mathbf{z}) + Y_b(\mathbf{z})) \quad (5)$$

and the noise residue frame R is calculated as the difference between the two estimated frames

$$R(\mathbf{z}) = Y_f(\mathbf{z}) - Y_b(\mathbf{z}). \quad (6)$$

Given the forward and backward flows, this interpolation scheme is quite simple, but the process is independent of the flow calculations, so one could improve results by using a more sophisticated optical flow algorithm without altering its internal mechanics, e.g. a flow calculated using the excellent framework proposed by Xu et al. [19] would likely produce significantly better interpolations than the method used here [14].

A number of alternative flow based interpolation methods have been compared to the one presented here. Interpolation based on direct interpolation along the flow lines and subsequent calculation of Y as a locally weighted average of the two intermediate frames, with weights determined from information about occluded/disoccluded areas, proved to give results that were marginally worse than (5). The more elaborate scheme presented in [20], where motion and interpolation are estimated simultaneously has been implemented, but while it provides interpolated frames that are visually more pleasing than (5), because the forward and backward flow will converge to a common interpolation, they are quantitatively inferior in terms of difference to the real frames.

B. Multi-hypothesis Soft-input and Reconstruction based on Joint Weighted Distribution

With the obtained noise distribution $f_{x|y_k}$ for each individual side information observation Y_k , a joint weighted distribution F is defined as:

$$F_j = \sum_{k=1}^M w_{jk} f_{x|y_k} \quad (6)$$

where $j, j \in [1, N]$ denotes the index of candidate joint weighted distribution, N is the total number of joint distributions employed candidates, and w_{jk} denotes the j th predefined weighting parameter on side information k , $k \in [1, M]$ and $\sum_{k=1}^M w_{jk} = 1$. M is the total number of distinct side information frames available at the decoder. (As the example shown in Fig.2, $M=2, N=6$).

As the OBMC based frame interpolation scheme [6] gives better results on the different test sequences compared to the other side information techniques employed in this paper (shown in Table I), the soft input calculation is only based on the joint weighted distribution within a specific unreliable region specified by the set map . Outside of the map region, the side information is given by the OBMC based scheme. The values of the Laplacian parameters should express the reliability of the corresponding side information frame, thus an unreliable set S_k of each single side information estimation Y_k in band b_l can be determined by evaluating the individual Laplacian parameters and their corresponding mean value as:

$$S_k = \{(m, n) | \alpha_k^{b_l}(m, n) < E(\alpha_k^{b_l})\} \quad (7)$$

where $\alpha_k^{b_l}(m, n)$ is the estimated Laplacian parameter of side information Y_k at position (m, n) in band b_l and E is the expectation operator. The overall unreliable region map is defined as a union of the sets S_k :

$$map = \bigcup_{k=1}^M S_k \quad (8)$$

The multi-hypothesis soft-input are calculated as:

$$Pr_j = \begin{cases} P(b_i | y_1, f_{x|y_1}, b^-), & \text{if } i \notin map \\ P(b_i | y_1..y_M, F_j, b^-), & \text{if } i \in map \end{cases} \quad (9)$$

where Pr_j is the j th candidate soft-input fed into LDPCA decoder, b_i denotes the i th bit in one bit plane, i is the one-dimensional presentation of the coordinate (m, n) , and $y_1..y_M$ denote different side information values in transform domain based on diverse side information generation schemes. Particularly, y_1 and $f_{x|y_1}$ denote the corresponding side information value for bit b_i and the estimated noise distribution based on the OBMC based frame interpolation scheme.

