Multi-hypothesis Based Multi-view Distributed Video Coding

Yongpeng Li[1], *Hongbin Liu*[2], *Xianming Liu*[2], *Siwei Ma*[3], *Debin Zhao*[2], *Wen Gao*[3]

[1]Graduate University, Chinese Academy of Sciences, Beijing 100039, China

[2] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

[3] School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

{ypli, hbliu, xmliu, dbzhao,wgao}@jdl.ac.cn

ABSTRACT

This paper proposes a multi-hypothesis based Wyner-Ziv (WZ) decoder for the multi-view distributed video coding (MDVC). Two hypotheses, the intra-view SI and the inter-view SI, are fed together into the WZ decoder in the proposed scheme. A multi-hypothesis based correlation model (MHBCM) is presented to fully exploit the redundancy between these two SI frames and the original frame. The MHBCM is also applied on the optimal minimum mean-square error reconstruction of the quantized samples. The simulation results show that the proposed algorithms are able to significantly improve the coding efficiency of the MDVC system.

***Index Terms*—**multi-view video coding, distributed video coding, correlation model, reconstruction

1. INTRODUCTION

In the state of art video coding schemes like MPEG and H.264/AVC, the high complex motion compensation is performed at the encoder to achieve compression. The decoding is relatively low-complexity. Therefore, this coding paradigm suits for the applications like broadcasting and video-on-demand system etc., in which the video is compressed once but has to be decoded several times. This coding paradigm, however, is being challenged by the emerging applications such as the low power wireless video surveillance and the multimedia sensor networks etc., as the low–complexity encoders are significantly desirable in these applications.

These requirements lead to a novel video coding paradigm, distributed video coding (DVC), in which the majority computational bulk is shifted to the decoder. DVC is built on the Slepian-Wolf theory [1] and the Wyner-Ziv theory [2]. Slepian and Wolf first proved that although two statistical sources X and Y are independently encoded, the similar performance can be achieved as long as the joint decoding of them is allowed for the lossless coding. Wyner and Ziv extended the theory to the lossy coding with side information at the decoder.

The Wyner-Ziv (WZ) coding utilizing multiple SI is referred to as the multiple-hypothesis WZ coding. The authors in [3] explain why the coding efficiency can be improved when using multiple SI, and they also presented an efficient solution for the multi-hypothesis based DVC. In their proposed solution, the multiple-hypothesis based correlation model (MHBCM), the core of the multiple-hypothesis based WZ coding, is obtained by training on several sequences. Besides, the authors in [4] also investigate the scenario with multiple SI in the WZ coding. They make use of the error probability of the decoded block to select the best SI among those that are available at the decoder. All the above works, however, are focusing on the mono-view distributed video coding. It is expected that the multi-hypothesis based solution would bring promising performance improvements in the multi-view distributed video coding (MDVC) [5], where multiple hypotheses can be obtained by employing the inter-view correlation as well as the temporal correlation. Based on this inspiration, we extend the mono-view multi-hypothesis to the multi-view scenario. A novel MHBCM is presented to fully exploit the correlation between the original frame and the multiple hypotheses. It is also applied on the optimal minimum mean-square error reconstruction of the quantized samples. The experiment results demonstrate that the proposed multiple-hypothesis based decoding algorithms can significantly improve the coding efficiency.

This paper is organized as follows. In Section 2, the framework of the proposed decoding structure is first presented, and the two kinds of SI generation methods, the proposed MHBCM and the reconstruction technique are also included in this section. Experiments results are shown in Section 3 and conclusions are given in Section 4.

