

Technical University of Denmark



## Improved side information generation for distributed video coding

**Huang, Xin; Forchhammer, Søren**

*Published in:*

IEEE Int'l Workshop om Multimedia Signal Processing

*Link to article, DOI:*

[10.1109/MMSP.2008.4665079](https://doi.org/10.1109/MMSP.2008.4665079)

*Publication date:*

2008

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Huang, X., & Forchhammer, S. (2008). Improved side information generation for distributed video coding. In IEEE Int'l Workshop om Multimedia Signal Processing (pp. 223-228). IEEE. DOI: 10.1109/MMSP.2008.4665079

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Improved Side Information Generation For Distributed Video Coding

Xin Huang<sup>1</sup> and Søren Forchhammer<sup>2</sup>

*DTU Fotonik, Technical University of Denmark*

*Building 343, Lyngby 2800, Denmark*

<sup>1</sup>xin@com.dtu.dk

<sup>2</sup>sf@com.dtu.dk

**Abstract**—As a new coding paradigm, Distributed Video Coding (DVC) deals with lossy source coding using side information to exploit the statistics at the decoder to reduce computational demands at the encoder. The performance of DVC highly depends on the quality of side information. With a better side information generation method, fewer bits will be requested from the encoder and more reliable decoded frames will be obtained. In this paper, a side information generation method is introduced to further improve the Rate-Distortion (RD) performance of transform domain distributed video coding. This algorithm consists of a variable block size based Y, U and V component motion estimation and an adaptive weighted Overlapped Block Motion Compensation (OBMC). The proposal is tested and compared with the results of an executable DVC codec released by DISCOVER group (DIStributed COding for Video sERvices). RD improvements on the set of test sequences are observed.

## I. INTRODUCTION

Distributed Video Coding (DVC) has been proposed in [1] to avoid using complex motion estimation and motion compensation at the encoder and only explore the video statistics at the decoder side. In many emerging applications, e.g. wireless video surveillance, wireless PC cameras and mobile cameras, due to limited memory and computational power at the encoder side, DVC might be more suitable than conventional video coding like ISO MPEG-x and ITU-T H.26x which have one highly complex encoder and (one or) many simpler decoders. DVC is based on two major information theoretic results: the Slepian-Wolf [2] and Wyner-Ziv [3] theorems. According to the Slepian-Wolf theorem, it is possible to achieve the same rate as a joint encoding system by independent encoding but joint decoding of two statistically dependent signals. The Wyner-Ziv theorem extends the Slepian-Wolf theorem to a lossy case, which becomes the key theoretical basis of DVC.

A Low-Density Parity-Check (LDPC) based transform domain Wyner-Ziv codec released by DISCOVER [4][5] is one of the best DVC codecs available. It improves on the work based in [6] by introducing an advanced frame interpolation for side information generation [7][8], a finer correlation noise modeling [9][10], and an optimal reconstruction algorithm [11]. However, there are still significant RD performance gaps between DVC and conventional video coding schemes as H.264/AVC. Since the quality of side information frames is a very natural element influencing the coding efficiency of DVC,

there are several different side information generation schemes in the literature including interpolation [7][8] and extrapolation [12][13] based algorithms. In this paper, an interpolation based side information generation scheme is introduced and applied to a transform domain DVC to improve the RD performance. This new scheme improves on the work in [7] and [8], by introducing Y, U and V based motion estimation with variable block size to take advantage of more information and obtain more accurate motion vectors first, combined with an adaptive weighted Overlapped Block Motion Compensation (OBMC) to generate better side information.

The rest of this paper is organized as follows: Section II briefly describes the architecture of the LDPC based transform domain Wyner-Ziv video coding. In Section III, the proposed side information generation scheme is introduced. Test conditions and results are presented in Section IV.

