WAVELET-DOMAIN DISTRIBUTED VIDEO CODING WITH MOTION-COMPENSATED REFINEMENT

Anhong Wang[1,2], *Yao Zhao*[1], *and Lei Wei*[1]

[1]Institute of Information Science, Beijing Jiaotong University, Beijing, China
[2]Taiyuan University of Science and Technology, Taiyuan, China
Wah_ty@yahoo.com.cn, yzhao@center.njtu.edu.cn

ABSTRACT

In this paper, we propose a distributed video coding (DVC) paradigm based on lattice vector quantization in wavelet domain. In this framework, we use a fine and a coarse lattice vector quantizer to wavelet coefficients and the difference of two lattice quantizers is coded by Turbo encoder. At decoder, side information is gradually updated by motion-compensated refinement. The refinement is built on partially decoded current frame and we give its matching strategy. Due to the refinement, burden of Turbo encoding is cut down, bit rate is saved and the reconstruction is improved to some degree.

Index Terms—distributed source coding, distributed video coding, video coding processing

1. INTRODUCTION

DVC is an application of DSC (distributed source coding) from 1973's Slepian-Wolf [1] and 1976's Wyner-Ziv [2] theories, which state that we can get almost the same compression performance using dependences only at decoder as at both encoder and decoder. DVC offers some supports for easy encoder and consequently can meet the demands of friendly up-linking multimedia serves.

Recently, among some practical DVC system [3][4][5], Anne Aaron and Bernd Girod have provided an easy scheme with low-complexity encoder [3][4]. They encode the key frames in conventional intra-frame mode. To other frames, the Wyner-Ziv ones, a Slepian-Wolf codec based on Turbo coding is used. Anne's scheme [4] is built on DCT domain. While, it has been proved that DWT can overcome the 'block-effect' brought by block-wise DCT and achieve better coding performance in image coding. So, in our scheme, we want to setup our DVC scheme in wavelet domain. Additionally, we use a pair of lattice vector quantization (LVQ) to subtract the dependence between wavelet coefficients, which is different to [3][4][5] based on scalar quantization (SQ).

Side information is important in DVC. The more dependent side information is, the less burden Slepian-Wolf codec will have and less bits are sent. Now, to get better side information, the average and motion-compensated interpolation are both presented [4], but in these interpolations, the side information is only built on past or future frames without any information from current frame. Then, in this paper, we present to use partially decoded current frame to refine the reconstruction and update the side information. Also, we give a new search strategy, modified weighted minimum absolutes distance (MWMAD), for motion searching of vector reconstruction of wavelet coefficients. This is the so-called motion compensation refinement, which first presented in [6] but we extend it to wavelet domain and get partial decoding from frequency band instead of bit-plane. And because of the dependence between coefficients of different frequency band, our method seems more valid. Additionally, this refinement at decoder is different from the idea of [5][7] which is based on a helpful 'hash' from encoder. Our method is based on decoded current frame, which does not bring increase of bit rate naturally.

This paper is organized as following: in Section 2, we will give our DVC framework. In Section 3, the details in our scheme are presented. Then, some comparison experiments are presented in Section 4. Finally, Section 5 is the conclusion including the future work.

2. THE PROPOSED CODING FRAMEWORK

Here, the proposed framework is mainly based on the coding architecture of [4]. But there are differences, such as we replace DCT with DWT and SQ with LVQ, the other contribution is that we extend the motion compensation refinement concept of pixel domain to wavelet domain and give the new searching strategy for vector reconstruction.

As shown in Fig. 1, the system consists of intra-frame encoder and inter-frame decoder. To odd frame, X_{2i-1}

This work was supported in part by National Natural Science Foundation of China (60373028 , 90604032), Specialized Research Fund for the Doctoral Program of Higher Education and Program for New Century Excellent Talents in University.