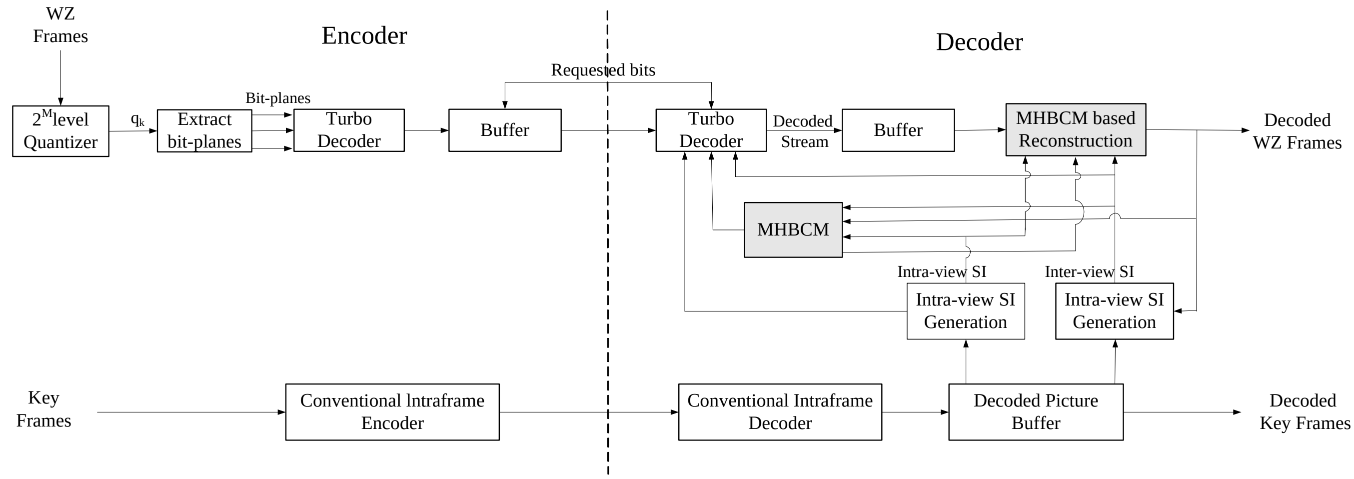


Fig.1 Proposed system

2. PROPOSED SYSTEM

The proposed system is illustrated in Fig. 1. In the system, each view sequence is independently encoded without communicating with each other. In each view sequence, the even and odd frames are encoded as the key frame and the WZ frame, respectively. The key frame is coded using the conventional intra coding method. The pixels of WZ frames are first quantized into 2^M levels using a uniform scalar quantizer, and then the quantized pixels are fed into a turbo codec, which consists of two identical constituent convolution codecs [6]. The parity bits generated by the turbo codec are stored in the buffer.

At the decoder, the key frames are independently decoded using the intra decoding method, and the corresponding reconstruction frames are used to generate the SI for the WZ frames. The proposed system uses two kinds of SI: the intra-view SI which is generated using the temporal interpolation (TI) and the inter-view SI that is generated using the disparity-guided temporal interpolation (DGTI), the scheme to generate the SI by employing the adjacent views' information.

These two SI frames are converted into the soft-input information needed for the turbo decoding by the proposed MHBCM. And then the parity bits, used for correcting the errors in the SI frames, are successively requested by the turbo decoder, which is composed of two soft-input soft-output (SISO) decoders implemented using the Maximum A-Posteriori (MAP) algorithm, through a feedback channel until a pre-defined threshold (10^{-3} in our MDVC system) of bit error rate is satisfied. Finally, the current frame is reconstructed by applying the proposed multi-hypothesis based reconstruction technique on the decoded stream and the two SI frames.