## II. ARCHITECTURE OF DVC

The architecture of a state-of-art DVC codec [5][6] is depicted in Fig. 1. A fixed Group of Pictures (GOP=2) is adopted. The video sequence is first split into odd (key) frames and even (Wyner-Ziv) frames. The odd frames are intra coded by using a conventional video coding like H.264/AVC while the even frames are Wyner-Ziv coded.

In the Wyner-Ziv encoder, Wyner-Ziv frames are partitioned into non-overlapped 4x4 blocks and an integer discrete cosine transform (DCT) [14] is applied on each of these. The transform coefficients within a given band  $b_k, k \in \{0 \dots 15\}$ , are grouped together and then quantized with  $2^{M_k}$  levels. DC coefficients and AC coefficients are uniformly scalar quantized and dead zone quantized, respectively. After quantization, the coefficients are binarized, each bitplane is transmitted to a rate-compatible LDPC accumulate encoder [15] starting from the most significant bitplane. For each encoded bitplane, the corresponding accumulated syndrome is stored in a buffer together with an 8-bit Cyclic Redundancy Check (CRC). The amount of bits to be transmitted depends on the requests from the decoder through a feedback channel.

In the Wyner-Ziv decoder, based on two intra coded frames  $X_{2i-1}$  and  $X_{2i+1}$ , a motion estimation and compensation based frame interpolation algorithm is adopted to create a side information frame  $Y_{2i}$  and a motion estimated residual frame  $R_{ME}$  (i.e. the difference between  $X_{2i-1}$  and  $X_{2i+1}$  along

