Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **2**, March **2017**

Bayesian Multi-Hypothesis Wyner-Ziv Video Coding

Lili Meng, Jingxiu Zong, Guina Sun, Jie Cheng, Jia Zhang and Mengchen Zhao

Department of Information Science and Engineering
Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology
Shandong Normal University
No.88 Wenhua East Road, Jinan, 250014, China
mengll_83@hotmail.com

Received August 2016; revised December, 2016

Abstract. *The performance of distributed video coding (DVC) relies heavily on the quality of the side information (SI), and better performance can be expected if multiple SIs are employed. In this paper, we consider the scenario with two SIs, which are obtained by forward and backward predictions respectively. In this case, a couple of ad hoc approaches have been proposed to approximate the conditional probability density function (pdf) of a Wyner-Ziv-coded frame given the two SIs. However, the optimal conditional pdf and the relative performances of these ad hoc approaches with respect to the optimal solution have not been studied. In this paper, using Bayes' formula, we derive the closed-form expression of the conditional pdf. Experimental results show that the Bayesian solution has similar performance to previous methods. Moreover, simulation results reveal that when better SIs are available, the Bayesian solution could outperform other methods. This analysis sheds some lights on developing future DVC schemes.*
Keywords: Wyner-Ziv video coding, multi-hypothesis, correlation noise model, Bayesian theory.

1. **Introduction.** State-of-the-art video coding schemes such as MPEG-2 and H.264/AVC are based on motion-compensated transform coding. The compression is mainly achieved by motion compensation at the encoder, which has high complexity, whereas the complexity of the decoder is much lower. Therefore, these conventional codecs are suitable for applications where the video is only compressed once but decoding is performed several times, such as broadcasting. However, they are not ideal if low complexity is desired at the encoder, such as distributed sensor networks.

Distributed Source Coding (DSC) represents a major paradigm shift, where the complexity is shifted from the encoder to the decoder. It is based on the theoretical foundation laid out by Slepian-Wolf (SW) theory [1] and Wyner-Ziv (WZ) theory [2]. SW coding is a lossless compression technique, which states that when two or more correlated sources are seperately encoded and jointly decoded, the same compression ratio as joint encoding can be achieved. A special case of SW coding is when one of the sources is only available at the decoder as a side information (SI). The WZ coding is an extension of this case to lossy compression. It can be viewed as SW coding followed by a quantizer.

The DSC has drawn great attention in the video coding community [3, 4, 5]. In many practical distributed video coding (DVC) schemes, such as the DISCOVER codec [3, 6], a video sequence is divided into two parts: key frames and WZ frames. The key frames are encoded and decoded by conventional intra video coding methods. At the encoder,

the WZ frames are intra-encoded without using motion estimation, and parity bits are produced as the compressed bitstream. At the decoder, the WZ frames are inter-decoded by using both the parity bits and the SI produced from the decoded key frames. Therefore, the model of statistical correlation between WZ frame and SI is needed at the decoder. The model between WZ frame and SI can be expressed by $Y = X + Z$, where Y is the SI, X denotes the WZ frame and Z is the correlation noise. The correlation noise is usually assumed to have Laplacian distribution. As a result, the quality of the SI and the estimation of correlation noise determine the decoder efficiency.

The optimal reconstruction method using one SI is developed in [7]. In [8], it is suggested from the information theory perspective that the DSC coding efficiency can be improved by using multiple SIs, because they can reduce the conditional entropy of the source. In the typical setting of the DVC, two SIs for each WZ frame can be readily obtained from the neighboring key frames, using forward and backward motion estimation respectively. To take full advantage of multiple SIs, the conditional pdf given all the SIs is needed, whose closed-form expression is still unknown. In the past, some ad hoc approaches have been proposed to approximate the conditional pdf. In [7], the simple average of two individual conditional pdfs is used. In [9], a weighted average is proposed, and the weight is determined by the quality of the SI.

This paper sheds more lights on the conditional pdf and its impact to the performance of the system. First, using Bayes' formula, we show that the closed-form expression of the conditional pdf can be obtained. Experimental results show that the Bayesian solution has comparable performance to the method in [9]. Moreover, simulation results reveal that when better SIs are available, the Bayesian solution could potentially outperform other methods.