All the hypotheses of soft-input, $Pr_j, j \in [1, N]$, are fed into a modified LDPCA decoder. The first converging soft-input is chosen thus reducing the rate of LDPCA decoding. Subsequently, with the information of chosen soft-input, the corresponding joint weighted distribution $F_j, j \in [1, N]$, in the unreliable region map is determined. Given this information to a proposed joint weighted distribution based on reconstruction module, the minimum mean-square error reconstructed value, x' , in the unreliable region map is obtained as:

$$\begin{aligned} x' &= E[x | x \in [L, U], y_1..y_M] = \frac{\int_L^U x F_j(x) dx}{\int_L^U F_j(x) dx} \\ &= \frac{\sum_{k=1}^M \int_L^U x w_{jk} f_{x|y_k}(x) dx}{\sum_{k=1}^M \int_L^U w_{jk} f_{x|y_k}(x) dx} \end{aligned} \quad (10)$$

where $[L, U]$ are decoded quantization intervals. F_j denotes the most accurate joint weighted distribution available at decoder. w_{jk} (where $\sum_{k=1}^M w_{jk} = 1$) are predefined weighting parameters corresponding to F_j for side information k , $k \in [1, M]$. The reconstructed value outside the map region is calculated following the single side information reconstruction technique as in [11].

IV. EXPERIMENTAL RESULTS

In order to make a fair evaluation of the proposed Wyner-Ziv video coding, the test conditions adopted in this paper are the commonly used DISCOVER project [7] test conditions. The test sequences are *Foreman*, *Soccer*, *Coastguard* and *Hall* at 15 frames per second, QCIF, GOP size 2. Key frames are coded with H.264/AVC intra and QPs are chosen as in [7] so that the average quality of Wyner-Ziv frames is similar to the average quality of the key frames.

First of all, in order to evaluate the performance of the optical flow based frame interpolation scheme described in Section III-B, the quality of interpolated frames is measured by average Peak Signal-to-Noise Ratio (PSNR) over the set of test sequences and compared with block based frame interpolation [6] and extrapolation [11] techniques in Table I. It can be seen that the OBMC based frame interpolation method gives better performance overall. However, optical flow based frame interpolation outperforms OBMC on the high motion sequences, e.g. *Soccer*, and on the individual interpolated frames with high motion, as shown in Fig. 2. On the other hand, block based methods typically perform better in low motion sequences. Taking these diverse side information generation schemes as input, the proposed Wyner-Ziv video decoder is able to combine the different estimated side information adaptively.

The RD performance of the proposed TDWZ video coding is evaluated. Only the luminance component is coded, allowing for comparison with the DISCOVER codec [7] and

TABLE I
THE AVERAGE PSNR RESULTS FOR DIFFERENT SIDE INFORMATION GENERATION METHODS. KEY FRAMES ARE INTRA CODED WITH FIXED QP

	OBMC based Interpolation [7]	Optical Flow based Interpolation	Block based Extrapolation [12]
Foreman, QP=25	29.26	29.28	25.20
Soccer, QP=25	21.30	22.43	19.26
Coast, QP=26	31.83	30.92	28.55
Hall, QP=24	36.46	32.28	33.24

the best available single side information mode TDWZ codec [13]. The performances of benchmark codecs, H.264/AVC Intra and Inter no motion, are also included. The proposed Wyner-Ziv decoder is implemented in two different versions, by employing two frame interpolation schemes as shown in Fig. 1 (i.e. $M=2$) and three frame generation schemes (two frame interpolation shown in Fig. 1 plus the frame extrapolation technique employed in [12], i.e. $M=3$), respectively. In order to make the comparison fair, the number of candidate hypothesis is constrained to $N=6$ (also allowing for fair comparison with the similar complexity as in [11]). For the case $M=2$ and $N=6$, the weighting parameters used are, $w_{lj}=\{1;0.8;0.6;0.4;0.2;0\}$ and $w_{2j}=1-w_{lj}, j \in [1, 6]$. For the case $M=3$ and $N=6$, the weighting parameters are empirically predefined as: $w_{lj}=\{1;0;1/2;1/2;0;1/3\}$, $w_{2j}=\{0;1;1/2;0;1/2;1/3\}$, and $w_{3j}=\{0;0;0;1/2;1/2;1/3\}$.

