Modeling Correlation Noise Statistics at Decoder for Multi-view Distributed Video Coding

Yongpeng Li[1], *Siwei Ma*[2], *Debin Zhao*[1], *Wen Gao*[2]

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

[2] School of Electronics Engneering and Computer Science, Peking University

{ypli, swma, dbzhao, wgao}@jdl.ac.cn

Abstract— Recently, multi-view distributed video coding (MDVC) receives more and more attention, as its low-complexity encoder and high-complexity decoder coding paradigm suits for many applications such as sensor networks, in which several view sequences are required to be coded by a few power-constraint encoders. Modeling the correlation noises between original frame and side information frame is a hot research issue in distributed video coding (DVC), since it is a vital factor affecting the coding efficiency. This paper firstly proposes a novel method to model the correlation noises in MDVC. And an algorithm to online estimate the model at decoder using the knowledge of adjacent views is also presented. Experiment results show that the proposed correlation model can significantly improve coding efficiency.

Index Terms—Multi-view video coding, distributed video coding, correlation model

I. INTRODUCTION

In conventional hybrid video coding paradigm, high complex motion compensation is performed at encoder to exploit temporal correlation. Hence, it suits for the applications such as broadcasting and video-on-demand system etc. in which video is compressed once but decoded several times. However, this paradigm is being challenged by emerging applications such as low power wireless video surveillance and multimedia sensor networks etc., as a low power consummation is significantly required at encoder in these applications.

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

Multi-view distributed video coding (MDVC) [3] is the extension of DVC strategy to multi-view video. In MDVC, all view sequences are encoded independently and all kinds of redundancies are exploited only at decoder. This makes MDVC an elegant solution for the applications like sensor networks and video surveillance, etc, in which several correlated view sequences have to be encoded and the resources like battery-power and computation-capability for the encoder are limited.

Pixel domain turbo code WZ coding (PDTCWZC) [4] is one of the most interesting DVC schemes due to its simplicity. Different from transform domain turbo code WZ coding, pixels are directly quantized and fed into turbo codec in PDTCWZC. This can fully alleviate the burden of encoder. The coding efficiency of PDTCWZC critically depends on the ability to model the correlation noises between original frame and side information frame. However, as stated by C. Brites, J. Ascenso and F. Pereira in [5][6] that since side information quality varies not only along sequence but also within the frame and the decoder does not have access to the original frame, correlation noise statistic modeling at decoder becomes a complex problem. To tackle such a problem, they proposed several algorithms to estimate the correlation model based on temporal information at frame level, block level and pixel level, respectively. Promising performance is achieved in their proposed algorithms. However, to the best of our knowledge, there is no work focusing on estimating correlation noises between original frame and side information with the help of adjacent view information in MDVC so far. This paper firstly proposes a novel correlation noises model based on pixel level partition of the frame in MDVC. And an algorithm to online estimate the model at decoder using adjacent inter-view information is also presented. Experiments show that the proposed method can significantly improve coding efficiency.

The rest of the paper is organized as follows. In Section Ⅱ, the framework of MDVC is first presented, and then, two kinds of side information generation methods based on intra-view information and inter-view information are introduced. The proposed correlation noises model is presented and analyzed in Section Ⅲ. Experiments results are shown in Section Ⅳ and conclusions are given in Section Ⅴ.