The outline of the paper is as follows. Sec. 2 overviews our proposed DVC framework. Sec. 3 derives the Bayesian conditional pdf. Experiment and simulation results are given in Sec. 4, and Sec. 5 summarizes the paper.

2. **Overview of the Proposed System.**

Our DVC architecture is similar to the DISCOVER system [3] and is depicted in Fig. 1. The video frames are first split into key frames (I frames) and Wyner-Ziv frames (WZ frames). In this paper, we only consider the case with one WZ frame between two key frames.

The key frames are encoded and decoded by H.264 Intra codec, whereas the WZ frames are encoded independently and decoded by WZ decoder, with the help of the SIs generated from the decoded key frames.

The WZ frame is first transformed by 4×4 DCT, followed by scalar quantization (SQ). The quantization matrix in [3] is adopted. After that, the bit planes of the quantized coefficients are transmitted to the Low-Density-Parity-Check-Accumulate (LDPCA) encoder [10], which computes the parity bits and transmits them to the decoder adaptively according to the request of the LDPCA decoder.

At the decoder, after decoding the key frames, the two SIs required by the decoding of WZ frames are obtained from forward prediction (SI_f) and backward prediction (SI_b), using the decoded key frames. The generation of the two SIs is illustrated in Fig. 2. The two SIs are then transformed by DCT, and the conditional pdf $Pr(x|y_1, y_2)$ is obtained, which will be used in the final reconstruction, as will be described in Sec. 3.

To improve the accuracy of the correlation noise model, the classification method in [11] is used, where all the coefficients in each DCT subband in a frame is divided into several groups, based on the residual energy between the two SIs. Each group uses its own Laplacian parameter.

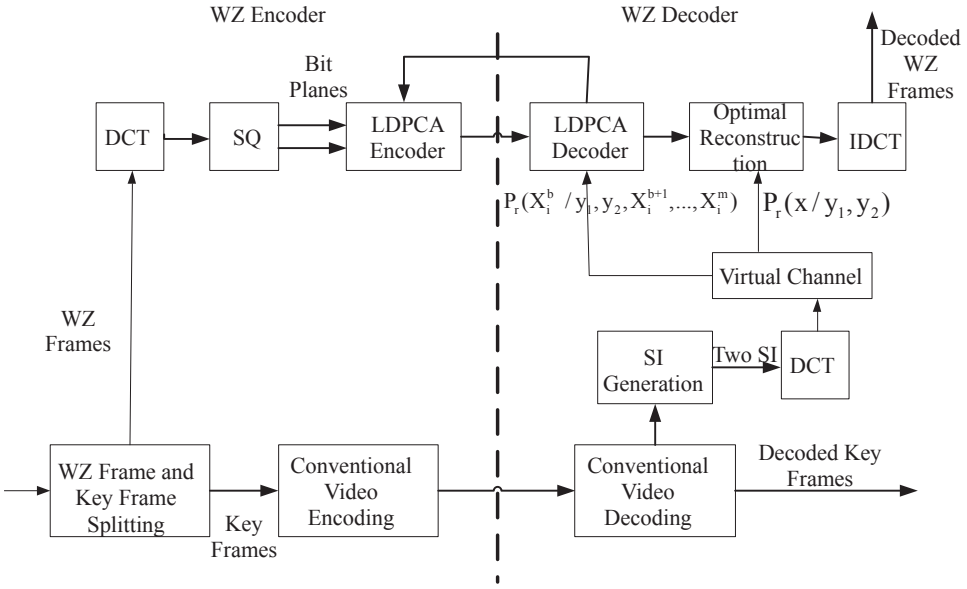

FIGURE 1. Proposed DVC Scheme.

The LDPCA decoder uses the belief propagation algorithm to decode the WZ frame from the received parity bits, the already decoded higher bit-planes and the conditional pdf. At the beginning, only a small portion of the parity bits is transmitted to the LDPCA decoder. If the decoding fails, more parity bits will be requested by the feedback channel until the decoded result is satisfactory.

