Distributed Video Coding with Spatial Correlation Exploited Only at the Decoder

Mei Guo[1,*],Yan Lu[2],Feng Wu[2],Shipeng Li[2],Wen Gao[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[2]Microsoft Research Asia, Beijing, 100080, China

Abstract—**A new pixel-domain Distributed Video Coding (DVC) scheme is proposed in this paper, in which both the temporal and the spatial correlations are exploited only at the decoder. A video is treated as a collection of data correlated in temporal and spatial directions. Besides splitting a video into frames at different time instants, a frame is further split by spatially sub-sampling. Each yielded part is then encoded individually. At the decoder, the side information signals are from both adjacent frames and the spatially decoupled signals. To utilize these multiple side information signals, a new probability model is proposed, in which the transitional probability is calculated from the conditional probabilities on the multiple side information signals. The coding efficiency is enhanced by further removing the spatial redundancy, while the encoding complexity remains the same as the previous pixel-domain DVC techniques that only consider the temporal correlation.**

I. INTRODUCTION

Distributed Video Coding (DVC) technologies are committed to compress videos with the limited resources [1]. Slepian-Wolf theorem has suggested that the two correlated independent identically distributed (i.i.d) sequences X and Y can be encoded losslessly with the same rate as that of the joint encoding as long as the collaborative decoders are employed [2]. Wyner and Ziv have extended this theorem to the lossy coding of continuous-valued sources [3]. Based on the Wyner-Ziv theorem, the task of exploiting correlation can be shifted to the decoder, and the low complexity encoding is then achieved. This feature is quite desirable in mobile devices and distributed sensor networks.

A turbo-code based Wyner-Ziv (WZ) video coding scheme performed in pixel domain is presented in [1]. The WZ frame is encoded with turbo code. The delivered parity bits are decoded jointly with the side information generated from the previously reconstructed adjacent frames. To further exploit the spatial correlation, WZ coding is further applied to transform domain such as DCT [4] or wavelet [5]. In particular, some high frequency DCT coefficients as hash words can be transmitted to the decoder for the purpose of assisting motion estimation [1]. Besides the turbo code based DVC schemes, another DVC system based on syndrome coding is proposed in [6]. Within the coset specified by the received syndrome bits, the codeword is selected depending on the side information derived from the adjacent decoded frames.

Actually, most of the previous work has been focuses on removing the spatial redundancy by using transform at the encoder. Is it possible to exploit the spatial correlation only at the decoder? The approach introduced in [7] considers one frame as a two-dimensional stationary Markov random field and utilizes the spatial correlation at the decoder. When it deals with the continuous-tone source, it leads to too many states with the computation on a trellis to be used in practice. Actually, the correlation in a video inherently exists in both temporal direction and spatial directions. Besides splitting a video into frames at different time instants, a frame is further split by spatially sub-sampling. At the decoder, there are multiple side information signals derived from the adjacent frame and the neighboring pixels in one frame. Since the decoder can not know which one is better without the knowledge of the original signal to be decoded, the key issue is how to use the multiple side information at the decoder.

In this paper, we propose a WZ video coding scheme, in which both the temporal and the spatial correlations are exploited only at the decoder. Each symbol to be decoded has multiple side information including one from temporal direction and some decoded neighboring pixels. Since side information is usually considered as a noisy copy of source, the multiple side information signals are taken as the multiple copies of source transmitted through the different virtual channels. The turbo decoder is to combine all the conditional probabilities derived from the multiple side information signals in the decoding process. In summary, the exploitation of spatial correlation enhances the coding efficiency, whereas the encoding complexity remains the same as the original turbo code based WZ coding performed in pixel domain.

The remainder of this paper is arranged as follows. In Section II, the exploitation of the spatial correlation is analyzed. In Section III, the proposed DVC scheme is described in detail. The experimental results are shown in Section Ⅳ. Finally, Section V concludes this paper.

---

*This work was done when M. Guo was with Microsoft Research Asia as a Research Intern.

1-4244-0921-7/07 $25.00 © 2007 IEEE.

II. THE EXPLOITATION OF THE SPATIAL CORRELATION

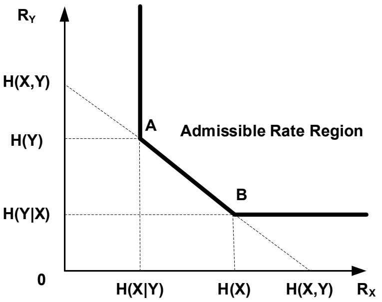


Figure 1. The admissible rate region for separate encoding and joint decoding X and Y. $R_X \geqslant H(X|Y)$, $R_Y \geqslant H(Y|X)$, and $R_X+R_Y \geqslant H(X,Y)$.