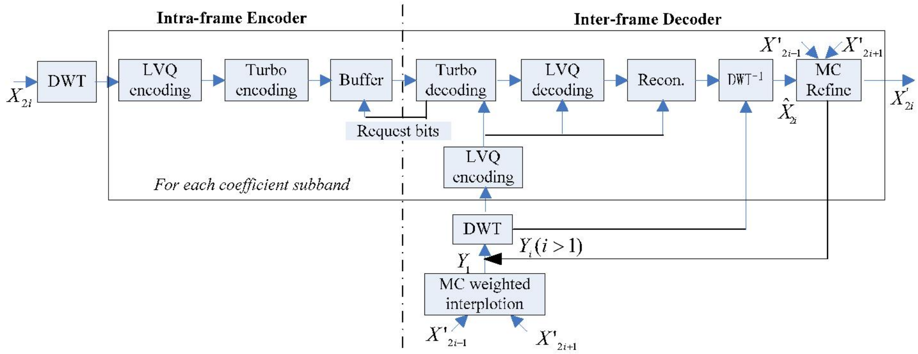

Fig. 1 the proposed DVC framework in wavelet domain

and X_{2i+1}, conventional intra-frame coding and decoding methods can be used. While to even X_{2i}, coding process is: after DWT, its wavelet coefficients are grouped into vectors and quantified by a pair of lattice vector quantizer. Then, for each wavelet subband, Turbo encoder codes the symbol sequence of the difference vector. The parity bits from RCPT (rate-compatible punctured turbo code) are stored in buffer and ready for being sent on decoder's demands. At decoder, the side information is refined gradually. i.e., firstly, Y_1 is decomposed by DWT and its lowest frequency coefficients are used as side information for the lowest frequency symbol of X_{2i}. Then, organizing the restored lowest frequency and the other frequency coefficients from Y_1, after DWT^{-1}, we get partially decoded current frame $\hat{X}_{2i}$ for the first time. Based on $\hat{X}_{2i}$, the restored X'_{2i-1} and X'_{2i+1}, a MC refinement is implemented and we get more accurate X'_{2i}; Then, for second frequency, X'_{2i} is used as side information Y_2. So, side information is updated and the second Wyner-Ziv decoding starts. This refinement is used to reconstruction step by step until all frequency coefficients are decoded. Here, inner loops are developed in refinement for side information.

With DWT and a pair of lattice quantizers we can get as good as or better efficiencies than that with DCT and SQ, additionally, the refined reconstruction brings about the quality enhancement for reconstructed frame and meanwhile, the burden of Turbo coding is cut down due to updated side information. Besides, lattice vector quantization based on fast algorithm [8] will not bring much higher complexity than [4] at encoder, thought the refinement will make decoder more complex because of motion searching, but this will not affect the friendly up-link application.

3. IMPLEMNT DETAILS

3.1. Lattice vector quantization to wavelet coefficients

In LVQ, an input vector x is coded by an index corresponding to the lattice point $Q(x)$ nearest to x [8]. While, in LVQ encoding, we mainly base the scheme: LVQ with side information in [9] [10]. But in [9] and [10], this scheme is only studied in theory. To every input vector x, we use a pair of lattice quantizers with the same dimension, i.e., the fine quantization $Q(.)$, and the coarse $Q_1(.)$, Then, send the symbol of difference vector, $T(x) = Q(x) - Q_1(x)$, to decoder. If x and side information y exist in an identical coarse Voronoi cell, then, $Q(x)$ can be computed losslessly at decoder. But with above method we can only get $Q(x)$, an approximation to original x .So, in reconstruction, we use side information again and take the nearest to side information in Voronoi cell indicated by $Q(x)$ as the final reconstruction. Here, the difference vector represents a set of lattice points so it gets compression again [9] [10].

In our scheme, A_2 lattice is used to quantify wavelet coefficients. Then in each frequency subbands, 2-dimension vectors are formed and quantified with 2^{M_k} quantization level (Here, the quantization level is determined by the number of difference vector of two lattice quantizers), where, $2^{M_k} \in$ {0, 8, 16, 32, 64, 128, 256, 512, 1024}. Here, $2^{M_k} = 0$ means no bits are sent and the according coefficients from side information are used as the reconstruction. In fact, the above quantization accords to 0,1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 bits per pixel respectively and the according bit rate of LVQ, the number of difference vector, is N=7, 13, 31,61,127,217, 469,919.

The combination of above quantization determines the quantization for whole image. To determine the combination, we trained several sequences and try all possible combination certainly with a goal that is to decrease quantization level with frequency level increases. We record rate-distortion points, and finally get 'optimal' quantization label in Fig. 2, and we find this combination makes PSNR improved. Where, we adopt 3-level wavelet decomposing and the data in Fig. 2 show number of bit that every pixel is quantified in its frequency subband.