After the LDPCA decoder, we know the quantized DCT coefficients of the WZ frames, which specify the ranges of the unquantized coefficients. A final reconstruction step is then applied to find the best estimate of the unquantized coefficients in this range, given the knowledge of the two SIs and the corresponding correlation noise parameters. This will be explained in Sec. 3. Finally, the decoded WZ frames are obtained by IDCT.

3. **The Bayesian Conditional pdf and the correlation model.**

3.1. **Bayesian conditional pdf.** For each DCT coefficient x in the WZ frame, after the LDPCA decoder, we know that its value is in the range of $[z_i, z_{i+1}]$. With the additional knowledge of the two SIs y_1 and y_2, the optimal reconstruction of x is proposed by [7]. The optimal reconstruction of x using minimum mean-squared error (MMSE) is expressed as follows.

$$\hat{x}_{opt} = E[x|x \in [z_i, z_{i+1}], y_1, y_2] = \frac{\int_{z_i}^{z_{i+1}} x f_{X|y_1,y_2}(x)dx}{\int_{z_i}^{z_{i+1}} f_{X|y_1,y_2}(x)dx}. \tag{1}$$

As a result, we need to find the expression of the conditional pdf $f(X|y_1, y_2)$. Some ad hoc solutions have been proposed to approximate it. In [7], it is simply assumed that

$$f_{X|y_1,y_2}(x) \approx \frac{1}{2}(f_{X|y_1}(x) + f_{X|y_2}(x)), \tag{2}$$

where $f_{X|y_i}(x)$ is given by the following Laplacian model.

$$f_{X|y_i}(x) = \frac{\alpha_i}{2} e^{-\alpha_i|x-y_i|}, \quad i = 1, 2. \tag{3}$$

The Laplacian parameter α_i can be estimated from the variance σ_i^2 of the side information Y_i by $\alpha_i^2 = 2/\sigma_i^2$.

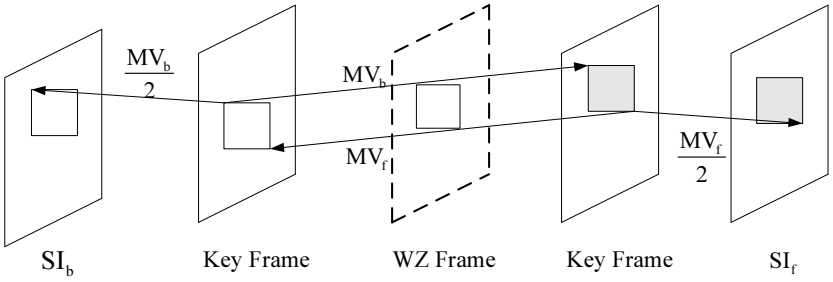

FIGURE 2. The generation of multiple SIs using forward and backward predictions.

In [9], the following improved weighted average is proposed.

$$f_{X|y_1,y_2}(x) \approx w_1 f_{X|y_1}(x) + w_2 f_{X|y_2}(x), \tag{4}$$

where

$$w_i = \frac{\alpha_i^2}{\alpha_1^2 + \alpha_2^2}. \tag{5}$$

The motivation is to assign more weights to $f_{X|y_i}(x)$ if Y_i is more accurate.

Clearly, Eq. (2) is a special case of Eq. (4). However, both are approximations of the conditional pdf $f_{X|y_1,y_2}(x)$. In the following, we will derive a closed-form expression of $f_{X|y_1,y_2}(x)$ directly.

By Bayes' formula, $f_{X|y_1,y_2}(x)$ can be written as (for simplicity, we drop the subscript of the pdf)

$$f(x|y_1,y_2) = \frac{f(x,y_1,y_2)}{f(y_1,y_2)} = \frac{f(y_1,y_2|x)f(x)}{f(y_1,y_2)}. \tag{6}$$