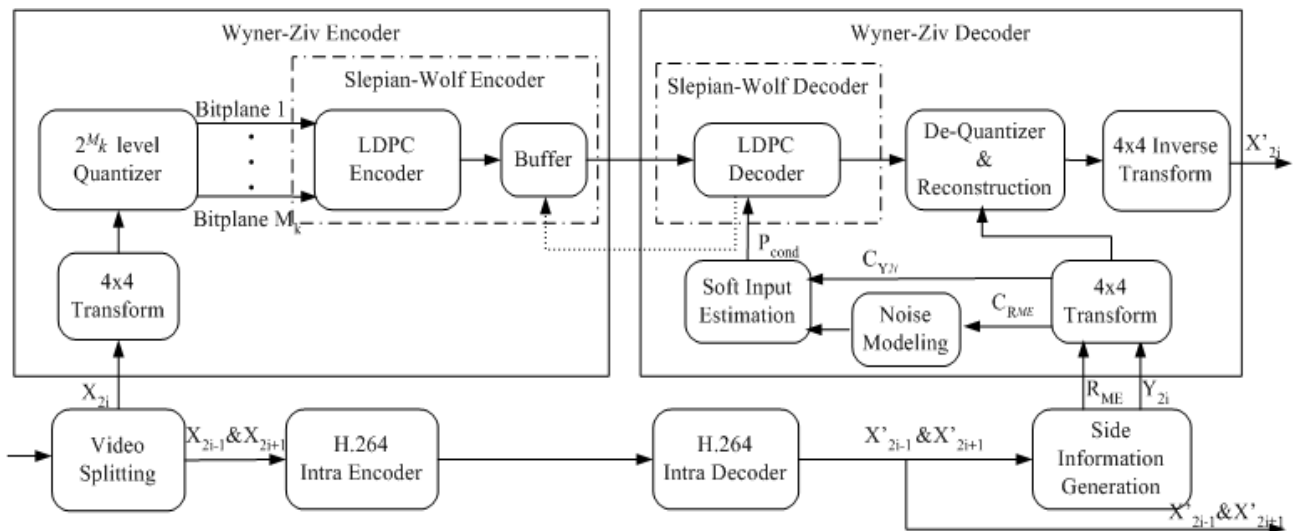


Fig. 1. Diagram of LDPC based transform domain Wyner-Ziv codec architecture

the motion vectors).  $Y_{2i}$  and  $R_{ME}$  undergo the same 4x4 integer DCT to obtain coefficients  $C_{Y_{2i}}$  and  $C_{R_{ME}}$ .  $C_{R_{ME}}$  is utilized on-line to roughly model [9] the noise distribution between corresponding DCT bands of side information frame and Wyner-Ziv frame (i.e.  $C_{Y_{2i}}$  and  $C_{X_{2i}}$ ). By using the obtained noise distribution, coefficient values of the side information frame  $C_{Y_{2i}}$  and previous successfully decoded bitplanes, soft information (conditional bit probabilities  $P_{cond}$ ) for each bitplane is estimated. With a given soft-input information  $P_{cond}$ , the LDPC decoder starts to process the corresponding bitplanes to correct the bit errors. Convergence is tested by computing the Hamming distance between the received syndrome and the one obtained by the decoded bitplane. If the Hamming distance is different from zero after a certain amount of iterations, the LDPC decoder requests more accumulated syndrome bits from the encoder buffer via the feedback channel. If the Hamming distance is equal to zero, then the 8-bit CRC sum is requested from the buffer to verify successful decoding. A decoded bitplane with correct CRC sum is sent to an optimal reconstruction module [11], a bitplane with incorrect CRC sum requests more accumulated syndrome bits from encoder buffer to correct the existing bit errors until a low error probability is guaranteed. For more details refer to [5][15].

### III. SIDE INFORMATION GENERATION

Based on the architecture of the DVC, the output of side information generation not only influences the soft input estimation module but also the reconstruction module. Therefore, the choice of the adopted side information generation scheme can significantly influence the RD performance. Generally, more accurate side information frames means that fewer bits are requested from the encoder for the same decoding quality. An advanced motion compensated interpolation algorithm [7] is reportedly adopted in the executable DVC codec [4][5] released by DISCOVER. It includes forward motion estimation,

bi-directional motion estimation, spatial smoothing of motion vectors and bi-directional motion compensation. The work has been improved to extend motion estimation and compensation to sub-pixel accuracy [8].

Although this scheme can generate good side information frames, there are some limitations: First of all, it does not utilize all the information available at the decoder side. Secondly, the block size used for motion estimation and compensation might not be an optimal choice. Thirdly, only a simple bi-directional motion compensation is employed. Overcoming these limitations will improve the side information generation and further improve RD performance of DVC. Therefore, an improved side information generation scheme is proposed as shown in Fig. 2. It is divided into two parts: Y, U and V based motion estimation with variable block size is applied to get accurate motion vectors at first. Then an adaptive weighted Overlapped Block Motion Compensation (OBMC) is employed to generate better side information frames.

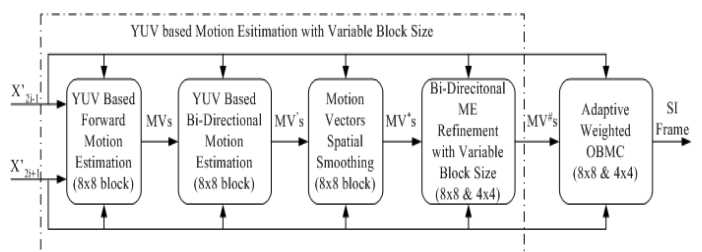


Fig. 2. Improved side information generation scheme

#### A. YUV based motion estimation with variable block size

In order to take advantage of more information available at the decoder, the chroma components (U and V) in intra decoded key frames are utilized to assist luma component (Y) based motion estimation. Thus the Mean Squared Error (MSE)

based motion estimation is determined by:

$$\begin{aligned} \operatorname{argmin} \{ & \xi_{(m,n) \in \text{block}} \{ (X_{2i-1}^Y(m,n) - \\ & X_{2i+1}^Y(m + \Delta m, n + \Delta n))^2 \} \\ & + \lambda \cdot \xi_{(m',n') \in \text{block}} \{ (X_{2i-1}^{UV}(m',n') \\ & - X_{2i+1}^{UV}(m' + \Delta m', n' + \Delta n'))^2 \} \} \end{aligned} \quad (1)$$