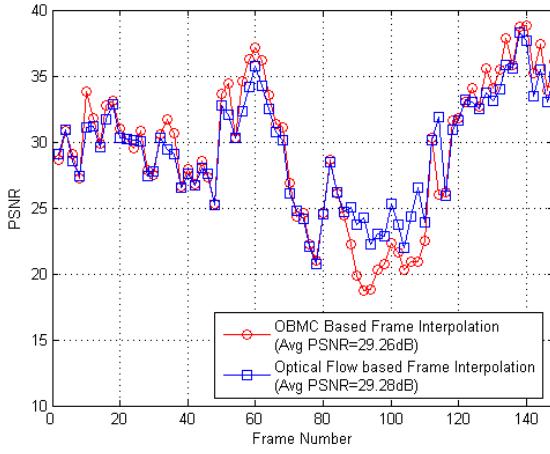


Fig. 2. PSNR for interpolated frames of Foreman, (Key frames QP=25) using OBMC and optical flow

Fig. 2 shows that the frame by frame interpolation performance may differ even though the overall performance is similar.

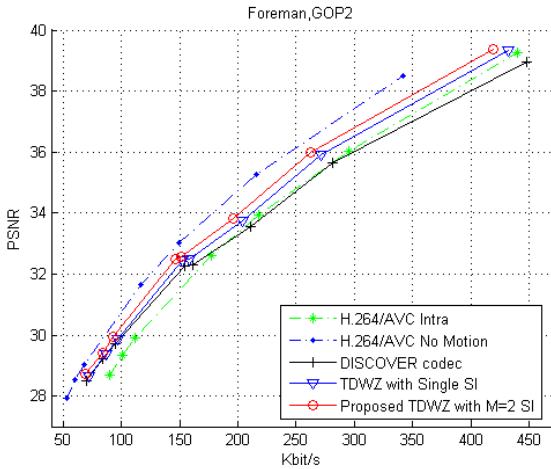


Fig. 3. Overall RD performance comparison for Foreman

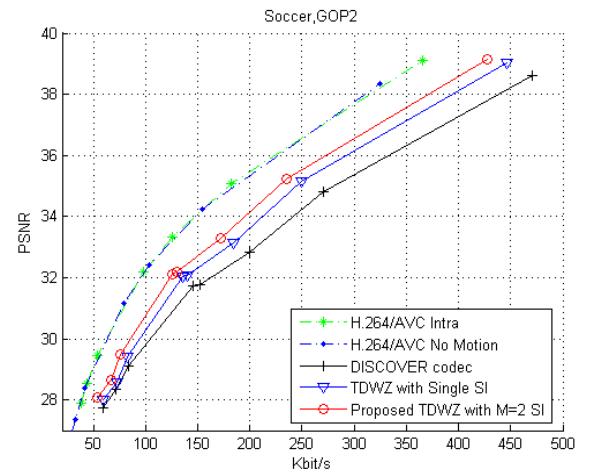


Fig. 4. Overall RD performance comparison for Soccer

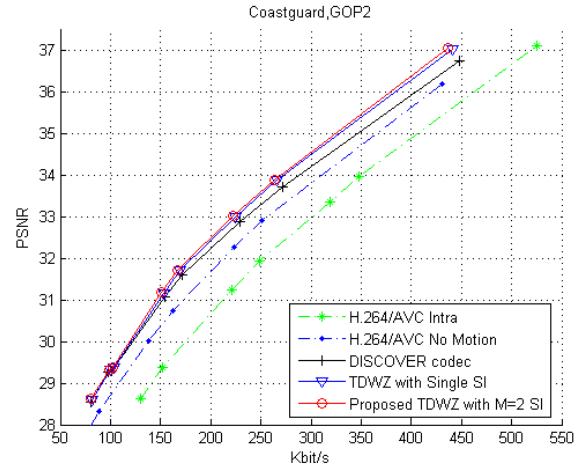


Fig. 5. Overall RD performance comparison for Coastguard

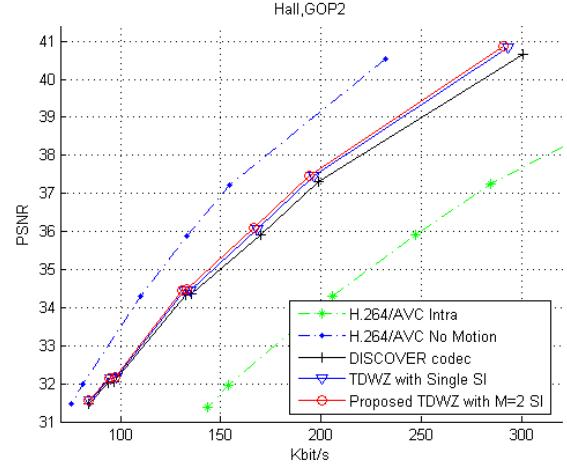


Fig. 6. Overall RD performance comparison for Hall

As shown in Figs. 3-6, the performance of the best available TDWZ with single side information [13] employed in this paper is significantly better than the DISCOVER video codec. With the proposed Wyner-Ziv video codec (with $M=2$ and $N=6$ mode), the overall RD performance of TDWZ can be improved by up to 0.6 dB at high bitrate for the sequence

Soccer. The accumulated improvement compared to the DISCOVER codec is up to 2 dB at high bitrate. Compared with H.264/AVC Intra coding, the proposed TDWZ codec gives a better RD performance for relative low motion sequences, *Foreman*, *Coastguard* and *Hall*. For the high motion sequence *Soccer*, the performance gap compared to H.264/AVC Intra coding has been substantially reduced but not eliminated yet. It is worth to note that, the proposed TDWZ significantly outperforms H.264/AVC Inter no motion coding for *Coastguard* but still is not competitive for the other test sequences.