As usual, we assume $y_i = x + e_i$, where e_i has zero-mean Laplacian distribution. We also assume that given x, e_1 and e_2 are independent. Therefore, we have $f(y_1,y_2|x) = f(y_1|x)f(y_2|x)$. Invoking the Bayes' formula again, the conditional pdf $f_{X|y_1,y_2}(x)$ can be converted into

$$f(x|y_1,y_2) = \frac{f(x|y_1)f(x|y_2)f(y_1)f(y_2)}{f(x)f(y_1,y_2)}. \tag{7}$$

Plugging this into Eq. (1), the terms $f(y_1)$, $f(y_2)$, and $f(y_1,y_2)$ can be canceled, and the optimal reconstruction of x becomes

$$\hat{x}_{opt} = \frac{\int_{z_i}^{z_{i+1}} x \frac{f_{X|y_1}(x) f_{X|y_2}(x)}{f(x)} dx}{\int_{z_i}^{z_{i+1}} \frac{f_{X|y_1}(x) f_{X|y_2}(x)}{f(x)} dx}. \tag{8}$$

Note that this solution needs $f(x)$, which can be estimated by the bidirectional motion estimation-based SI (Y_0) which is better than forward prediction SI (y_1) and Backward prediction SI (y_2). Our experiments show that the result is not sensitive to $f(x)$, so we can also model it by a Laplacian distribution with parameter α. As a result, the above equation can be expressed as follows:

$$\begin{aligned} \hat{x}_{opt} &= \frac{\int_{z_i}^{z_{i+1}} x \frac{\frac{\alpha_1}{2} e^{-\alpha_1|x-y_1|} \frac{\alpha_2}{2} e^{-\alpha_2|x-y_2|}}{\frac{\alpha}{2} e^{-\alpha|x|}} dx}{\int_{z_i}^{z_{i+1}} \frac{\frac{\alpha_1}{2} e^{-\alpha_1|x-y_1|} \frac{\alpha_2}{2} e^{-\alpha_2|x-y_2|}}{\frac{\alpha}{2} e^{-\alpha|x|}} dx} \\ &= \frac{\int_{z_i}^{z_{i+1}} x e^{-\alpha_1|x-y_1| - \alpha_2|x-y_2| + \alpha|x|} dx}{\int_{z_i}^{z_{i+1}} e^{-\alpha_1|x-y_1| - \alpha_2|x-y_2| + \alpha|x|} dx}. \end{aligned} \tag{9}$$

3.2. **Classified Correlation Noise Model.** As shown in [11], classifying the data into different groups and obtaining different Lapacian parameters in different groups can improve the R-D performance of the DVC system. This method is adopted in this paper.

In this process, the α_1 and α_2 can be founded in the different classified tables of the correlation noise parameter according to the accuracy of the SIs. The training data for α_1 and α_2 are obtained from four sequences *news*, *coastguard*, *carphone* and *highway* offline, as in [11]. The tables of α_1 and α_2 are calculated by the residual frame between the original WZ frame and the corresponding SIs. However, the thresholds which are used to decide the class of each coefficient are obtained by the residual frame between Y_0 and the corresponding SI. In this paper, the coefficients of residual frame are classified in to 8 groups which almost have the same number of elements. For every group, we can get the corresponding α_1 and α_2. Because the thresholds will be used in the decoding and only Y_0 and the corresponding SI are available at the decoder. The residual frames which are used to achieve the correlation noise parameters (α_1 and α_2) can be gained by the following formulas:

$$R_i(x,y) = WZ(x,y) - Y_i(x,y), \quad i = 1,2. \tag{10}$$

The residual energy which is used to decide the class of the α_1 and α_2 is given by

$$E_i(x,y) = \frac{1}{M*N}\sum_{x=1}^{M}\sum_{y=1}^{N}[Y_i(x,y) - Y_0(x,y)]^2, i = 1,2, \tag{11}$$

where M and N represent the block size. In this paper, $M = 4$ and $N = 4$.

The threshold values are based on the residual energy. In addition, there are different-quality SIs (including Y_1, Y_2 and Y_0) for different QP. So, we choose different thresholds for different QPs and different SIs. For example, when $QP = 24$, the thresholds for Y_1 and Y_2 are

$$\begin{aligned}&[0,0.0130,0.0521,0.2005,0.7204,3.3979,65.2484],\\ &[0,0.0131,0.0528,0.2031,0.7294,3.3965,62.9480].\end{aligned} \tag{12}$$

4. **Experiment and Simulation Results.** In this section, we compare the performances of H.264 intra mode, DISCOVER [6], our DVC scheme using two types of SIs, namely, two SIs with the weighted pdf in Eq. (4) [9] and the Bayesian pdf derived in this paper. The test video sequences are $foreman$ (with the Siemens logo). The resolution is qcif and the frame rate is 15 fps.

