Chinese Journal of Electronics
Vol.18, No.1, Jan. 2009

# Residual Distributed Video Coding Based on LQR-Hash*

WANG Anhong[1,2], ZHAO Yao[1] and PAN Jeng-Shyang[3]

(1.*Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*)

(2.*Taiyuan University of Science and Technology, Taiyuan 030024, China*)

(3.*Kaohsiung University of Applied Sciences, Kaohsiung 807, Taiwan*)

**Abstract — Distributed source coding (DSC) and Distributed video coding (DVC) have aroused high research interest due to their property of low-complexity encoding. This paper presents a residual DVC based on Low quality reference (LQR) hash compressed by 0-motion H.264/AVC. SW-SPIHT is applied to the residual between Wyner-Ziv frame and the decoded LQR hash, which exploits the temporal and spatial correlations so improves the coding efficiency. At the decoder, the decoded LQR is used for more accurate motion estimation and better side information. And a refinement reconstruction with side information is proposed for better SPIHT decoding. The experimental results indicate the proposed scheme achieves better rate-distortion performance than current literatures under the constraint of similar encoding complexity.**

**Key words — Distributed video coding, Low quality reference (LQR) hash, Residual coding, Slepian-Wolf-SPIHT (SW-SPIHT).**

## I. Introduction and Our Motivation

In traditional video coding system, such as standard H.26x algorithms, it is the encoder that performs inter-frame prediction to exploit the correlation among successive frames. Since predictive coding makes use of motion estimation, the complexity of the encoder is typically 5 to 10 times more than that of the decoder. This case is desirable for the applications where video is compressed once and decoded many times, as in broadcasting. However, some new friendly uplink communication applications including mobile camera phones and sensor network cameras may require low-complexity encoders. DSC and DVC provide theory foundations for easy encoding, thus they can meet the demand of these applications.

DSC's foundation is based on two information theories, Slepian-Wolf theory[1] for lossless cases and Wyner-Ziv[2] for the lossy ones. These theory state that the compression of two separated source can be made as efficient as if they are compressed together as long as joint decoding is done at the receiver. Wyner-Ziv's source coding with Side information (SI) accessible at the decoder is the counterpart of these theories, *i.e.*, assuming $X$ and $Y$ are statistically dependent Gaussian sources and $Y$ is SI, the Rate-Mean Squared Error Distortion function of $X$ is the same whether $Y$ is known only at the decoder, or both at the encoder and decoder. DSC suggests a source coding framework which exploits correlation of sources only at the decoder so the encoder can be simplified.

DVC is DSC's application in video[3−8]. Generally in DVC system, for Wyner-Ziv frames, the quantized pixels or transform coefficients $X$ is channel encoded and only the parity bits $P$ are transmitted. At the decoder, the error-correcting decoding using the parity bits $P$ and SI $Y$ is implemented to recover $X$. The reconstruction is also to recover the original pixels or transform coefficients with the help of SI.

As an estimation to the main source, SI $Y$ is very important in DSC and DVC. Generally, the more dependent $Y$ is, the less parity bits and the higher compression as well as the better reconstruction can be achieved. Among some methods, hash-based SI is the best attractive method[5]. The main idea is, except the Wyner-Ziv bits, the encoder also sends some representing information on encoding to assist the motion estimation at the decoder. Because the hash bits take more information about current frame, the hash-based motion estimation is more accurate than that without hash. While, hash-based DVC is enhanced at the cost of excessive hash bits. Therefore the hash should be compressed efficiently. Specially, in Ref.[5], the compression of hash bits is based on the entropy code so the hash bit rate is $R(h) \geq H(h)$. Ref.[8] proposes to use the LQR frame to generate efficient hash. The LQR hash refers to the low quality version of the current Wyner-Ziv frame compressed by prediction coding similar to DPCM/DCT with 0-motion-searching and coarse quantization. Just because the temporal dependence is exploited, the bit rate of LQR hash approaches ideally to conditional entropy $R(h) \geq H(h|Y)$. But, LQR-based method of Ref.[8] has apparent disadvantage. For example, LQR hash bits and Wyner-Ziv bits are two different streams for the same frame. So there is redundancy between LQR and Wyner-Ziv bits especially at high bit rate.