where  $X_{2i-1}^Y(m,n)$  and  $X_{2i-1}^{UV}(m',n')$  are the corresponding luma and chroma values at coordinates  $(m,n)$  and  $(m',n')$  in key frame  $X_{2i-1}$ , respectively.  $(\Delta m, \Delta n)$  and  $(\Delta m', \Delta n')$  represent the motion vectors. For 4:2:0 video sequences,  $\Delta m = 2\Delta m', \Delta n = 2\Delta n', m = 2m'$  and  $n = 2n'$ .  $\lambda$  is a parameter to balance the weight between luma and chroma values.

Besides YUV based motion estimation, the first three modules in Fig. 2 are similar to [7][8]. With the given two decoded key frames  $X_{2i-1}$  and  $X_{2i+1}$ , an  $8 \times 8$  block based motion estimation is applied with full-pixel accuracy first. Since the rigid block based motion estimation results in overlapped and uncovered areas after the frame interpolation, the obtained motion vectors are only seen as candidates. Motion vectors,  $MV$ , are selected from the candidates that intercepts the interpolated frame closest to the center of each  $8 \times 8$  block. In order to obtain more accurate motion vectors  $MV's$ , a bidirectional motion estimation scheme [7] with sub-pixel accuracy is applied with a smaller search range. This bidirectional motion estimation selects a linear trajectory by using  $MV's$  as initial values, then the refined  $MV's$  are obtained by a bidirectional symmetric motion search. Afterwards,  $MV's$  are smoothed by using a weighted vector median filter [7]. A six tap Wiener filter [14] is used to interpolate key frames and consequently motion estimate in sub-pixel accuracy.

Since an  $8 \times 8$  block based motion estimation is applied in [5][7][8], it may not perfectly match the true motion especially around object boundaries. Variable size block based motion estimation is more efficient in representing irregular motion. Therefore, a bi-directional motion estimation with variable block size ( $8 \times 8$  and  $4 \times 4$ ) is adopted after the motion vector smoothing module. Selecting two predefined thresholds  $\tau_{mse}$  and  $\tau_\sigma$ , each  $8 \times 8$  block is evaluated to decide whether to divide it into  $4 \times 4$  sub-blocks based on:

$$MAP_{4 \times 4} = \begin{cases} True & \text{if } MSE_{8 \times 8} \geq \tau_{mse} \\ & \text{and } Var(MV) \geq \tau_\sigma \\ False & \text{otherwise} \end{cases} \quad (2)$$

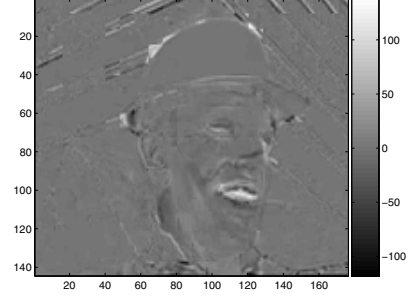
where  $MSE_{8 \times 8}$  is the YUV based MSE value between  $X_{2i-1}$  and  $X_{2i+1}$  over the corresponding  $8 \times 8$  block,  $Var(MV)$  is a function to calculate the variance of the relevant motion vectors for the current block in an  $3 \times 3$  window.

$$Var(MV(m,n)) = \frac{\sum_{i=-1}^1 (MV(m+i, n+i) - \bar{MV})^2}{9} \quad (3)$$

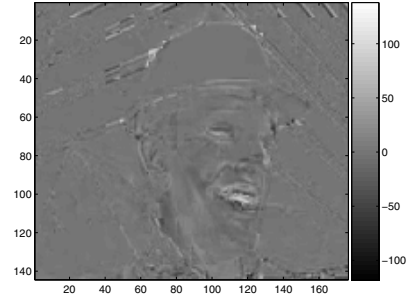
where  $\bar{MV}$  is the mean value of  $MV's$ . If an  $8 \times 8$  block satisfies the above conditions, its  $MV$  is taken as initial  $MV$  of each  $4 \times 4$  sub-blocks and the relevant  $MSE_{4 \times 4}$  are

calculated. A small refinement search range  $\rho$  is chosen to find the best matching  $4 \times 4$  sub-block with minimum  $MSE_{4 \times 4}$ .