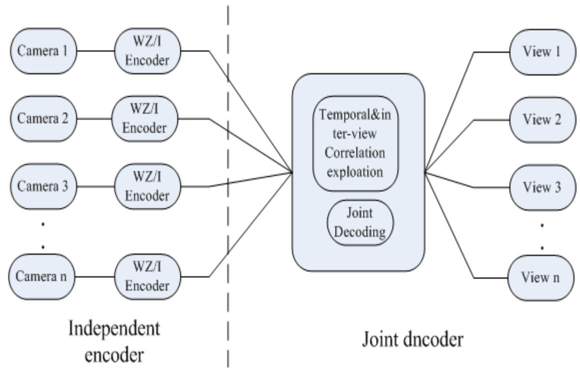


Fig. 1 Framework of DMVC

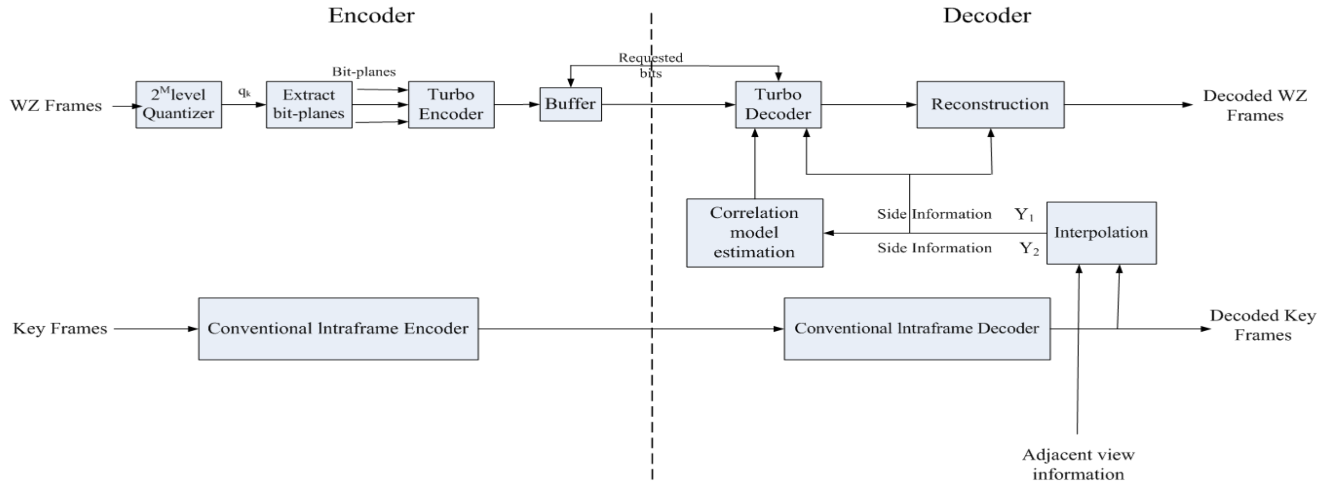


Fig.2 Block diagram of MDVC system

II. THE MDVC FRAMEWORK

A. MDVC framework

The framework of DMVC is illustrated in Fig. 1. Each view sequence is encoded independently without communicating with each other. In each view sequence, each frame is also encoded independently. GOP size 2 is adopted for each view sequence. Even frames are coded as Key frame, and odd frames are coded as WZ frame. The block diagram representation of the MDVC system is shown in Fig. 2. At encoder, key frames are coded using traditional DCT-based intra coding method. 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. The turbo codec consists of two identical constituent convolution codecs and it uses a generator matrix to generate parity bits. Generated parity bits are stored in a buffer. At decoder, key frames are decoded independently. And their reconstruction frames are used to generate side information for WZ frames. In our scheme, two kinds side information are generated for WZ frame. As shown in Fig. 2, side information Y_1 for current WZ frame is generated by carrying out temporal interpolation (TI) on its adjacent forward and backward key frames and Y_2 is generated using disparity-guided temporal interpolation (DGTI), which works with the help of adjacent view information. The detail techniques of TI and DGTI will be presented in the following subsections. After Y_1 and Y_2 are generated, they are used to estimate the correlation noises model. The decoder consists of two soft-input soft-output (SISO) decoders, which are implemented using the Maximum A-Posteriori (MAP) algorithm. The decoder, which recovers the original frame based on side information, will successively request parity bits from the encoder until a pre-defined threshold (10^{-3} in our MDVC system) of bit error rate is satisfied.

B. Temporal interpolation

Temporal interpolation (TI) is a widely used side information generation method in mono-view DVC. To estimate motion vectors without the knowledge of original frame, a general assumption of a linear motion model in video sequence is made in TI. TI involves symmetrical bidirectional motion compensated prediction. As shown in Fig. 3, current WZ frame at time index *n* is denoted by W_n, and its backward and forward key frames are termed I_{n-1} and I_{n+1}, respectively, and b_i denotes the block in W_n.When applying the linear motion model, the backward motion vector of b_i is obtained by scaling motion vector of its co-located block I_{n+1} with the equation $MV_b = (MV_{n+1})/2$, where MV_{n+1} is obtained by performing motion estimation between I_{n+1} and I_{n-1}. Similarly, the forward motion vector of b_i is obtained by scaling motion vector of its co-located block in I_{n-1} with the equation $MV_f = (MV_{n-1})/2$. Side information for b_i is yielded by averaging backward prediction and forward prediction. This progress is performed on all blocks so that side information for whole frame is obtained.

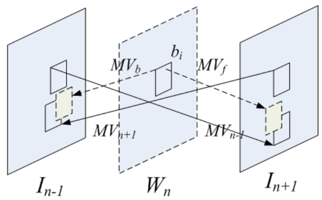


Fig. 3 TI illustration

Generally speaking, TI can work efficiently in the regions with low motion, such background, since linear motion model is usually satisfied in these regions. However, in the rest regions, especially the regions with high motion, it is hard to get true motion by TI.