*Manuscript Received Nov. 2007; Accepted June 2008. This work is supported in part by the National Natural Science Foundation of China (No. 90604032, No. 60373028), Program for New Century Excellent Talents in University, 973 Program (No.2006CB303104) and Youth Science Technology Research Foundation of Shanxi Province (No. 2008021020).

On the other hand, it has shown that residual DVC can give better performance than pure pixel DVC[9]. Residual DVC means Wyner-Ziv encoding the residual $D = W - W_{re}$, where $W_{re}$ is a simple estimation to $W$ and accessible both at the encoder and decoder. $D_y = Y - W_{re}$ acts as SI in Wyner-Ziv decoding. Finally, $W_{re}$ is added to obtain the recovered $W'$. Because there are correlations between $W_{re}$ and $W$, residual DVC brings improvement to pure DVC.

Motivated by the above analysis, in this paper, some improvements will be given to the residual DVC based on LQR hash. We firstly propose to use the decoded LQR hash as $W_{re}$ for residual DVC. The decoded LQR is just the coarse version of $W$ so it is dependent with $W$. Also, SW-SPIHT is used to the residual $D$, which can further decrease the spatial redundancy in residual frame. Our contributions mainly include: (1) combining the residual DVC with LQR hash to improve the rate-distortion performance; (2) exploiting SW-SPIHT to the residual; (3) refining wavelet coefficients by SI.

This paper is arranged as follows, Section II proposes the residual DVC based on LQR hash in detail, in Section III, the experimental results are given, and finally, Section IV concludes the paper.

# II. The Proposed Residual DVC Based on LQR Hash

The proposed scheme is in Fig. 1. To key frames, H.264/AVC intra-encoding and decoding is implemented. While to Wyner-Ziv frames, the low complexity encoding and high complexity decoding are implemented.

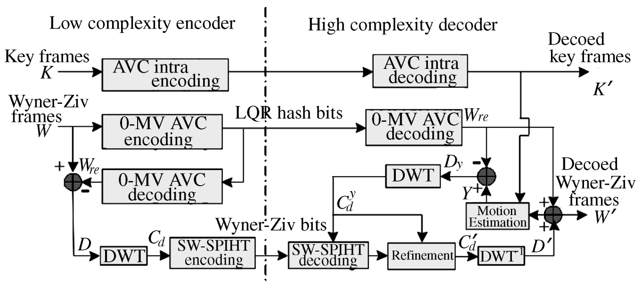


Fig. 1. Proposed residual DVC based on LQR

**1. Low-complexity encoding**

The encoder generates two kinds of bit stream for a Wyner-Ziv frame $W$; one is LQR hash, *i.e.*, the bit stream from 0-motion H.264/AVC with coarse quantization. The other, the Wyner-Ziv bit, is from the residual DVC. That is, the residual $D = W - W_{re}$ is decomposed by Discrete wavelet transform (DWT). Then, the coefficient $C_d$ is encoded by SW-SPIHT encoding[6] and the outputted bit forms the so-called Wyner-Ziv bit stream.

Note that, LQR encoding does not use any computational motion estimation. And, SW-SPIHT encoding consists of DWT with SPIHT and channel encoding, which means the complexity is similar to that of conventional DCT-domain DVC. These mentioned bring low-complexity encoding to the whole system.

**2. High complexity decoding**

At the decoder, the decoded LQR is first used to do motion estimation to generate $Y$, which is explained in the Subsection II.4. Then, the difference $D_y = Y - W_{re}$ is decomposed by a DWT similar to the encoder. The wavelet coefficient $C_d^y$ of $D_y$ is used as SI for SW-SPIHT decoding and the refinement process, consequently the decoded $C_d'$ is achieved. Next, the inversed DWT is implemented and the difference pixel $D'$ is obtained. Finally, let $W' = W_{re} + D'$, then Wyner-Ziv frame is reconstructed.