With variable block size, the smaller blocks are used to describe irregular motion around the edges of objects, the larger blocks are used for homogeneous motion. As shown in Fig. 3, the energy of the motion estimated residual  $R_{ME}$  with variable block size is smaller than the one with a fixed  $8 \times 8$  block. Thus providing an advantage by introducing fewer inaccurate pixels into the side information frame.



(a) With  $8 \times 8$



(b) With  $8 \times 8$  and  $4 \times 4$

Fig. 3. Comparison of motion estimated residuals  $R_{ME}$  with fixed block size ( $8 \times 8$ ) and adaptive block size ( $8 \times 8$  and  $4 \times 4$ )

## B. Adaptive Weighted Overlapped Block Motion Compensation

Overlapped Block Motion Compensation (OBMC) is usually applied to reduce blocking artifacts and improve subjective quality in frame rate up-conversion. However, it also has a higher risk of over-blurring the interpolated side information frame compared with the simple bi-directional motion compensation used in [7]. Since the MSE value over each block of the YUV based motion estimation approximately reflects the reliability of its relevant motion vectors, an adaptive OBMC [16] weighted by MSE is employed to reduce the interpolated errors and control the blurring. Let  $j \in [0, k]$  denote the index of the neighboring blocks. As shown in Fig. 4, the value of  $k$  is varying due to variable block size adopted.

$$Y_{2i}(m_0, n_0) = \frac{\sum_{j=0}^k \omega_j \hat{Y}_j}{\sum_{j=0}^k \omega_j} \quad (4)$$

$$\hat{Y}_j = \frac{1}{2} \times (X_{2i-1}(m_0 + \Delta m_j^*, n_0 + \Delta n_j^*) + X_{2i+1}(m_0 + \Delta m_j^\#, n_0 + \Delta n_j^\#)) \quad (5)$$

$$R_{ME}(m_0, n_0) = \frac{\sum_{j=0}^k \omega_j \hat{R}_j}{\sum_{j=0}^k \omega_j} \quad (6)$$

$$\hat{R}_j = (X_{2i-1}(m_0 + \Delta m_j^*, n_0 + \Delta n_j^*) - X_{2i+1}(m_0 + \Delta m_j^\#, n_0 + \Delta n_j^\#)) \quad (7)$$

where  $(m_0, n_0)$  belongs to current block,  $(\Delta m_j^*, \Delta n_j^*)$  and  $(\Delta m_j^\#, \Delta n_j^\#)$  are backward and forward motion vectors of  $Block_j$  in  $X_{2i-1}$  and  $X_{2i+1}$ , respectively. The relation,  $(\Delta m_j^*, \Delta n_j^*) = -(\Delta m_j^\#, \Delta n_j^\#)$ , i.e. linear motion (with  $GOP=2$ ) is assumed.  $\omega_j$  is the weight of  $Block_j$  obtained by calculating the inverse proportion of the YUV based MSE:

$$\omega_j = (\xi_{(m_j, n_j) \in Block_j} ((X_{2i-1}^{YUV}(m_j + \Delta m_j^*, n_j + \Delta n_j^*) - X_{2i+1}^{YUV}(m_j + \Delta m_j^\#, n_j + \Delta n_j^\#))^2))^{-1} \quad (8)$$

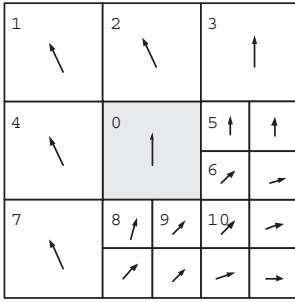


Fig. 4. Utilized neighboring motion vectors and blocks for adaptive weighted OBMC