RD improvements are also measured by average Bjøntegaard bitrate savings [21] over the DISCOVER codec and reported in Table II. It shows that the proposed Wyner-Ziv video coding scheme (either $M=2$ or $M=3$ mode) outperforms DISCOVER codec and the best available single side information mode TDWZ [13]. By adding one more frame extrapolation scheme ($M=3$ mode) in the proposed Wyner-Ziv decoder, the performance can be further improved. Compared to the related previous work in [11] with the same N , the proposed scheme is also better. The proposed Wyner-Ziv decoder provides larger gains for high motion sequences like *Foreman* and *Soccer* with average rate gain up to 44.5% for Wyner-Ziv frames (24.2% for overall performance) compared to DISCOVER codec. Although the OBMC based frame interpolation is quite efficient (see Table. 1) in single side information mode TDWZ for low motion sequences, e.g. *Hall*, the other side information generation methods can still contribute to RD performance improvement in the proposed multi-hypothesis Wyner-Ziv decoder.

TABLE II

BJØNTEGAARD AVERAGE BITRATE SAVINGS IN PERCENTAGE COMPARED WITH DISCOVER TDWZ VIDEO CODEC

	RD improvement measured in average bitrate savings (in %) over DISCOVER codec for WZ frames and for overall performance in parentheses			
	Best Available TDWZ with Single SI [13]	Related Previous Work in [11]	TDWZ with M=2 SI	TDWZ with M=3 SI
Foreman	13.3 (5.71)	20.3 (8.50)	27.0 (10.9)	31.4 (12.6)
Soccer	23.1 (12.8)	29.5 (16.5)	41.2 (22.6)	44.5 (24.2)
Coast	11.5 (3.91)	16.4 (5.22)	17.4 (5.51)	19.1 (5.96)
Hall	8.77 (2.72)	14.1 (3.88)	13.3 (3.78)	14.3 (3.94)