Denote R_X and R_Y as the minimum number of bits per symbol needed to encode the sequences X and Y respectively so that they can be reproduced perfectly at the decoder side. When X and Y are separately encoded but jointly decoded, the admissible rate region of R_X and R_Y established in [2] is depicted in Figure 1. The compression techniques intend to encode X and Y with the sum of R_X and R_Y approaching the line AB namely the joint entropy H(X, Y). The vertex points A and B correspond to the asymmetric case in DVC, which means X (Y) is encoded in the presence of the correlated Y(X) at the decoder. As for point A, in an ideal situation,

$$\begin{aligned} R_x &= H(X\,|\,Y) \\ R_y &= H(Y) \\ R_x + R_y &= H(X\,|\,Y) + H(Y) = H(X,Y) \end{aligned} \quad . \tag{1}$$

Other points on the line AB except A and B correspond to the symmetric case, where both encoders of X and Y send partial information to the decoder. It is can be achieved by binary linear code [8]. The most-likely codeword pair is searched for from the pair of cosets which are determined by the received syndromes from two encoders. Due to the consideration of complexity, most of the existing DVC techniques utilize temporal correlation in asymmetric manner [1] [4] [5] [7]. The decoder takes the previously decoded frames as the side information.

A video can be viewed as a collection of data correlated in temporal and spatial directions. How to split the 3-dimensional data determines the correlation that can be exploited. Besides the temporal splitting, a frame can be further divided spatially by sub-sampling. Each yielded part is encoded separately. Thus, the spatial correlation can be exploited in asymmetric manner. At the decoder side, there will be multiple available side information including the one from previously decoded adjacent frames and the neighboring pixels in one frame. Since the current part to be decoded is unknown to the decoder, the decoder can not specify which side information is more correlated. Our proposed scheme which utilizes the multiple side information will be introduced in the next section.

III. PROPOSED SCHEME

The proposed WZ video coding scheme is illustrated in Figure 2. The intra frames, i.e. the odd frames, are encoded and decoded with conventional video coding system. The inter frames, i.e. the even frames, are WZ coded.

A. Encoder

The WZ frame is split into four parts as shown in Figure 2(b). Four neighboring pixels are arranged in four sets P1, P2, P3 and P4 respectively. They are encoded and transmitted successively. We adopt the WZ encoder as depicted in Figure 2(a). The pixel values are scalar quantized first. Then, the symbols are fed into a turbo encoder. The generated parity bits are punctured and delivered according to the correlation. To estimate correlation at the encoder is quite a tough task since independent encoding is performed. Thus, the rate adaptation is realized through the feedback channel from the decoder. If the decoder can not decode the symbols correctly, a bit-request is sent back to the encoder. Then, some additional punctured parity bits are transmitted [1]. The process is repeated until the frame is successfully decoded.

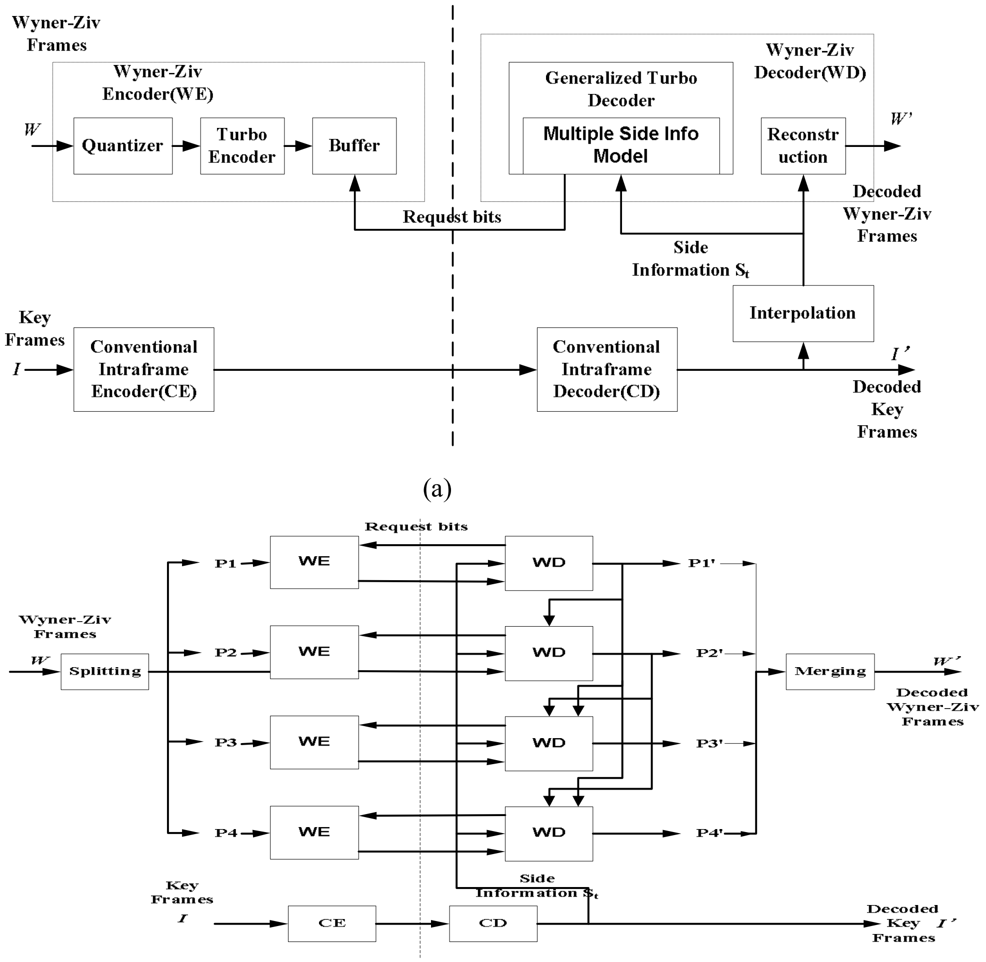
B. Probability Calculation with Multiple Side Information

