CORRELATION ESTIMATION FOR DISTRIUTED WIRELESS VIDEO COMMUNICATION

Xiaoliang Zhu, Na Zhang, Xiaopeng Fan, Ruiqin Xiong, Debin Zhao

Dept. of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
zxl328659866@gmail.com, hitnzhang@hit.edu.cn, fxp@hit.edu.cn
rqxiong@pku.edu.cn, dbzhao@hit.edu.cn

ABSTRACT

One important problem in distributed video coding is to estimate the variance of the correlation noise between the video signal and its decoder side information. This variance is hard to estimate due to the lack of the motion vectors at the encoder side. In this paper, we first propose a linear model to estimate this variance by referring the zero motion prediction at the encoder based on a Markov field assumption. Furthermore, not only the prediction noise from the video signal itself but also the additional noise due to wireless transmission is considered in this paper. We applied our correlation estimation method in our recent distributed wireless visual communication framework called DCAST. The experimental results show that the proposed method improves the video PSNR by 0.5-1.5dB while avoiding motion estimation at encoder.

***Index Terms*—**distributed video coding, wireless video multicast, correlation noise estimation, confidence interval

1. INTRODUCTION

Distributed source coding (DSC) [2] stands for independent encoding and joint decoding of multiple correlated sources. Distributed video coding (DVC) [8] [9] or Wyner-Ziv video coding (WZVC) is one of the earliest and most advanced applications of DSC. In WZ coding, the encoder does not know or use side information (SI), which make it possible to perform predictive coding without encoder motion estimation (ME) and motion compensation (MC). In a word, a block of pixels or coefficients in the current frame is WZ encoded into a stream without using any information of previously encoded data, and the decoder use all available information to generate side information to decode the current compressed stream.

Note that an important difference between conventional WZ video coding and DVC is that the estimation of the correlation noise are known to both the encoder and decoder and does not change over time in the former. In DVC, however, the decoder has to generate SI using what the information it has. No matter how SI is generated, the estimation of correlation noise is unknown and dynamically changes over time. Indeed, due to the nonstationarity of real scenes, WZ video coding in DVC has to deal with varying correlation noise estimation.

The estimation of correlation noise has been regarded as an important challenge in DVC. Usually, the authors used a Laplacian distribution to model the correlation noise between the original frame and the corresponding SI frame [5] [6]. It has been shown however that these models are not accurate enough if there are occluded regions in the scene [7]. Under this situation, the correlation noise should be concentrated in the occluded areas, which are usually at the moving objects' boundaries.

In [10], the correlation noise is modeled as Laplacian, but to capture the nonstationary nature of the scene, the correlation parameter was varied from pixel to pixel. Similarly, in [11], the Laplacian distribution is used with the parameter estimated online at the sequence level, frame level, block level, and pixel level from decoded frames at the decoder. In [12] and [13], improved online channel estimators are proposed that attempt to address the issue of difficulty in adaptive correlation in smaller spatial regions due to the difficulty in acquiring sufficient statistics.

The above correlation estimation methods are just generating more accurate SI at decoder, and ignore the encoding efficiency. In a recent distributed wireless visual communication frame work, which is called DCAST [1] [4], the correlation estimation is also needed at encoder. Unlike the previous transmission mode does, DCAST does not perform quantization, FEC and modulation to the video signal. To use the inter frame redundancy, DCAST introduced the DVC at encoder to achieve high inter frame compression efficiency and avoid error drifting. More than those, DCAST utilizes coset coding [3] and transmit the coset code of the video signal by raw OFDM channel. At the decoder, the video signal can be decoded with the help of inter frame prediction.

In DCAST, quantization step size in coset coding is calculated according to the correlation noise between video signal and its side information. As in DVC the side information is not available at encoder, to solve this problem, DCAST performs motion estimation and motion compensation to generate current frame's prediction, i.e. the side information. The ME and MC module makes the coset coding can be accomplished at encoder, it also increases the coding complexity significantly. To support more and more convenient mobile devices, we should reduce the encoding complexity and remove the ME and MC module from the encoder. To support the coset coding at encoder, we have to find another way to estimate the correlation noise at encoder.