2.1. The generation of multi-hypothesis

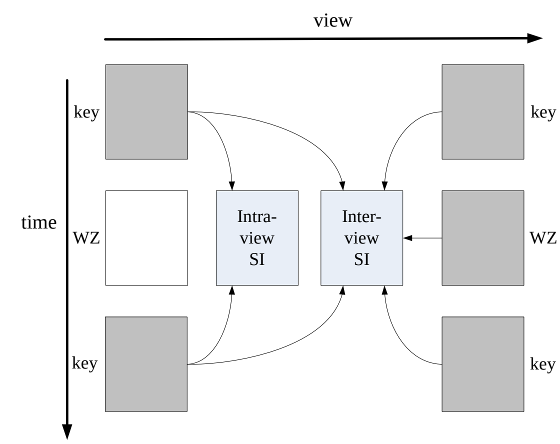


Fig. 2 Intra-view and inter-view side information generation

The proposed system utilizes the intra-view SI and the inter-view SI together to recover the original frame. As shown in Fig. 2, the intra-view SI is generated using the scheme known as temporal interpolation, in which the SI for the current WZ frame is predicted by its adjacent forward and backward key frames. The prediction is based on the translational model, which assumes that the motion in the successive frames is linear. Obviously, the linear motion assumption is not always satisfied. The TI consequently can not predict the SI well enough in the regions with high motion, especially the regions where the translational model is not satisfied. The inter-view SI is generated using the scheme called disparity-guided temporal interpolation (DGTI). The prediction of this scheme is based on the fact that the motion of the current view can be estimated by that of the adjacent views. The DGTI scheme can be completed by three steps: firstly, the disparity estimation is performed on the decoded key frames of the adjacent views so that the corresponding elements (block) in these inter-view adjacent frames can be determined; secondly, the motion in the decoded view is estimated and recorded; thirdly, the motion of the decoded view is applied on estimating the motion of the current view, which are then used to generate the SI for the current view.

2.2. The multi-hypothesis based correlation model

The correlation model (CM), which describes the relationship between the SI frames and the original frame, is one of the most vital factors that determine the coding efficiency of the DVC system. The more accurate the CM is, the better coding efficiency the DVC system will achieve. A mechanism, which can distinguish the relatively well predicted parts from the bad ones within the SI frame, is introduced in our proposed MHBCM. This mechanism is realized by the method called the set based segmentation, which is described in (1).

$$(x,y) \in Set_i, if\ BP(IntraVSI(x,y),i) == BP(InterVSI(x,y),i) \quad i \in [0,8] \tag{1}$$

where the $IntraVSI$ and the $InterVSI$ represent the intra-view SI frame and the inter-view SI frame, respectively, and (x,y) is the pixel coordinate. $BP(value,i)$ is a function that returns the top i MSB of $value$. It should be noted that once a pixel belongs to the set_i, it does not belong to the set_j for $j<i$ anymore. The set segmentation is based on the assumption that the more MSB of the co-located pixels of the intra-view SI frame and the inter-view SI frame match with each other, the better the corresponding pixel will be predicted. It implies consequently that the pixels belonging to the set_i are better predicted than those belonging to the set_j when $j<i$. We also assume that in each set the differences between the original frame and the intra-view (inter-view) SI frames can be characterized by the Laplacian distribution. For each single hypothesis the set based CM is described in (2),

$$f(WZ(x,y), SI(x,y)) = \frac{\alpha_i}{2} e^{\alpha_i |WZ(x,y)-SI(x,y)|}, if\ (x,y) \in Set_i \tag{2}$$

where the WZ and the SI represent the original frame and the SI frame respectively. The α_i is the parameter of the Laplacian distribution that models the distribution of the differences between the WZ and the SI in set_i.

The proposed MHBCM is approximated by the linear weighted addition of each single hypothesis correlation model. Such CM is presented in (3),

$$\begin{aligned} & f(WZ(x,y), IntraVSI(x,y), \mathrm{int}\,erVSI(x,y)) \triangleq \\ & w_1 * \frac{\alpha_{i,interV}}{2} e^{-\alpha_{i,interV}|WZ(x,y)-InterVSI(x,y)|} + \\ & w_2 * \frac{\alpha_{i,IntraV}}{2} e^{-\alpha_{i,ntraV}|WZ(x,y)-IntraVSI(x,y)|}, if\ \ (x,y) \in Set_i \\ & w_1 = \alpha_{i,interV}{}^2 / (\alpha_{i,\mathrm{int}\,erV}{}^2 + \alpha_{i,\mathrm{int}\,raV}{}^2) \\ & w_2 = 1 - w_1 \end{aligned} \tag{3}$$

where the WZ, the $IntraVSI$ and the $InterVSI$ represent the original frame, the intra-view SI frame and the inter-view SI frame, respectively, and (x,y) is the pixel coordinate. The $\alpha_{i,\mathrm{int}\,erV}$ and the $\alpha_{i,\mathrm{int}\,raV}$ are the Laplacian distribution parameters of set_i for the inter-view SI frame and the intra-view SI frame, respectively. The parameters for the current WZ frame are obtained by applying the parameter estimating method proposed in [7] on the reconstructed frame and the intra-view (inter-view) SI frame of the previously decoded WZ frame.