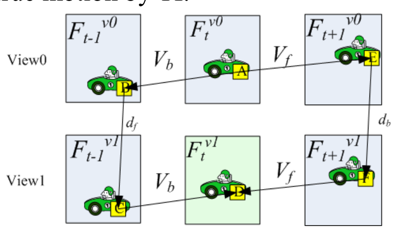


Fig. 4 DGTI illustration

C. Disparity-guided temporal interpolation

Disparity-guided temporal interpolation (DGTI) is an approach to generate side information using adjacent views' information. The basic idea of this scheme is depicted in Fig. 4. This generation method is composed of three steps: firstly, disparity compensation operation is performed on two key frames, F_{t-1}^{V0} and F_{t-1}^{V1}. The obtained disparity vectors help the blocks in F_{t-1}^{V1} find their most similar blocks in F_{t-1}^{V0}; secondly, motion vectors are estimated for the blocks in F_{t-1}^{V0}. Notably, F_t^{V0} is a reconstructed frame so that motion vectors between F_{t-1}^{V0} and F_t^{V0} can be estimated; thirdly, the blocks in F_{t-1}^{V1} learn the motion vectors of their corresponding blocks in F_{t-1}^{V0}, and these motion vectors are applied to generate side information for F_t^{V1}. This works when the two cameras lie in same plane and point to the same direction. In this method, inter-view information is used by means of motion correlation. For F_t^{V1}, the motion vectors derived from F_{t-1}^{V0} and F_t^{V0} are probably more accurate than those derived from F_{t-1}^{V1} and F_{t+1}^{V1} for the reason that the temporal correlation between F_{t-1}^{V0} and F_t^{V0} is much stronger than that between F_{t-1}^{V1} and F_{t+1}^{V1}. The above progress is called backward interpolation. Similarly to TI, DGTI involves symmetrical bidirectional interpolation. Accordingly, forward interpolation is performed on F_{t+1}^{V0}, F_t^{V0} and F_{t+1}^{V1}. And side information is yielded by averaging forward interpolation and backward interpolation.

III. CORRELATION MODEL

Side information frame is viewed as "corrupted" original frame, and it provides the input information for the SISO decoders, the core of turbo decoder in our MDVC system. The more accurate the input information is, the fewer bits the decoder will spend on correcting the same amount errors. Therefore, the statistical relationship between side information frame and original frame is very important.

Traditionally, the statistical relationship between side information frame and original frame is modeled with Laplacian distribution at frame level. This ignores the fact that side information quality is not constant within the frame. The errors in high motion region are usually much more than those in low/zero motion region.

In this paper, we propose a novel model to characterize the relationship between side information frame and original frame. In order to describe our algorithm conveniently, we call the pixels true pixels (TPs) that are correctly generated in side information, and the others false pixels (FPs). Motivated by the fact that the regions where side information generation is successful and the region where side information generation has failed co-exist, we propose to divide the frame into N sets depending on the ratio of TPs. That is, the ratio of TP in Set_i is larger that that of Set_j, if and only if $i > j$. Note that every set is composed of pixels instead of blocks in the proposed model. The reason is that we realize that certain amount pixels in one block can still be successfully generated even though this block locates in high motion region. Similarly, certain amount pixels in one block may fail to be generated although the block locates in low motion region. That is to say, TPs and FPs co-exist in a block no matter whether the block locates in high motion region or low motion region.

The proposed correlation model is presented in (1).

$$f(WZ_{(x,y)}, SI_{(x,y)}) = \frac{\alpha_i}{2} e^{\alpha_i |WZ_{(x,y)} - SI_{(x,y)}|}, if\ (x,y) \in Set_i \qquad (1)$$

Each set is characterized with a Laplacian distribution. Each parameter of them is estimated independently. The segmentation criterion of each set is shown in (2).

$$(y,x) \in Set_i, if\ BP(TISI(y,x), i) == BP(DGTISI(y,x), i)\ \ i \in [0,8] \qquad (2)$$

in which $TISI(y,x)$ and $DGTISI(y,x)$ indicate the value of pixel that locates at coordinate (y,x) in side information frame generated by TI and by DGTI, respectively. $BP(value, i)$ is a function that returns top i bit-planes of *value*. Notably, we define that if a pixel belongs to Set_i, it does not belong to Set_j for $j < i$ anymore.

In order to validate the efficiency of the proposed correlation model, we assume that the original frame, TISI (side information generated using TI) and DGTISI (side information generated using DGTI) are simultaneously available so that the parameters of each set can be estimated ideally. This is called an offline estimating method and it can give us insight of maximum or "ideal" performance that the proposed correlation model can achieve. We apply the offline estimating method described in [5] on estimating parameter for each set in proposed model. The calculation of parameter for each set can be formalized as (3):

$$\alpha_i^2 = \frac{2}{\sigma_i^2} \quad (3)$$

where σ_i^2 is the variance of the residual of pixels in Set_i between original frame and side information frame.