#### IV. EXPERIMENTAL RESULTS

First of all, in order to evaluate the proposed Side Information Generation (SIG) scheme, different methods were implemented and these are compared in Table I by measuring the average Peak Signal-to-Noise Ratio (PSNR) of the interpolated frames over the whole sequence: i) bidirectional motion search based SIG employed in [13]; ii) advanced SIG employed in [5]; iii) YUV based SIG; iv) YUV based SIG with Variable Block Size; v) YUV based SIG with fixed block size ( $8 \times 8$ ) and adaptive weighted OBMC; vi) Y based SIG with VBS and adaptive weighted OBMC; vii) YUV based SIG with VBS and adaptive weighted OBMC.

Then the RD performances of DVC implementation with the proposed SIG method (vii) and the SIG method (ii) based on [7][8] as in [5] are compared. For testing and comparison, a DVC codec was implemented in MATLAB 7 combined with C. The performance of the basic version was brought in-line with the executable DVC codec[4] (see Figs. 6-9) for the comparison. The conditions for the tests are:

- All (149) frames of "Foreman", "Coastguard", "Soccer" and "Hallmonitor" sequences are used. The sequences are in QCIF@15Hz format, and they are available at [4].

TABLE I  
THE AVERAGE PSNR RESULTS FOR DIFFERENT METHODS, KEY FRAMES ARE INTRA CODED WITH H.264/AVC WITH FIXED QUANTIZATION PARAMETER (QPs).

Sequence	Foreman	Coastguard	Soccer	Hallmonitor
Key frames	QP=25	QP=26	QP=25	QP=24
i)	27.8192	29.7681	20.6988	35.0267
ii)	28.9047	31.4664	20.8326	36.3338
iii)	28.9843	31.4681	20.8483	36.3339
iv)	28.9999	31.5371	20.8453	36.3735
v)	29.2358	31.7708	21.2874	36.3331
vi)	29.2296	31.8317	21.2961	36.4548
vii)	29.2537	31.8340	21.2967	36.4593

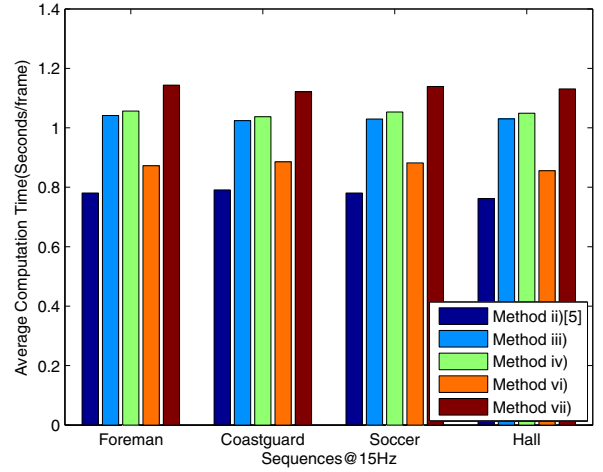


Fig. 5. Complexity comparison for different SIG methods

- As in [4][5], half-pixel accuracy motion estimation is used in the proposed side information generation for fair comparison.
- The most common GOP length ( $GOP=2$ ) in [5][6] is used. The key frames are encoded by H.264/AVC intra and the QP are chosen so that the average PSNR of Wyner-Ziv frames is similar to the quality of key frames as in [5]. The chosen QPs in Table I are corresponding to the eighth RD point.
- All the RD performance results are evaluated by the average of luminance components (Y) of key frames and Wyner-Ziv frames.
- Parameter  $\lambda$  in YUV based ME is chosen to be 5. The thresholds  $\tau_{mse} = \gamma \times Mean(MSE_{8 \times 8})$  and  $\tau_{\sigma} = 0$  are chosen for variable block size partition,  $\gamma = 6$  is chosen experimentally. The refinement search range  $\rho$  is defined in  $\pm 1$  pixels.