V. CONCLUSION

A novel multi-hypothesis TDWZ video coding including optical flow frame interpolation is proposed in this paper. The multiple side information is generated by both block based and optical flow based side information generation techniques. Multi-hypothesis soft-input is utilized for both decoding and reconstruction based on weighted joint distributions. In this way, the proposed scheme is not only able to reduce the required bitrate for decoding but also improve the quality of reconstructed Wyner-Ziv frame. Compared with the best available single side estimation mode TDWZ video coding, the overall RD performance can be improved up to 0.6 dB and up to 2 dB compared with DISCOVER TDWZ video codec.

ACKNOWLEDGMENT

The work presented is funded by the Danish Research Council (FTP Nr. 274-09-0249).

REFERENCES

- [1] 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*, pp. 339-352, 2008.
- [2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19 (4), pp. 471-480, Jul. 1973.
- [3] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Inform. Theory*, vol. 22 (1), pp. 1-10, Jan. 1976.
- [4] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform domain Wyner-Ziv codec for video," *Proc. SPIE VCIP*, San Jose, CA , USA, Jan. 2004.
- [5] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," *Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.
- [6] X. Huang and S. Forchhammer, "Improved side information generation for distributed video coding," *IEEE Int'l Workshop Multimedia Signal Proc.*, Cairns, Australia, Oct. 2008.
- [7] DISCOVER Project, www.discoverdvc.org, Dec. 2007.
- [8] X. Huang and S. Forchhammer, "Improved virtual channel noise model for transform domain Wyner-Ziv video coding," *IEEE Int'l Conf. on Acoustics, Speech, and Signal Proc.*, Taipei, Taiwan, ROC. April 2009.
- [9] R. Martins, C. Brites, J. Ascenso, F. Pereira, "Refining Side Information for Improved Transform Domain Wyner-Ziv Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19 (9), pp. 1327-1341, Sep. 2009.
- [10] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner-Ziv video coding with multiple side information," in *Proc. IEEE Int'l Workshop Multimedia Signal Process.*, pp. 183-186, Chania, Greece, Oct. 2007.
- [11] X. Huang, J. Ascenso, C. Brites, F. Pereira and S. Forchhammer, "Distributed video coding with multiple side information," *Picture Coding Symposium*, Chicago, USA, May 2009.
- [12] D. Varodayan, A. Aaron, B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Signal Processing*, vol. 23, pp. 3123-3130, 2006.
- [13] X. Huang and S. Forchhammer, "Cross-band noise model refinement for transform domain wyner-ziv video coding," *Signal Processing: Image Communication*, 2011, accepted.
- [14] S. Baker, D Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol 31, pp. 1-31, 2011.
- [15] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L¹ optical flow," *DAGM-Symposium*, 2007.
- [16] A. Wedel, C. Zach, T. Pock, H. Bischof, and D. Cremers, "An improved algorithm for TV-L¹ optical flow," *Dagstuhl Motion Workshop*, 2008.
- [17] L. Rakêt, L. Roholm, M. Nielsen and F. Lauze "TV-L¹ optical flow for vector valued images," *In Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2011 (to appear).
- [18] T. Brox, A. Bruhn, N. Papenberg and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," *Proceedings of ECCV*, vol. 4, pp. 25-36, 2004.
- [19] L. Xu, J. Jia, Y. Matsushita, "A unified framework for large- and small-displacement optical flow estimation" Technical report, *The Chinese University of Hong Kong*, 2010.
- [20] S. Keller, F. Lauze, and M. Nielsen, "Temporal super resolution using variational methods." *High-Quality Visual Experience: Creation, Processing and Interactivity of High-Resolution and High-Dimensional Video Signals*, eds. M. Mrak, M. Grgic, and M. Kunt, 2010.
- [21] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *VCEG Contribution VCEG-M33*, April 2001.