TWO-PASS RECONSTRUCTION IN DISTRIBUTED VIDEO CODING

Hongbin Liu[1], Yongpeng Li[2], Xianming Liu[1], Siwei Ma[3], Debin Zhao[1], Wen Gao[3]
{hbliu, ypli, xmliu, swma, dbzhao, wgao}@jdl.ac.cn

1Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R. China
2Graduate University, Chinese Academy of Science, Beijing, 100039, P.R. China
3Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, 100871, P.R. China

ABSTRACT

In this paper, we propose a novel two-pass reconstruction algorithm for the Wyner-Ziv (WZ) frames in distributed video coding (DVC), in which the traditional reconstructed WZ frame is utilized to perform motion estimation to obtain a more accurate motion field. During the motion estimation, the block, as well as its neighboring pixels are concerned. An overlapped block motion compensation is subsequently performed with the help of the motion field, consequently, an enhanced prediction for the WZ frame can be obtained, based on which an improved reconstruction can be achieved. Simulation results show that both the objective and subjective quality of WZ frames can be improved significantly.

Index Terms— distributed video coding, two-pass reconstruction

1. INTRODUCTION

Current video compression standards like MPEG-x and H.26x perform computationally intensive motion field estimation including motion estimation and mode decision for inter-pictures coding at encoder to efficiently exploit the temporal correlation. As a result, the encoder is much more complex than the decoder. This is reasonable for broadcasting or for the systems where videos are compressed once and decoded many times. However, in scenarios like the sensor network and the mobile communication where the encoder is not so powerful, the dual complexity allocation may be required. Distributed Source Coding (DSC) provides an opportunity to afford applications with simple encoder.

The Slepian-Wolf theorem [1] states that: assume X and Y are two statistically dependent discrete signals, when they are encoded independently and decoded jointly, the achievable rate region for probability of decoding error to approach zero is still $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$ and $R_X + R_Y \geq H(X,Y)$. On the other hand, Wyner-Ziv theorem [2] can further extend DSC to the lossy coding. Assume X and Y are two statistically dependent Gaussian random processes, the conditional Rate-Mean Squared Error Distortion function for X will be unchanged no matter Y is known only at decoder, or both at encoder and decoder. And these theorems demonstrate that DVC is feasible.

Recently, some practical DVC systems are presented. S. S. Pradhan and K.Ramchandran propose the DISCUS [3] architecture which uses syndrome-based encoding to perform WZ video coding. R.Puri and K.Ramchandran propose another DVC framework which is described as PRISM in [4][5]. A. Aaron and B. Girod also provide a asymmetric system with low-complexity encoder [6], in their scheme, H.263+ intra frame mode is used to encode the odd frames which are known as key frames and a Slepian-Wolf codec based on turbo coding is used to encode the left frames which are known as WZ frames.

In distributed video coding (DVC), a dual complexity allocation, where encoder is relatively simple and decoder is complicated, is provided. Although it seems to be a satisfactory method to solve problems in the scenarios like mobile wireless cameras where encoder is power-constrained, and is naturally robust against the channel errors, it degrades compression performance considerably when compared to H.264/AVC standard. One major reason is that neither the temporal correlation nor the spatial correlation can be exploited sufficiently at decoder in

DVC. This paper proposes a novel two-pass reconstruction algorithm to further exploit both the spatial and the temporal correlation, which provides a more promising compression performance while maintaining the complexity and bit-rate unchanged at the encoder.

This paper is organized as follows: Section 2 describes the proposed algorithm in detail. Section 3 shows simulations results and performance comparison. Conclusions and future work are presented in Section 4.

2. OVERVIEW OF THE TRADITIONAL DVC CODEC

The diagram of traditional DVC codec [6] is illustrated in Fig.1. The overall coding architecture is described as follows : a video sequence is classified into key frames and WZ frames, with key frames encoded in H.264 intra method and WZ frames encoded in DVC method.