As shown in Table I, each module of the proposed side information scheme generally improves the PSNR step by step. The proposed method (vii) gives the best PSNR result by increasing the complexity of decoder.

The complexity for different SIG methods are evaluated by calculating the average time (on a 3GHz PC) for generating one interpolated frame. As shown in Fig. 5, if the complexity

of the decoder should not be significantly increased, the proposed method (vi), which removes YUV based motion estimation from method (vii), gives a good balance between decoder complexity and PSNR performance. We choose method (vii) with the best PSNR results. According to RD results shown in Figs. 6-9, the performance of DVC implementation (with side information method (ii) as in [5][7][8]) is comparable with the results of the DISCOVER executable codec. Compared with the SIG used in [5], the RD performances of DVC for high bit-rates are improved up to 0.5dB with the proposed SIG scheme (vii).

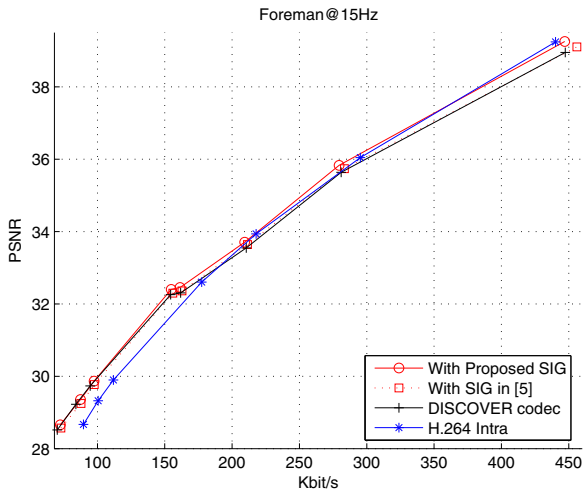


Fig. 6. RD comparison for sequence Foreman

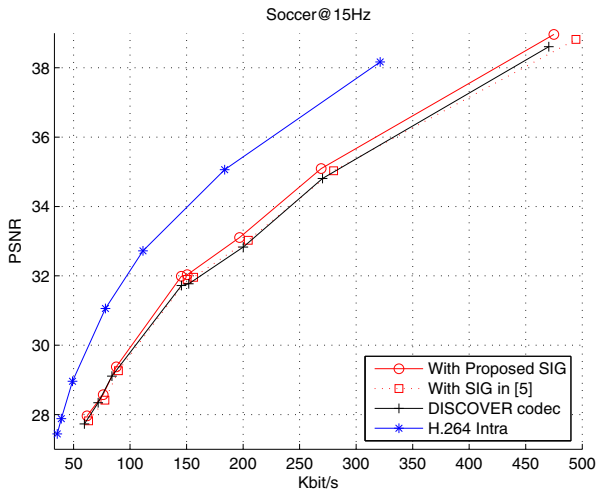


Fig. 7. RD comparison for sequence Soccer

## V. CONCLUSION

In this paper, an improved side information generation scheme is introduced in DVC. It overcomes three limitations of the current scheme utilized in the DISCOVER DVC executable codec. This is obtained by using more information (chroma)