The proposed MHBCM has the following advantages: firstly, it has the ability to evaluate and reflect the prediction accuracy of the SI frames. The well predicted parts are given a high conditional probability when converted into the soft-input information for the turbo decoding, since they are sorted into the sets with high index. This is done by the set-based segmentation of each single hypothesis; secondly, through a weighting manner the inter-view SI frame and the intra-view SI frame are both considered. The weights are determined by the parameters of the set that the pixel belongs to, and the parameter will take a large value when the pixels in the set of corresponding SI frame is more close to the original frame.

2.3. The multi-hypothesis based reconstruction

In this subsection we concern on the reconstruction technique. The DVC system can benefit from the reconstruction technique in term of the rate-distortion performance at almost no cost of spending extra bits. The authors present an optimal technique in [8] for the optimal minimum mean-square error reconstruction of the quantized samples. In their proposed technique, the optimal reconstructed value is computed as the expectation $E[x \mid x \in [z_i, z_{i+1}), y_1, y_2]$ of the random variable x given the quantization interval $[z_i, z_{i+1})$ and the SI y_1 and y_2. The efficiency of this technique significantly depends on the ability of the CM to describe the relationship between SI frames and the original frame. We suggest to use the set-based MHBCM as shown in (4) instead of the conventional frame-level correlation model (CFLCM) used in [8] to compute the reconstructed value. Correspondingly, the optimal reconstructed value is calculated as (4):

$$\hat{x}_{opt} = E[x \mid x \in [z_i, z_{i+1}], y_1, y_2] = \frac{\sum_{z_i}^{z_{i+1}} x f_{x|y_1,y_2}(x)}{\sum_{z_i}^{z_{i+1}} f_{x|y_1,y_2}(x)} \tag{4}$$

Where y_1 and y_2 represent the inter-view SI frame and intra-view SI frame, respectively, and $f_{x|y_1,y_2}(x)$ is given in (3).

3. EXPERIMENT RESULTS

The efficiency of the proposed MHBCM and the reconstruction technique based on it are evaluated in term of rate-distortion performance of WZ frames. Experiments are carried out on the multi-view video sequences *Ballroom* and *Race1*. Both of *Ballroom and Race1* are captured by 1-D parallel cameras in resolution of 640x480, and the frame rate for them are 25fps and 30fps, respectively. Inter modes P16x16, P16x8, P8x16, P8x8 and fast motion/disparity estimation algorithm are applied during motion/disparity compensation in TI and DGTI. The 1/4 pixel spatial interpolation in reference frame is also employed to improve motion/disparity accuracy. Key frames are about 38dB.

In order to evaluate the coding efficiency of the multi-hypothesis solution, we first compare it with the single hypothesis scenario. The multi-hypothesis based scheme with the proposed MHBCM is compared with the intra-view SI and inter-view with the conventional frame level correlation model (CFLCM), respectively. In this comparison, the luminance component and bit-rates of the first 50 WZ frames are evaluated, and all the WZ frames are reconstructed using the optimal reconstruction technique based on the CFLCM so that the efficiency of the proposed MHBCM in term of decoding performance can be clearly examined. The simulation results are shown in Fig.3 It can be observed that the proposed multi-hypothesis based solution can outperform the single intra-view SI and the single inter-view SI about 2.5dB and 0.8dB for the sequence *Ballroom*, respectively. The similar performance improvements can be observed in the sequence *Race1*.

We also have tested the MHBCM based optimal reconstruction technique and compared it with the optimal reconstruction technique based on CFLCM. The simulation results are also presented in Fig.3. As the reconstruction technique does not affect the bits rate, only PSNR is improved. It can be observed that in the sequence *Ballroom* the proposed method gains up to 0.5 dB at high bits rate compared to the optimal reconstruction method in which the CFLCM is assumed, and the improvement can be up to 0.8 dB at high bits rate for the sequence *Race1*.