Fig. 3 shows the rate-distortion performance for the sequence $foreman$. The key frames are encoded by H.264 intra mode, and the corresponding QPs for different rate-distortion points are obtained from [6].

The results show that the performance of multi-hypothesis Wyner-Ziv video coding proposed in this paper is more efficient than H.264 Intra mode and DISCOVER. The result of one-SI version of our scheme is better than DISCOVER. The reason is different binary codec is exploited. The complement is used to be the binary codec of the index and corresponding Gray code is used in the LDPCA encoder. If using general binary codec and corresponding Gray code in our scheme, the result using one SI will be similar with DISCOVER. The corresponding results are given in Fig. 4.

From Fig. 3, we can see that the method with the weighted average of the individual conditional pdf as in (4) have slightly better performance than the Bayesian pdf. One possible reason is that $f(x)$ is not estimated accurately enough. This will be our future work.

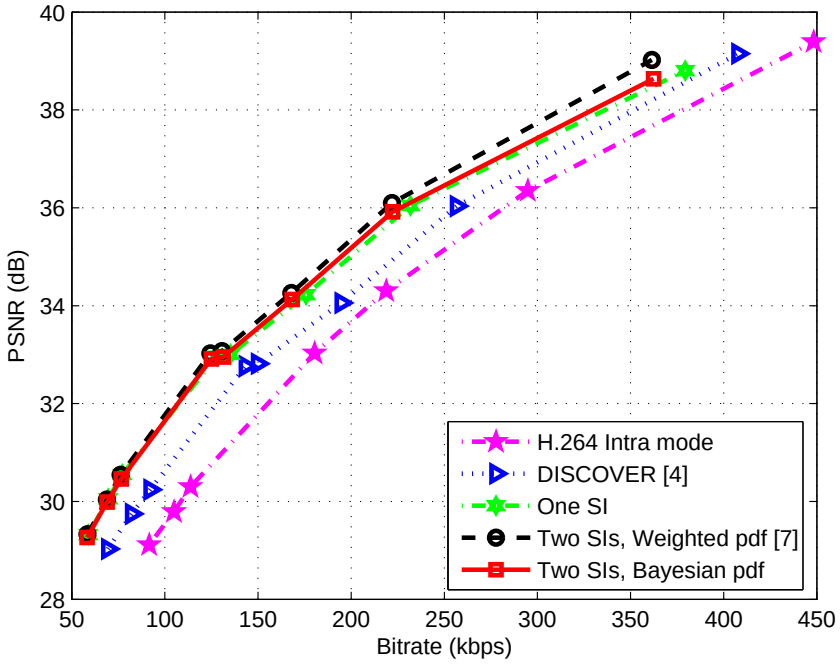

FIGURE 3. Rate-Distortion results for Foreman.

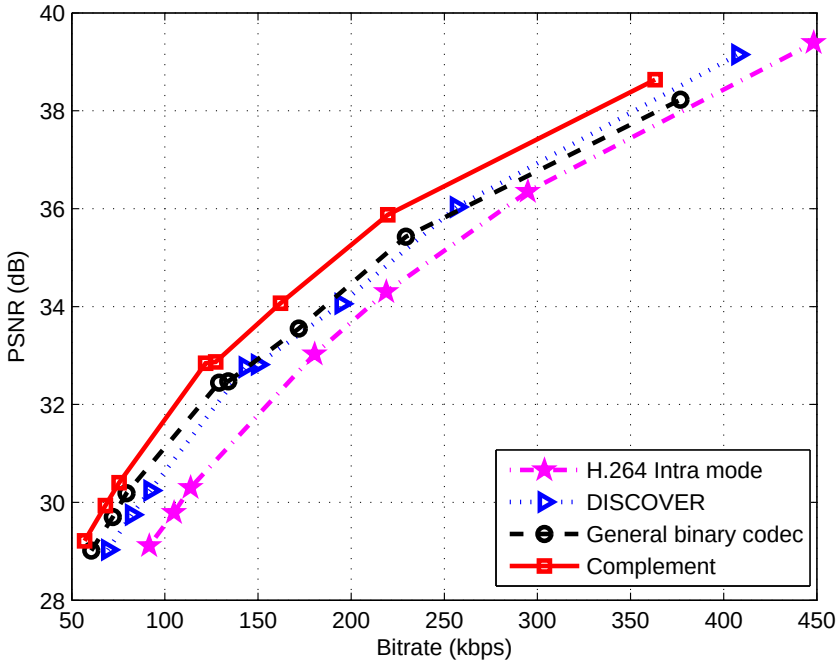

FIGURE 4. Rate-Distortion results using different binary codec.