The computational motion estimation and iterated channel decoding arouse high-complexity decoding though they do not influence the easy encoding.

**3. Refinement to the wavelet coefficient**

After SW-SPIHT decoding, SPIHT stream is recovered completely, the wavelet coefficients are refined based on the formula

$$C_d' = \begin{cases} v_{\max}, & \text{if } C_d^y \geq v_{\max} \\ C_d^y, & \text{if } C_d^y \in (v_{\min}, v_{\max}) \\ v_{\min}, & \text{if } C_d^y \leq v_{\min} \end{cases} \quad (1)$$

where $C_d'$ is the final wavelet coefficient value after refinement, $v_{\max}$ and $v_{\min}$ are the possible maximal and minimal value of $C_d$ if SPIHT decoding is implemented to all Bit planes (BP), on the assumption that the total number of BP is $m$, $m > n$, $n$ is the number of recovered BP. There are two reasons that this refinement gives improvement, one is that the recovered $n$ BPs take some information of $C_d$, which will correct the noisy BPs of $C_d^y$ and limit the distortion from $C_d^y$, the other is that some additional information from $m - n$ BPs of $C_d^y$ is added to the recovered $n$ BPs, which compensates the reconstruction of $C_d$ from $n$ BPs.

**4. Motion compensation based on LQR hash**

The motion compensation of LQR hash consists of two steps. Firstly, motion estimation is implemented to find the best matched block and the corresponding motion vector. Specially, let $L_t(i, j)$ denote the pixel value at row $i$ and column $j$ of the decoded LQR in time $t$ and let $R_t(i, j)$ denote the corresponding pixel in the referenced frame. Then the motion vector for the block at $(a, b)$ is

$$MAD(a, b) = \arg\min_{dx,dy} \sum_{j=0}^{7} \sum_{i=0}^{7} |L_t(a+i, b+j) - R_t(a+i+dx, b+j+dy)| \quad (2)$$

where $dx \in [-8, 8]$ and $dy \in [-8, 8]$, and the decoded previous frame is used as the referenced frame. Secondly, we use the weighted average value of $block_{LQR}$ at $(a, b)$ of LQR and its matched block $block_{matched}$ to generate the compensated block $block_{comp}$. That is,

$$block_{comp} = \alpha \times block_{matched} + (1 - \alpha) \times block_{LQR} \quad (3)$$

where, $\alpha$ is an experimental value in $[0, 1]$.

# III. Experimental Results

100 luminance frames of 15Hz QCIF sequences foreman and hall are tested to evaluate the performance of DVC codec proposed. Meanwhile, the optimized combination quantization for constant decoded frames is considered in Table 1,

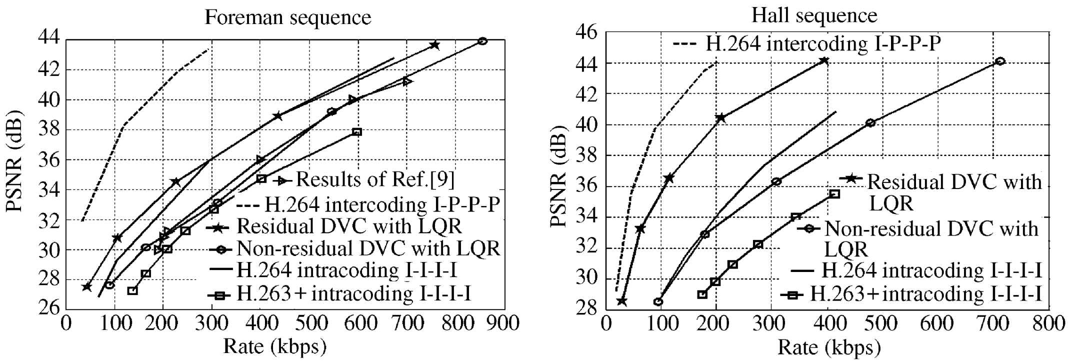


Fig. 2. Rate-distortion curve comparison (GOP=8)