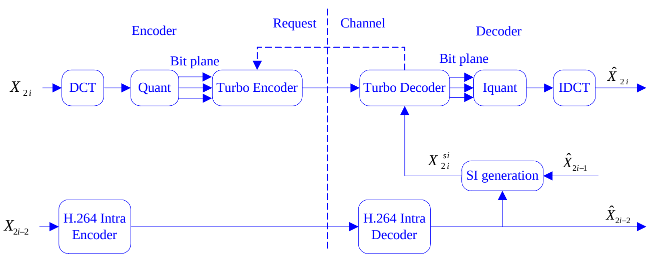


Fig.1. Block diagram of the traditional low delay DVC

At encoder, for each WZ frame, it is firstly applied a block based DCT transform; then, the DCT coefficients of the entire frame are grouped together to form DCT coefficient bands, according to the position in the DCT transformed block. After this, each coefficient band is uniformly quantized according to the quant parameters. Finally, coefficient bands are encoded from most significant band to least significant band, i.e., form low frequency to high frequency. When encode a given band, bitplanes of it are extracted and encoded from most significant bitplane to least significant bitplane successively. To encode extracted bitplanes, turbo encoder, in which the generated parity bits are successively sent to decoder until correct decoding is achieved, is adopted.

At decoder, firstly, motion interpolation method, in which the translational motion model is assumed, is employed to generate SI of the current WZ frame, with the help of the nearest two decoded key frames. Secondly, perform DCT, organization, and quantization on SI as the encoder does. Then, decode coefficient bands and bitplanes following the same order of encoding. To decode a given bitplane, the corresponding bitplane of SI and the parity bits transmitted from encoder are fed together into the turbo decoder, the decoder will request more parity bits from encoder through the request channel until it can decode the bitplane correctly. During the turbo decoding, the statistics of the residue between WZ frame and SI is assumed to be modeled by a Laplace distribution. For each coefficient band, after decoding all of the bitplanes of it, it can be reconstructed with the help of SI. After reconstruct all of the coefficient bands, they will be reorganized to form DCT coefficient blocks, subsequently, IDCT will be performed on all of the DCT coefficient blocks to reconstruct the WZ frame.

3. PROPOSED TWO-PASS RECONSTRUCTION ARCHITECTURE

It can be observed that when large motion exists, subjective quality of the reconstructed WZ frame in DVC can be very annoying though the objective quality is very well. This mainly attributes to inaccurate SI. To tackle this problem, we use the traditional reconstructed WZ frame to achieve a more accurate motion field, a better prediction as well as an improved reconstruction can thus be obtained. Based on this idea, a two pass reconstruction algorithm is designed.

Fig. 2 shows the architecture of the improved DVC codec, with the proposed algorithm highlighted in color parts. As can be seen in Fig.2, the encoder works same with the traditional DVC encoder, the proposed algorithm is performed at decoder only. Details of the proposed algorithm can be described as follows: firstly, the proposed refined codec decodes and reconstructs the WZ frame as the traditional DVC decoder does, here, denoting the reconstruction generated here as the first-pass reconstruction (FPR). Then, the FPR is utilized to obtain the true motion field of the current WZ frame, which is used to compensate an improved prediction. Finally, the stored decoded bit-planes of current WZ frame together with the improved SI are fed into the Iquant module, after Iquant and idct, a second-pass reconstruction can be obtained.