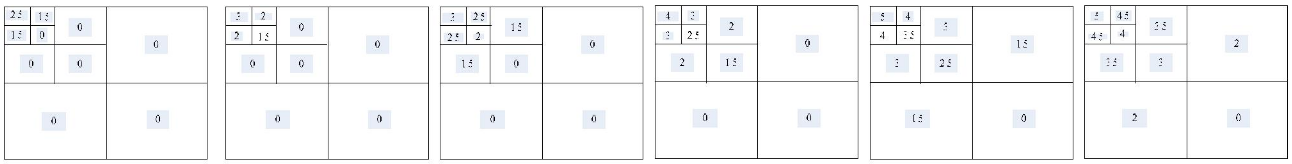

Fig. 2 Combination quantization table to wavelet coefficients

3.2. Motion-Compensated weighted interpolation

Here, the MC weighted interpolation is only used to get the original side information Y_1 for the lowest frequency subband. In DVC, averaged interpolation is an easy and practical method, but it does not give more consideration to motion area. In our system, $Y_1$is from motion compensated interpolation, but we use an easy weighted-interpolation for multiply frames to overcome the overlapped and vacancy in interpolated frame. The process is, firstly, fixing frame X'_{2i+1} , for a block b_{2i+1} in X'_{2i+1} , we search over X'_{2i-1} and find a most matched block b_{2i-1} and its motion vector (v_x, v_y) in it, then for the interpolated frame Y_{2ib} , in the position $(v_x/2, v_y/2)$, we get interpolated block $(b_{2i+1}+b_{2i-1})/2$, second, fixing X'_{2i-1}, for block b'_{2i-1} in X'_{2i-1}, searching the matched block b'_{2i+1} in X'_{2i+1} , we get the interpolated block $(b'_{2i-1}+b'_{2i+1})/2$ in the same position of interpolated frame Y_{2if} , finally, the interpolated side information is taken as :

$$y_{2i} = \alpha x'_{2i-1} + \beta x'_{2i+1} + \gamma y_{2ib} + \theta y_{2if} \qquad (1)$$

Where, $y_{2i}, x'_{2i+1}, x'_{2i-1}, y_{2ib}, y_{2if}$ are the pixels of Y_{2i} , X'_{2i+1} , X'_{2i-1} , Y_{2ib} , Y_{2if} respectively, and $\alpha, \beta, \gamma, \theta$ are the weighted coefficients accordingly. Here, we assume the motion between adjacent frames is uniform.

3.3. Motion compensated refinement

Intuitionally, the more information about current frame we can get, the more accurate reconstruction and side information can be achieved. The goal of motion compensated refinement is to get more accurate reconstructed and side information from decoded current frame. This is different from the original MC interpolation in Section 3.2 where decoder has not any information about current frame. Here, we use the decoded lower frequency coefficients and the other higher coefficients from side information to restore a partially decoded frame. And with this decoded current frame and its past and previous, we get a better reconstruction, which will in turn make the side information more accurate for next frequency, and consequently cut down the burden of Turbo decoding in higher frequency because of improved side information.

It is necessary to introduce refinement to decoded Wyner-Ziv frame. This is because: in Wyner-Ziv coding process, firstly, Turbo decoder correct quantified side information and get the almost error-free symbol of quantization ($P_e \leq 10^{-3}$) . Then, in reconstruction stage, decoder recovery the quantization, that is, if side information is in the same bin indicated by corrected quantization sequence, then, side information is taken as the reconstruction, otherwise, the nearest to side information in the bin is the reconstructed. This reconstruction process limits the maximum distortion from side information but brings other distortion because of inaccurate side information. So, it is necessary to introduce refinement to decoded frame.

The refinement works when the difference between the reconstructed and side information is bigger than the threshold [6], that is:

$$\sum_{(m,n)\in Block} |\hat{x}_{2i}(m,n) - y_{2i}(m,n)| \geq \tau \qquad (2)$$

Where, (m,n) is the coordination of pixel.

Then, what we should do is to find a more accurate block to compensate this block. This is another contribution of this paper. We base our block searching on a matching criterion:

$$MWMAD = MAD(x,y) \times |\hat{C}_{2i} - C_{ref}| / K \qquad (3)$$

Where, $\hat{C}_{2i}$ and C_{ref} are both wavelet coefficients in the same subband as just decoded subband. While, $\hat{C}_{2i}$ is from $\hat{X}_{2i}$ and C_{ref} from the reference frame. K is the number of same quantization output when quantify $\hat{C}_{2i}$ and C_{ref} using LVQ. This criterion, on one hand, limits the block witch has wavelet coefficients far away from the reconstructed in reference frame, on the other hand, promotes the blocks which have the most number of quantization output equal to that of $\hat{X}_{2i}$. The search is to find the minimal *MWMAD* to compensate the reconstructed frame $\hat{X}$, [6] adopt a *WMAD* which only consider *MAD* and the quantization results but not the reconstruction. Here, the reference frames refer to the past, future and the bi-directional searched frames in the past and future frames [6][11].