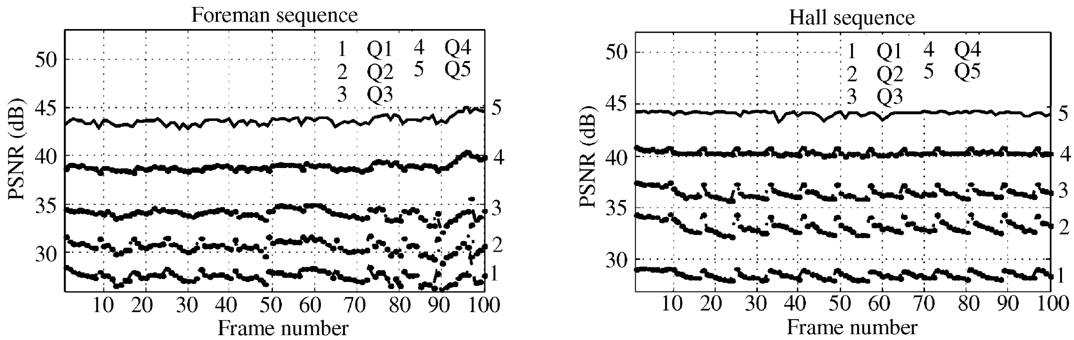


Fig.3. Temporal evolution curve for all distortion points (GOP=8)

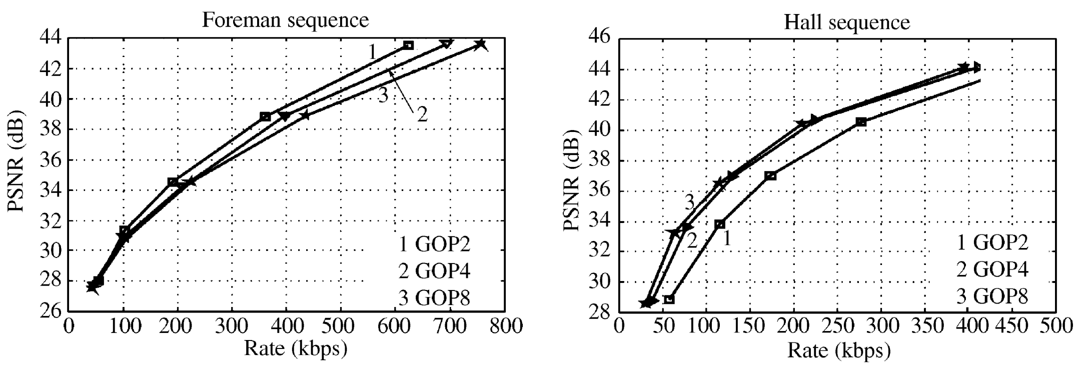


Fig. 4. Rate-distortion comparison of different GOP size

where, QP value means the quantization chosen in JM9.0 for key frames and the quantization of Wyner-Ziv frame is chosen by the number of BPs. Five rate-distortion points Q1, Q2, Q3, Q4, and Q5 is shown.

Firstly, the PSNR of the reconstructed video with GOP 8 is in Fig.2. The experiments also include the non-residual DVC, which just moves the subtraction operation from Fig.1, and the residual DVC in Ref.[9].

Compared with non- residual DVC, the proposed DVC gives 1.2–5.8dB improvement. This improvement originates from the fact that residual scheme can make use of second SI at the encoder[9]. When compared with the best results of Ref.[9], our scheme obtains 1.2–3dB improvements for foreman sequence. This is because of the efficient hash and SW-SPIHT which exploit the temporal and spatial correlation in residual frames. In addition, compared with intra-coding of H.264/AVC, the proposed scheme has up to 6dB enhancement. However, there is still a gap from the state-of-the-art inter-coding of H.264/AVC.