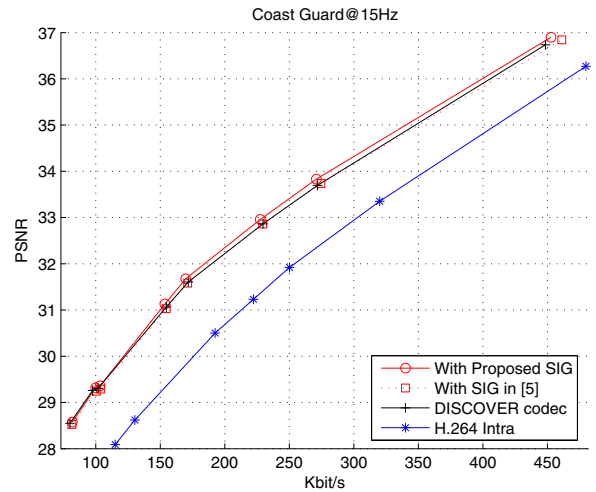


Fig. 8. RD comparison for sequence Coastguard

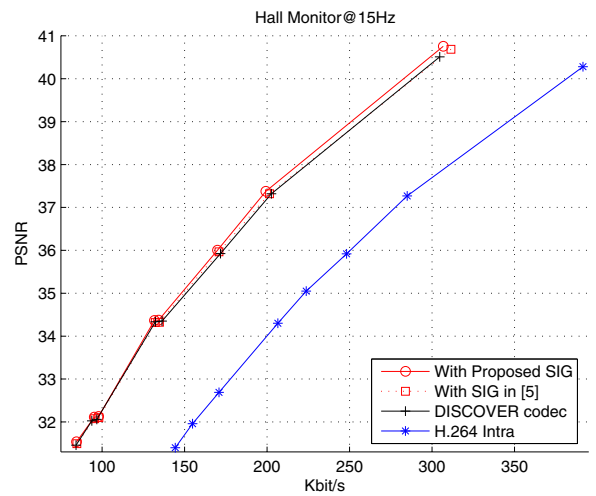


Fig. 9. RD comparison for sequence Hall Monitor

at the decoder side, utilizing variable block size for motion estimation and compensation and employing adaptive weighted OBMC. Experimental results show that the proposed scheme can improve coding efficiency of DVC. Compared with the current scheme employed in the DISCOVER executable codec, the proposed scheme improves the RD performance up to 0.5dB at the higher bit-rates.

## REFERENCES

- [1] A. Aaron, R. Zhang, and B. Girod, "Wyner-ziv coding of motion video," *Proc. Asilomar Conference on Signals and Systems*, Nov. 2002.
- [2] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, pp. 471–480, July 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, pp. 1–10, Jan. 1976.
- [4] Available on: [www.discoverdvc.org](http://www.discoverdvc.org).
- [5] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Knbasov, and M. Ouaret, "The discover codec: architecture, techniques and evaluation," *Picture Coding Symposium (PCS)*, Nov. 2007.

- [6] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform domain wyner-ziv codec for video," *Proc. SPIE VCIP*, Jan. 2004.
- [7] 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*, July 2005.
- [8] S. Klomp, Y. Vatis, and J. Ostermann, "Side information interpolation with sub-pel motion compensation for wyner-ziv decoder," *International Conference on Signal Processing and Multimedia Applications*, Aug. 2006.
- [9] C. Brites, J. Ascenso, and F. Pereira, "Studing temporal correlation noise modeling for pixel based wyner-ziv video coding," *IEEE Int'l Conf. on Image Proce.*, Oct. 2006.
- [10] C. Brites, J. Ascenso, J. Pedro, and F. Pereira, "Evaluating a feedback channel based transform domain wyner-ziv video codec," *Signal Processing, Image Communication* 23, pp. 269–297, 2008.
- [11] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in wyner-ziv video coding with multiple side information," *International Workshop on Multimedia Signal Processing*, Oct. 2007.
- [12] L. Lu, D. He, and A. Jagmohan, "Side information generation for distributed video coding," *IEEE Int'l Conf. on Image Proce.*, Oct. 2007.
- [13] Z. Li, L. Liu, and E. J. Delp, "Rate distortion analysis of motion side estimation in wyner-ziv video coding," *IEEE Trans. on Image processing*, vol. 16, pp. 98–113, Jan. 2007.
- [14] I. Richardson, *H.264 and MPEG-4 Video Compression*, John Wiley and Sons Ltd., West Sussex, England, 2003.
- [15] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, pp. 3123–3130, Nov. 2006.
- [16] S. Lee, O. Kwon, and R. Park, "Weighted-adaptive motion-compensated frame rate-up conversion," *IEEE Trans. Consum. Electron.*, vol. 49, pp. 485–492, Aug. 2003.