4. EXPERIMENTS RESULTS

We do comparison experiments as following for two QCIF video sequences: the first 101frames in Foreman and

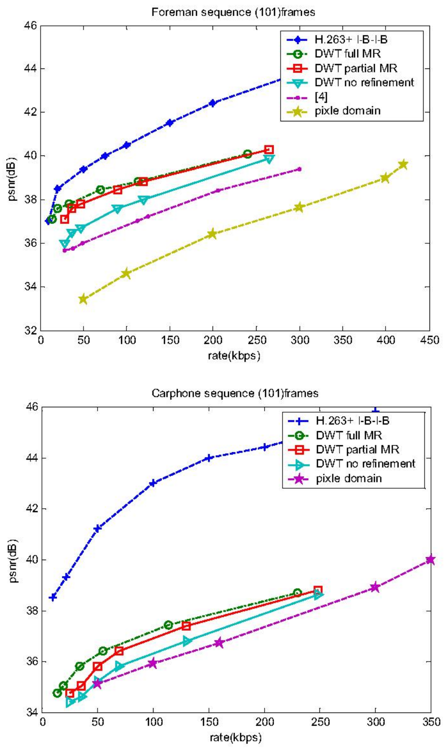

Fig. 3. Rate-distortion performance of Wyner-Ziv video for luminance

Fig. 4. (a) original (b) MC interpolated (c)reconstructed (40kbps) (d) refined reconstructed (40kbps)

Carphone. We only count the luminance of even frames assuming its frames rate is 15 frames per second. RCPT is from [12] and the experiment condition is similar to [4]. In our experiments, we use bi-orthogonal 9/7 wavelet transform. The rate-distortion is plotted in Fig. 3, meanwhile, the inter-frame coding efficiency, I-B-I-B, of standard H.263+ is shown. And 'pixel domain' refers to the coding system based on LVQ but only omitting the wavelet transform and motion compensation refinement. The partial MR means that only reconstruction is refined and full MR is implemented to both reconstruction and side information.

From the results, we can see our DVC of wavelet domain achieves up to 1dB improvement in PSNR than [4] of DCT, that is, with DWT and lattice quantization, we can get better results. These can be seen in both Foreman and Carphone sequence. Meanwhile, the wavelet transform brings better efficiency than pixel domain though the improvement is not evident in Carphone sequence.

To the motion compensated refinement technology, we can see it undoubtedly bring improvements to reconstruction of DVC system, specially in lower rate and high motion sequence. Besides, if the refinement is used to side information, it will cut down the burden of Turbo coding consequently bring up to 20Kb save in 250Kb/s.

From Fig. 4, we can see the reconstructed image is good and the refinement make the motion area clearly in eyes and month.

5. CONCLUSION

We implemented a new LVQ based DVC system with motion compensated refinement in wavelet domain. Our system outperforms the referenced. But the refinements is not so useful in less motion sequences and experiment results do not reach expectation. Also, there are some drawbacks, for example, the rate is determined by feedback and the scalable nature of wavelet transform is not included. Maybe these will be the future work included.

6. REFERENCE

[1] D.Slepian and J.K. Wolf, "Noiseless coding of correlated information sources", IEEE Transaction on information theory, vol.19,pp.471-480,July 1973.

[2] 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, January 1976.

[3] Aaron, S. Rane and B. Griod, "Toward practical Wyner-ziv coding of video" presented at the IEEE Int. Conf. Image Proceeding, Barcelona, Spain, 2003

[4] Aaron, S. Rane, E. Setton and B. Griod, "Transform-domain Wyner-Ziv codec for video", Proc. Visual Communications and Image Processing, VCIP-2004, San Jose, CA, January 2004.

[5] R. Puri and K. Ramchandran, "PRISM: an uplink-friendly multimedia coding paradigm," presented at the Int. Conf. Acoustics, Speech, and Signal Processing, 2003

[6] J.Ascenso,C.Beites,F..Pereira, "motion compensate refinement for low complexity pixel based on distributed video coding", http:// www.img.lx.it.pt/~fp/artigos/AVSS_final.pdf.

[7] Aaron, S. Rane,hash, and B. Griod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver", IEEE Int. Conf. Image Processing, Singapore, 2004.

[8] J.H. Conway and N.J.A. Sloane. "Fast quantizing and decoding algorithms for lattice quantizers and codes". IEEE Trans. on Information Theory. IT-28(2):227-232, March,1982

[9] Sergio. D. Servetto, "lattice quantization with side information", Data Compression Conference 2000

[10] Zhixin Liu and Zixiang Xiong,"Slepian-Wolf Coded Nested Quantization (SWC-NQ) for Wyner-Ziv Coding: Performance Analysis and Code Design," Data Compression Conference 2004.

[11] S.-W. Wu and A. Gersho, "joint estimation of forward and backward motion vector for interpolative prediction of video", IEEE Transaction on image processing, Vol. 3, No. 5, 1994.

[12] Milstein, "on the performance of hybrid FEC/ARQ system using rate compatible punctured turbo codes" IEEE Transaction on Communications, vol.48. no. 6, 2000.