The temporal evolution curve is in Fig.3. Constancy is acceptable for both high and low motion sequences due to the optimized combination quantization.

Secondly, GOP size is changed in Fig.4. The bigger GOP does improve the rate-distortion when motion is low, as for hall sequence. However, for high motion sequence foreman, the rate-distortion curve is degressive with the GOP size. This is because that the SI becomes bad especially for high motion sequences.

**Table 1. Combined quantization in residual DVC**

| | foreman | | | | | hall | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 |
| QP | 41 | 36 | 32 | 26 | 20 | 40 | 33 | 29 | 24 | 19 |
| BP | 2 | 3 | 4 | 5 | 6 | 2 | 3 | 4 | 5 | 6 |

Finally, the encoding time is measured in second. The hardware used is an ×86 machine Pentium (R) 4 processor at 2.5GHz with 256MB of RAM. The software condition is Windows XP operative system with the MATLAB version 6.5. The average encoding times of each frame for hall GOP=8 are 1.1025, 1.30, 1.41, 1.61 and 1.86 seconds for Q1, Q2, Q3, Q4 and Q5 respectively, while the corresponding encoding time of DCT domain DVC is 1.10, 1.30, 1.39, 1.59 and 1.7 seconds for the same quality; *i.e.*, the similar encoding complexity is achieved. The similar conclusions are also observed for foreman sequence.

## IV. Conclusions

In this paper, a novel residual DVC based on LQR is presented. This scheme benefits from the residual coding and the efficient LQR-hash-based motion compensation. It SW-SPIHT encodes the residual between Wyner-Ziv frame and a decoded LQR. The decoder recovers Wyner-Ziv frame using LQR. Experimental results show the proposed scheme has comparable rate-distortion performance and similar encoding complexity.

## References

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlatemation sources", *IEEE Transaction on Information Theory,* Vol.19, pp.471–480, 1973.

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder", *IEEE Transaction on Information Theory,* Vol.22, No.1, pp.1–10 , 1976.

[3] R. Puri and K. Ramchandran, "PRISM: an uplink-friendly multimedia coding paradigm", *Proc. of International Conference on Acoustics, Speech, and Signal Processing,* Hong Kong, pp.856–859, 2003.

[4] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video", *Proc. of Visual Communications and Image Processing,* California, USA, pp.120–123, 2004.

[5] A. Aaron, S. Rane and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver", *Proc. of IEEE International Conference on Image Processing,* Singapore, pp.3097–3100, 2004.

[6] A. Wang, Y. Zhao and J.S. Pan, "Multiple description image coding based on DSC and pixel interleaving", *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* Harbin, China, pp.1356–1359, 2008.

[7] A. Wang, Y. Zhao and L. Wei, "Wavelet-domain distributed video coding with motion-compensation refinement", *Proc. of IEEE International Conference of Image Processing,* Atlanta, USA, pp.241–244, 2006.

[8] E. Martinian, A. Vetro, J. Ascenso, A. Khisti and D. Malioutov, "Hybrid distributed video coding using SCA codes", *Proc. of IEEE 8th Workshop on Multimedia Signal Processing,* Canada, pp.258–261, 2006.

[9] A. Aaron, D. Varodayan and B. Girod, "Wyner-Ziv residual coding of video", *Proc. of Picture Coding Symposium,* Beijing, China, pp.28–32, 2006.

**WANG Anhong** is a Ph.D. student in the Institute of Information Science at Beijing Jiaotong University. She has worked in distributed video coding. (Email: wah_ty@yahoo.com.cn)

**ZHAO Yao** is a professor in the Institute of Information Science at Beijing Jiaotong University. He has worked in watermarking, image and video coding. (E-mail: yzhao@bjtu.edu.cn)