In our scheme, the turbo decoder is extended to handle multiple side information, since two categories of side information are available: S_t yielded temporally by the interpolation of the adjacent frames and S_s containing the decoded neighboring pixels. The first part P1 is decoded without any spatial redundancy removed. All the other three parts will make use of four decoded neighboring pixels.

The *multiple side info model* is committed to calculate probabilities conditioned on multiple side information. These multiple side information can be considered as the received copies when the identical source signal is transmitted through different channels. In digital communication system, code combing technique is devoted to operate on all the received repeated error-detected packets to obtain a more reliable estimate [9]. The estimate *m* is selected by

$$\max_{m} \sum_{i=1}^{L} w_i * c_i \tag{2}$$

where there are L received copies. w_i denotes the weighting factor of copy *i* and indicates the nature of the corresponding channel. c_i represents the metric of the agreement between the estimate and copy *i*. In turbo decoding, the conditional probability on the side information is usually regarded as the metric of agreement between the estimated value of the current symbol and the received side information. The weighting factor is determined by the correlation between the source and the side information. Since the knowledge of correlation can not be obtained, all side information have the



(a)

(b)

Figure 2. Our proposed WZ coding system exploits the spatial correlation only at the decoder.

identical weight. We extend the calculation of probability instead of just selection in terms of maximum likelihood. The conditional probability is expressed as

$$P(b=n \mid SideInfo) = P(b=n \mid s_t) + \sum_{j=1}^{4} P(b=n \mid s_{sj}) . \quad (3)$$

Denote *b* as the current symbol to be decoded and *n* as the estimated value. The statistical distribution of difference between source and side information is assumed to have a Laplacian model. Therefore, the conditional probability can be further calculated as

$$P(b=n \mid SideInfo) = P(\tilde{n}-s_t) + \sum_{j=1}^{4} P(\tilde{n}-s_{sj}) \quad (4)$$

where $\tilde{n}$ is yielded by de-quantization of *n*.

C. Decoding

Turbo decoder decodes the WZ bits in conjunction with the side information. The decoding process is based on the judgement of posterior probability (PP). Given a possible value *j*, PP is expressed as

$$PP = \sum_{\chi_j} \alpha_{i-1}(s')\gamma_i(s',s)\beta_i(s) \quad (6)$$

where χ_j is the set of all transitions from state s' to s with the input *j*. The probability functions $\alpha_i(s)$ and $\beta_i(s)$, which are the probabilities of approaching the state *s* at the instant *i* from the starting state and the ending state respectively, can be recursively calculated from the probability $\gamma_i(s', s)$ as introduced in [10]. The transitional probability is represented as

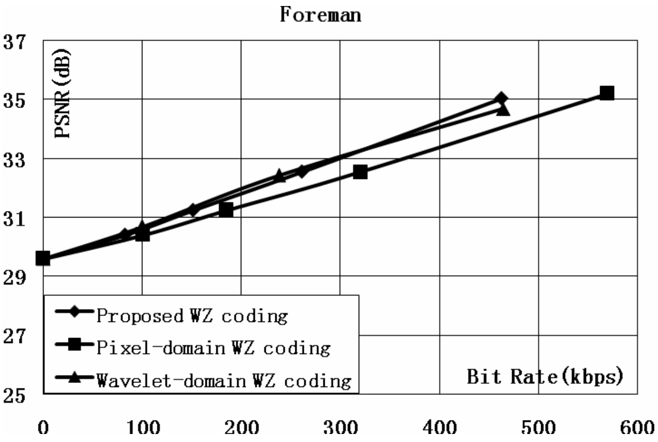
$$\gamma_i(s',s) = P(b=j)P(b=j \mid SideInfo)P(u_i \mid p_i) \quad (7)$$

where u_i is the output parity bit of the transition from state s' to *s* with the input *j* and p_i is the received parity bit correspondingly. The conditional probability given side information is calculated as (4).