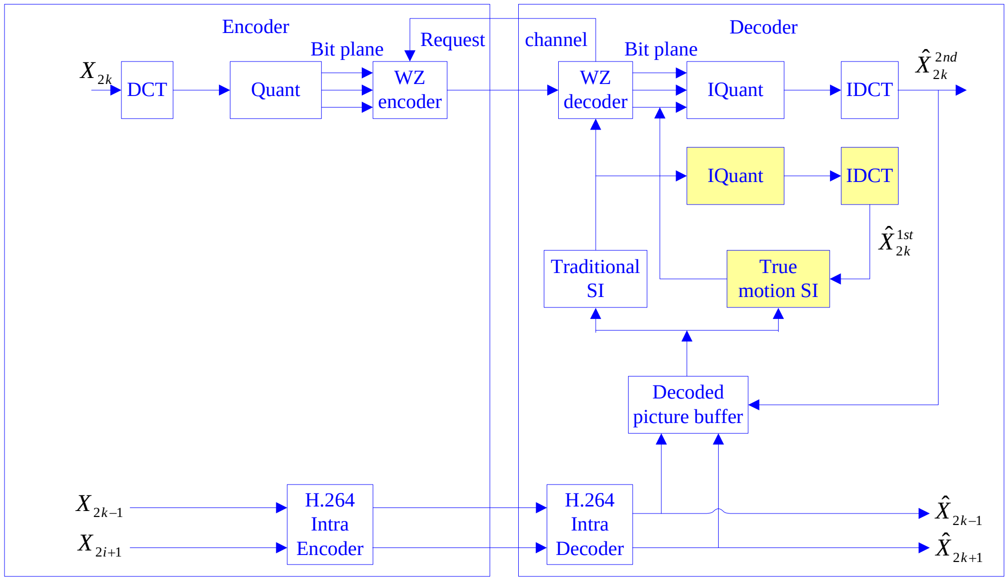


Fig.2. Block diagram of the proposed architecture

True motion of the WZ frame is generated as follows. During the estimation of the true motion vector (TMV), to avoid noise, the block, as well as its neighboring pixels which are shown in Fig.3, are used together to perform motion estimation.

The MV (dx, dy) which minimizes the cost function defined in (1) is chosen as TMV.

$$\text{cost}(x,y) = \sum_{(i,j)\in B(x,y)\cup N(x,y)} (FPR(i,j) - ref(i+dx, j+dy))^2 \quad (1)$$

Here, (i, j) and (x, y) are coordinates of the pixels, dx and dy are horizontal and vertical components of MV respectively, and $B(x, y)$ denotes current block with (x, y) as its top-left pixel, $N(x, y)$ denotes neighboring pixels of block $B(x, y)$. FPR represents the reconstructed WZ frame generated in the first-pass, and ref represents the reference frames decoded before current frame. $T(i, j)$ is the value of the pixel located at (i, j) in frame *T*.

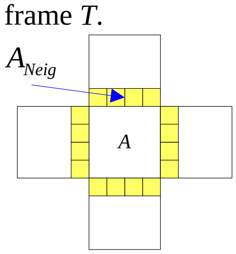


Fig.3. Neighboring pixels of one block A

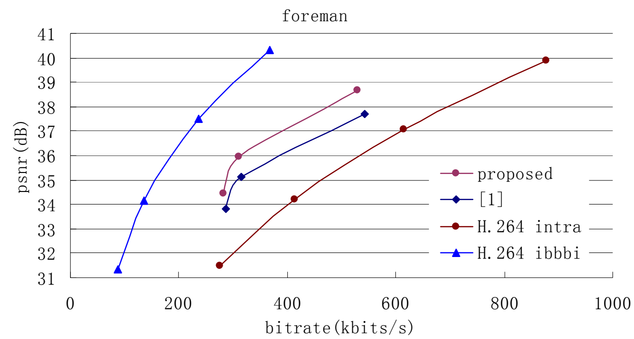
After obtaining TMV of each block, OBMC [8] is performed block by block to further improve the quality of SI.