4. CONCLUSIONS

This paper presents a multi-hypothesis solution for the multi-view distributed video coding. A multi-hypothesis based correlation model is proposed to fully exploit the correlation between the multiple hypotheses and the source information. It is also applied on the optimal minimum mean-square error reconstruction of WZ frames. Experimental results show that the proposed algorithms can significantly improve the coding efficiency of the MDVC system.

ACKNOWLEDGEMENT

This work was supported in part by National Science Foundation (60672088 and 60736043) and National Basic Research Program of China (973 Program, 2009CB320905).

REFERENCES

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

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

[3] K. Misra, S. Karande, H. Radha, “Multi-Hypothesis Distributed Video coding using LDPC Codes”, *Proc. Allerton Conference on communication, control, and computing, Sep. 2005* .

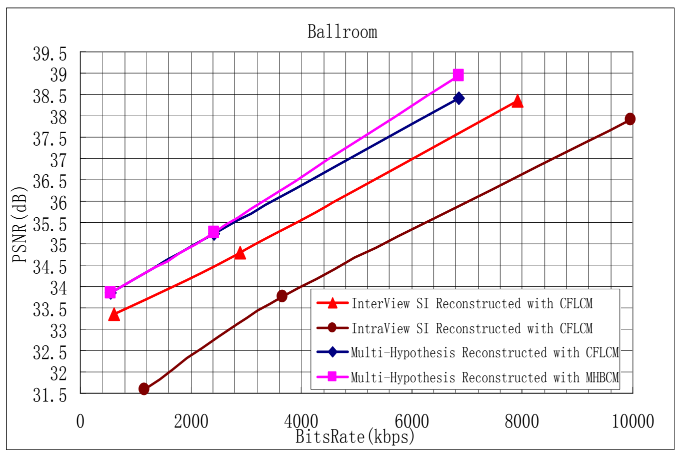
[4] A.B.B. Adikari, W.A.C. Fernando, H. Kodikara Arachchi, W.A.R.J. Weerakkody, “Multiple Side Information Streams for Distributed Video Coding”, Electronics Letters, Dec. 2006.

[5] X. Guo, Y. Lv, F. Wu, W. Gao, “Distributed Multi-View Video Coding”, *Proc. of SPIE, VCIP 2006, San Jose, California, USA, Vol. 6077, pp. 290-297, January 2006*.

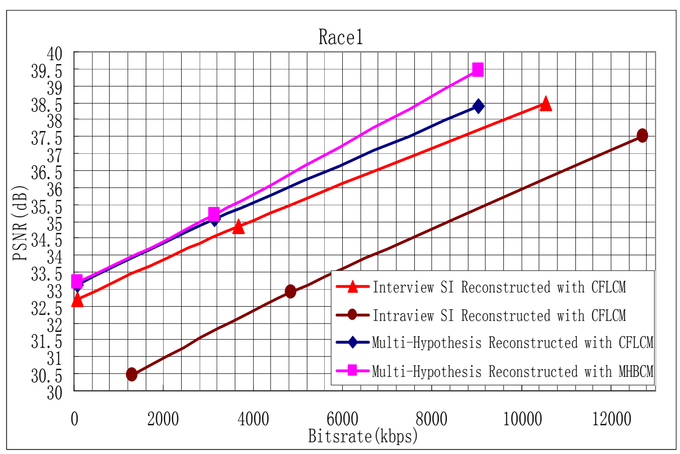
[6] B.Girod, A. Aaron, S.Rane, D. Rebollo-Monedero, “Distributed video coding”, *Proceeding of the IEEE, vol. 93, January 2005*.

[7] C. Birtes, J. Ascenso, F. Pereira, “Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding”, *IEEE International Conference on Image Processing , 2006.*

[8] D. Kubasov, J. Nayak, C. Guillemot, “Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side information”, *IEEE MultiMedia signal Processing Workshop, Chania, Oct. 2007.*



(a) *Ballroom*



(b) *Race1*

Fig. 3 Simulation results for *Ballroom* and *Race1*