To make the proposed correlation model more practical, we propose an approach to estimate σ_i^2 at decoder (online). In online estimating, we regard $F(x,y) = \frac{1}{2}(F_b(y,x) + F_f(y,x) + 1)$ as original frame to estimate α_i^2, where $F_b(y,x)$ and $F_f(y,x)$ denote the reconstructed backward and forward key frames, respectively. After the parameters of each set are estimated, the condition probability is calculated depending on which set the pixel belongs to during turbo decoding. This is shown as (4):

$$p(x \mid y) = \frac{\alpha_i}{2} e^{\alpha_i |y-x|}, if \ x \in Set_i \quad (4)$$

IV. EXPERIMNET RESULTS

The efficiency of the proposed correlation model and the correlation model estimating algorithm are evaluated in term of overall RD performance of WZ frames. If the proposed correlation model can more accurately reflect the relationship between original frame and side information frame, promising coding gains will be observed. Besides, if the correlation model estimating algorithm is efficient, the coding performance achieved will be comparable to that estimated offline.

Experiments are carried out on multi-view video sequences *Ballroom* and *Race1*. Both *Ballroom* and *Race1* are captured by 1-D parallel cameras in resolution of 640x480. The frame rate for *Ballroom* and *Race1* 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. 1/4 pixel spatial interpolation in reference frame is also employed to improve motion/disparity accuracy. Key frames are about 38dB. The authors stated that the algorithms they proposed in [5] can achieve performance that is equivalent to traditional frame level model with parameter estimated offline. Hence, three techniques, traditional frame level correlation model with parameter estimated offline, proposed model with parameters estimated offline and proposed model with parameters estimated online, are compared in our experiments. Because improving side information quality using a fusion method is beyond the topic of this paper, we apply the three techniques on TISI and DGTISI instead of a fused side information of them. And PSNR and bitsrate of luminance component of WZ frames are evaluated. Simulation results are presented in Fig.5. For both sequences, compared to traditional frame-level correlation model with parameter estimated offline, promising coding gain is obtained when applying proposed correlation model with parameters estimated offline on either DGTISI or TISI. This demonstrates that the proposed model can more accurately describe the relationship between original frame and side information frame. Besides, it is can be observed that comparable coding performance can be achieved when using proposed correlation model with parameters estimated online. This verifies that the proposed parameter estimating algorithm is also efficient.

V. CONCLUSIONS

This paper proposes a novel correlation noises model to describe the relationship between original frame and side information frame. In the proposed correlation noises model, WZ frame is segmented into several sets according to side information quality. Besides, an algorithm to estimate the proposed model using adjacent view information at decoder (online) is also proposed. Experiment results show the coding performance gain can be observed when applying the proposed correlation model. And the performance achieved using the proposed online estimation algorithm is closed to that achieved by using the offline estimation method.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation (60672088 and 60736043) and Major State Basic Research Development 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.

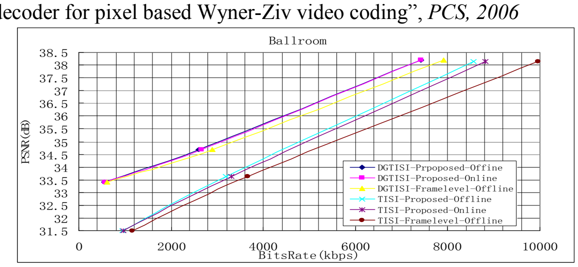
[2] A.D. Wyner and J. Ziv, “The rate-distorition function for source coding with side information at the decoder”, *IEEE Transaction on Information Theory*, vol. 22, pp.1-10, Jan. 1976.

[3] 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*.

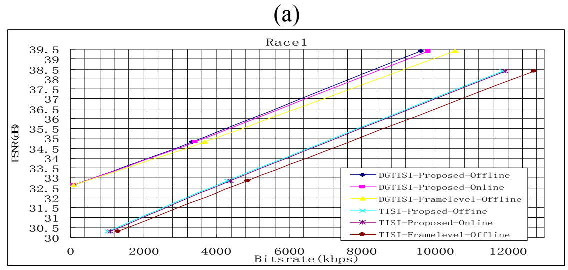
[4] B.Girod, A. Aaron, S.Rane, D. Rebollo-Monedero, “Distributed video coding”, Proceeding of the IEEE, vol. 93, January 2005

[5] 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*

[6] C. Birtes, J. Ascenso, F. Pereira, “Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding”, *PCS, 2006*



(a)



(b)

Fig. 5 Simulation results for *Ballroom* and *Race1*