Next, we study the impacts to different schemes when better SIs are available. To simulate SIs of different qualities, we scale the previous SIs as follows:

$$\begin{aligned} e_i &= y_i - x, \\ y_i' &= x + s * e_i, \end{aligned} \tag{13}$$

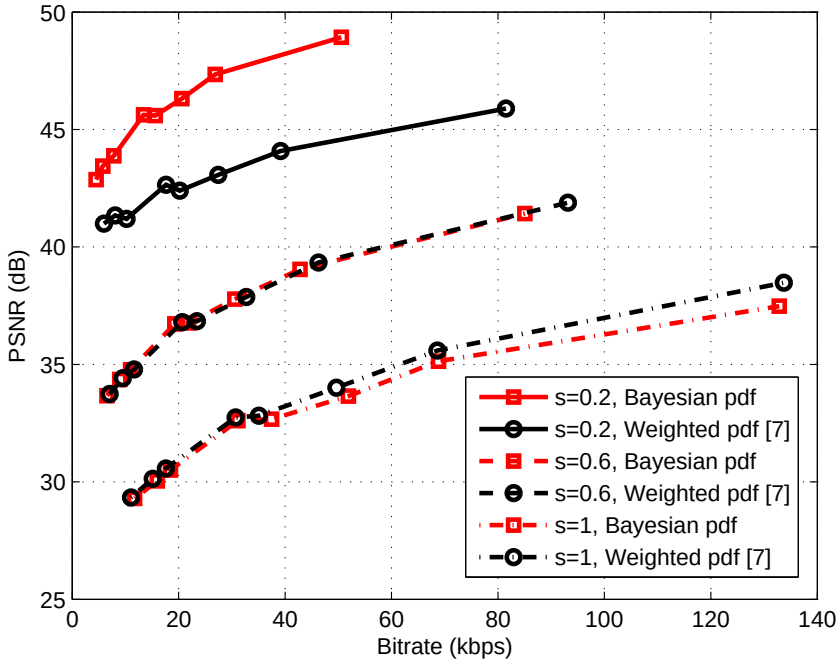

FIGURE 5. Rate-Distortion results using different-quality SI for Foreman.

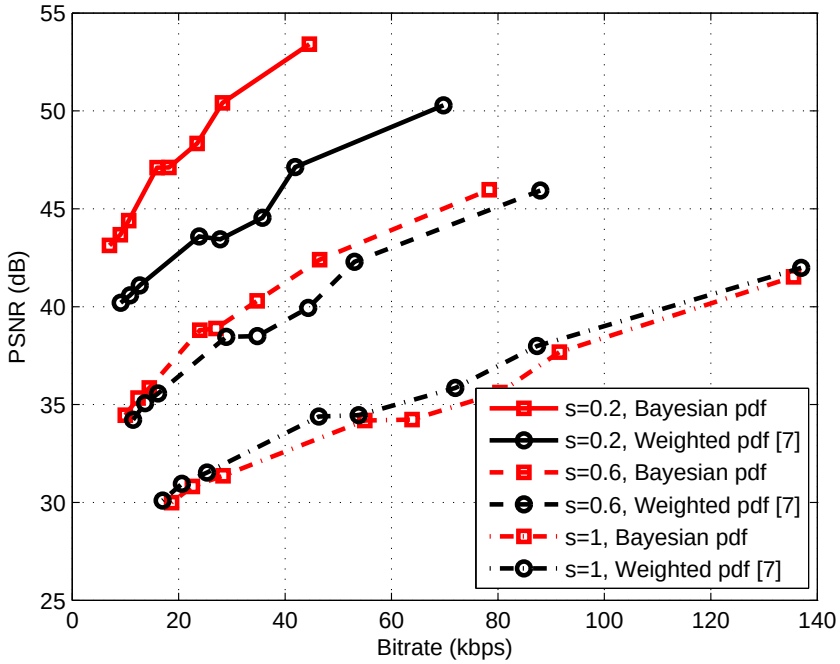

FIGURE 6. Rate-Distortion results using different-quality SI for Mother.

where the scaling factor s controls the quality of the new SI. When $s = 1$, it becomes the SI obtained using forward or backward motion estimation.