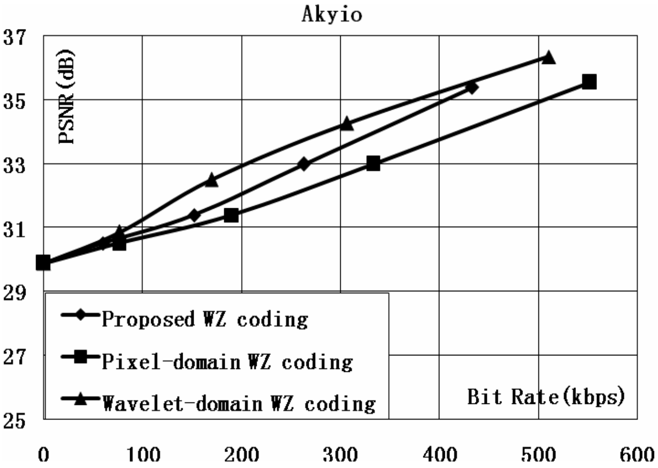
Four parts from P1′ to P4′ are reconstructed from quantized symbols and corresponding side information as introduced in [1]. The final reconstructed frame *W′* is yielded by merging those four reproduced parts.

IV. EXPERIMENTAL RESULTS

In order to verify the coding efficiency, we implemented our proposed WZ video coding scheme. Two sequences Foreman and Akyio in QCIF format with 30 frames per second are used in the test. In each sequence, the first 100 frames are selected. We set the GOP structure as IWIW, where I represents intra frame and W represents WZ frame. Intra frames are encoded with H.263 intra coding and decoded independently at the decoder to generate side information. We use a symmetric motion estimation algorithm which interpolates two adjacent decoded frames to generate the side information. In some areas with non-linear motion, the side information may not be accurate and the WZ bits can be used to compensate for the loss.



(a)



(b)

Figure 3. The coding performance of luminance component (a) Foreman sequence (b) Akyio sequence.

Figure 3 gives the coding performance for luminance component of the two sequences and only WZ frames are illustrated. Our proposed scheme is compared with WZ coding in pixel domain proposed in [2]. It is also taken as a benchmark of WZ video coding. As indicated in Figure 3, our scheme achieves around 1dB gain in terms of rate-distortion performance. The exploitation of spatial correlation improves the overall coding performance of WZ video coding. Furthermore, we compare our proposed scheme with the WZ coding using wavelet, which is proposed in [5]. It is can be observed that our scheme can achieve comparable coding performance to the method that removes the spatial redundancy by performing wavelet transform at the encoder. Accordingly, the spatial correlation can be exploited only at the decoder so that the complexity of encoding system is the same as that of pixel-domain WZ encoding system which doesn't explore the spatial correlation.

V. CONCLUSIONS

In this paper, we have presented that the spatial correlation can be exploited only at the decoder instead of performing transform at the encoder side. A practical WZ video coding system is proposed, in which the probabilities derived from both temporal and spatial side information are investigated by using the multiple side information model. The exploitation of correlation is only involved in the decoder, so no complexity burden is added to the encoder.

ACKNOWLEDGEMENTS

The authors wish to thank Professor Da Peng Wu for the valuable discussion with him.

REFERNCES

[1] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed Video Coding", Proc. IEEE, vol. 93, no. 1, Jan. 2005.

[2] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. on Inf.Theory, vol. IT-19, 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 Inf.Theory, vol.IT-22, no, 1, Jan. 1976.

[4] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video", Proc. of SPIE, Visual Communications and Image Processing, 2004

[5] X. Guo, Y. Lu, F. Wu, W. Gao, "Distributed Video Coding Using Wavelet", International Symposium on Circuits and Systems, Island of Kos, Greece, May 2006

[6] R. Puri and K. Ramchandran, "PRISM: A "Reversed" Multimedia Coding Paradigm", presented at the IEEE Int. Conf. Image Processing, Barcelona, Spain, 2003.

[7] D. Varodayan, A. Aaron, and B. Girod, "Exploiting Spatial Correlation in Pixel-Domain Distributed Image Compression", Proc. International Picture Coding Symposium, Beijing, P.R. China, Apr. 2006

[8] S. S. Pradhan and K. Ramchandran, "Distributed source coding: symmetric rates and applications to sensor networks", Proc. IEEE Data Compression Conference, pp. 363-372, Mar. 2000

[9] D. Chase, "Code Combing------A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets", IEEE Trans. on Commun. vol. Com-33, no. 5, May 1985

[10] C. Berrou, A, Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-Codes," Proc. IEEE International Conference on Communications, Geneva, Switzerland, pp. 1064-1070, May 1993