4. SIMULATION RESULTS

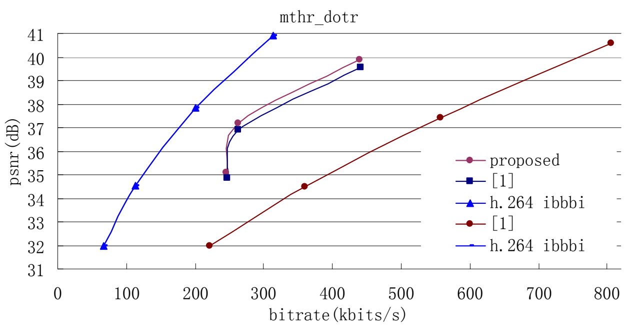
In the simulation, the transform domain turbo based DVC codec is adopted, and GOP (group of pictures) length is set to 4. The first frame in each GOP is H.264 intra coded, and other frames are WZ coded. Motion compensated interpolation (MCI) [7] is employed to generate initial SI for each WZ frame in the first-pass reconstruction. The block size in the true motion estimation process is set to 8x8.

Several video sequences are selected to evaluate the performance of the proposed algorithm, and the objective results of foreman and mthr_dotr qcif@30Hz are shown in Fig 4. As can be observed, up to 1 dB gain can be achieved, and the proposed algorithm performs well from low bit-rate to high bit-rate for *foreman*; while only about 0.3dB gain can be achieved for *mthr_dotr*, this is because the initial SI for most WZ frames are good enough, which decreases the effect of the proposed algorithm. Besides, the subjective quality of WZ frames with poor initial SI can be improved significantly for both *foreman* and *mthr_dotr*, which is shown in Fig.5. It can be seen that, the face part in the 2nd frame of *foreman*, where large motion exits, can be very annoying when reconstructed using algorithm of [7], however, this can be gracefully improved with the proposed algorithm. It also can be observed that, when reconstruct the 14th frame of *mthr_dotr* by [7], the hand part in it looks like fragment, and the right eye part of the little girl is lost, however, when it is reconstructed by the proposed algorithm, the hand part can be very smooth, and the right eye part of the little boy comes out.

It should be noted that further improvement can be achieved through more than two pass reconstruction, however, it is limited.



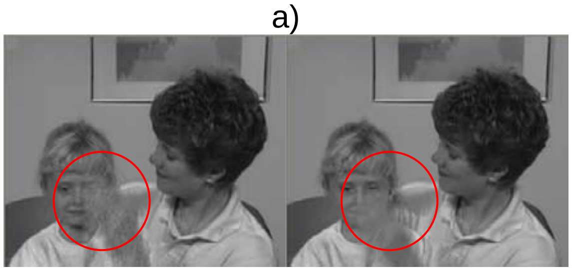
a)



b)

Fig.4 Rate-distortion curve for a) *foreman* qicf@30Hz, b) *mthr_dotr* qicf@30Hz

a)



b)

Figure 5. Subjective quality comparison of left : [6], right: proposed. a) 2^{nd} reconstructed frame of *foreman*, b) 14^{th} reconstructed frame of *mthr_dotr*.

5. CONCLUSIONS AND FUTURE WORK

We propose an efficient two-pass reconstruction algorithm for DVC which can further exploit both temporal and spatial correlation. Simulations results show that the proposed method works well on sequences with high motions where initial SI are usually poor at the first-pass reconstruction, though the effect relatively degrades on silent sequences. Another advantage of the proposed method is that no extra complexity or bit-rates spending is required at the encoder.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” in *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[2] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” in *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[3] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction”, in *Proc. IEEE Data Compression Conference*, 1999.

[4] R. Puri and K. Ramchandran, “PRISM: a new robust video coding architecture based on distributed compression principles,” *Allerton Conference on Communication, Control and Computing*, 2002.

[5] R. Puri and K. Ramchandran, “PRISM: a new robust video coding architecture based on distributed compression principles,” Allerton Conference on Communication, Control and Computing, 2002.

[6] A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video,” in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, Nov. 2002.

[7] A. Aaron, S. Rane, E. Setton, and B. Girod, “Tansform-domain Wyner-Ziv codec for video,” in Proc. SPIE Visual Communications and Image Processing, San Jose, CA, Jan. 2004.

[8] Michael T. Orchard and Gary J. Sullivan , “Overlapped Block Motion Compensation: An Estimation-Theoretic Approach”, IEEE Trans on Image Processing, 1994.