Fig. 5 and Fig. 6 show the results of different methods with different scalaing factors s. The test sequence are *foreman* and *mother*. At the decoder, we can only get the real SIs ($s = 1$). Therefore, only average PSNR of WZ frames at different bit rates are

given in the Fig. 5 and Fig. 6. The key frames are encoded by H.264 Intra mode and the corresponding quantization steps are the same as the DISCOVER [6].

From Fig. 5 and Fig. 6, it can be seen that when $s < 0.6$, the Bayesian pdf method can outperform other methods, and the gain increases with the improvement of the side information. This shows the potential of the Bayesian approach.

5. **Conclusion.** In this paper, a Bayesian multi-hypothesis Wyner-Ziv video coding scheme is developed. Experimental results show that the Bayesian solution has similar performance to previous methods. Moreover, simulation results reveal that when better SIs are available, the Bayesian solution could outperform other methods. Our future work will focus on how to improve the quality of the side information, as well as how to estimate the source pdf $f(x)$ more accurately.

Acknowledgment. The work is partially supported by the National Natural Science Foundation of China (No. 61402268, 61373081, 61401260, 61572298, 61601269, 61602285, 61601268), the Technology and Development Project of Shandong (No. 2013GGX10125), the Natural Science Foundation of Shandong China (No. BS2014DX006, ZR2014FM012, ZR2015PF006, ZR2016FB12) and the Taishan Scholar Project of Shandong, China.

REFERENCES

[1] J. D. Slepian and J. K. Wolf, Noiseless coding of correlated information sources, *IEEE Trans. on Information Theory*, vol. IT-22, pp. 471–480, Jul. 1973.

[2] A. D. Wyner and J. Ziv, The rate-distortion function for source coding with side information at the decoder, *IEEE Trans. on Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

[3] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, The DISCOVER codes: architecture, techniques and evaluation, *in Proc. Picture Coding Symp.*, Lisbon, Portugal, Nov. 2007.

[4] A. Abou-Elailah, F. Dufaux, J. Farah, M. Cagnazzo, Fusion of Global and Local Motion Estimation Using Foreground Objects for Distributed Video Coding, *IEEE Trans. on Circuits & Systems for Video Technology*, vol. 25, no.6, pp. 973–987, 2015.

[5] H. Bai, A. Wang, Y. Zhao, J. S. Pan, and A. Abraham, (2011). Distributed multiple description coding: principles, algorithms and systems. *Springer Science & Business Media*, 2011.

[6] http://www.discoverdvc.org/

[7] D. Kubasov, J. Nayak and C. Guillemot, Optimal reconstruction in Wyner-Ziv video coding with multiple side information, *IEEE Multimedia Signal Processing Workshop*, Oct. 2007.

[8] K. Misra, S. Karande and H. Radha, Multi-hypothesis distributed video coding using LDPC codes, *Proc, Allerton Conference on conmmunication, control and computing*, Sep. 2005.

[9] Y. Li, H. Liu, X. Liu, S. Ma, D. Zhao and W. Gao, Multi-hypothesis based multi-view distributed video coding, *Picture Coding Symp.*, May 2009.

[10] D. Varodayan, A. Aaron, and B. Girod, Rate-adaptive codes for distributed source coding, *EURASIP Signal Processing Journal*, vol.86, pp.3123C-3130, Nov. 2006.

[11] G. Esmaili and P. Cosman, Wyner-ziv video coding with classified correlation noise estimation and key frame coding mode selection, *IEEE Trans. Image Processing*, vol.20, no.9, pp.2463C-2474, Sept. 2011.