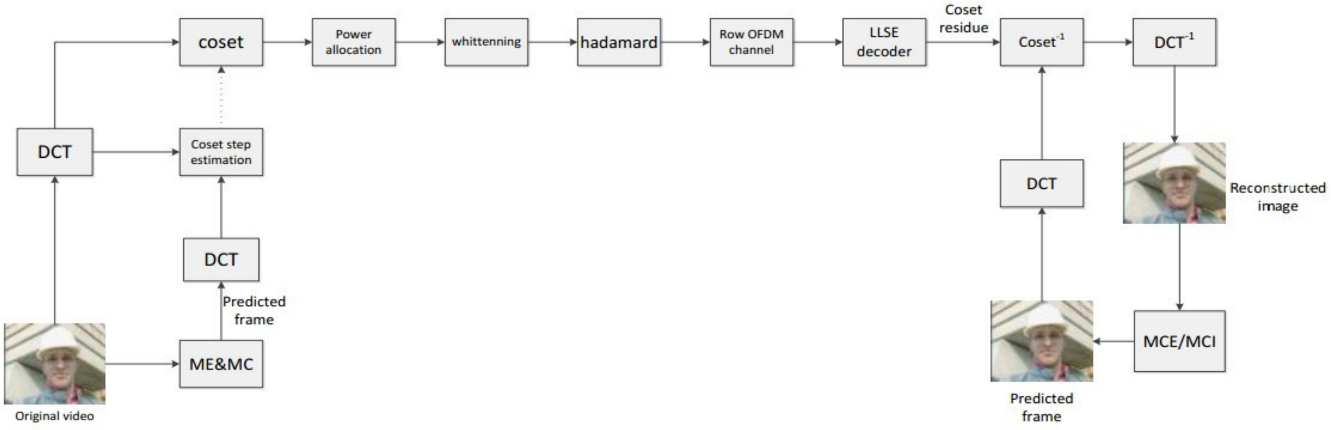


Fig.1. DCAST codec framework

As the decoder needs side information to reconstruct the received video signal in DCAST, in this paper, we propose a new method to estimate the variance of the correlation noise between the video signal and its decoder side information at encoder. Firstly, we propose a linear model to estimate the variance with zero motion prediction at the encoder based on a Markov field assumption. Secondly, we take the noise caused by wireless transmission into account.

The rest part of the paper is organized as follows: section 2 introduces the correlation noise estimation in DCAST. The proposed method is presented in section 3. Experimental results are given in section 4. Conclusions are given in section5.

2. DCAST CODEC FRAMEWORK

At the encoder, DCAST utilizes linear transform and distributed source coding to remove both intra frame redundancy and inter frame redundancy. Fig.1 describes the codec framework of DCAST.

On the encoding side, DCAST performs traditional DCT, quantization, coding and transmitting on the first frame of original video. For the rest of the frames, DCAST preforms motion estimation (ME) and motion compensation (MC) on the previous frame to get the predicted frame of current frame. Then makes DCT transform of current frame and the predicted frame, use the two DCT coefficients to estimate the step size which is used in the following coset coding. As the step is estimated, then performs coset coding on the DCT coefficients of current frame. The coset values are scaled for optimal power allocation. Then, DCAST applies Hadamard transformation on the signal to redistrict energy. In the end, the signal is directly transmitted by the raw OFDM channel.

On the decoding side, the channel superimposes the raw signal and channel noise in received signal. To decode the signal, three steps are performed: Firstly, DCAST performs motion compensated extrapolation (MCE) to get the predicted frame of current reconstructed frame, and transform the predicted frame into DCT domain. Secondly, Linear Least Square Estimator (LLSE) is performed to reconstruct the coset value with minimum distortion. Finally the decoder refines the frame by syndrome decoding.

At the encoder side in DCAST, the correlation noise is just the predicted noise between the predicted frame and current frame, actually the correlation noise also contains the noise caused by the wireless transmission, so we need a new method to estimate the correlation.

3. PROPOSED METHOD FOR CORRELATION ESTIMATION

As we introduced in section 2, both the encoder and decoder has ME and MC module, so we can just keep the decoder's and removes the encoder's to reduce the coding complexity significantly. The encoder of DCAST without ME and MC is shown in Fig.2.

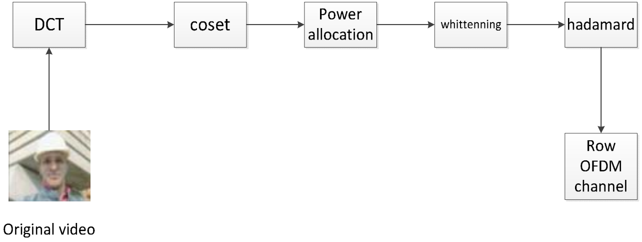


Fig.2. DCAST encoder without ME &MC

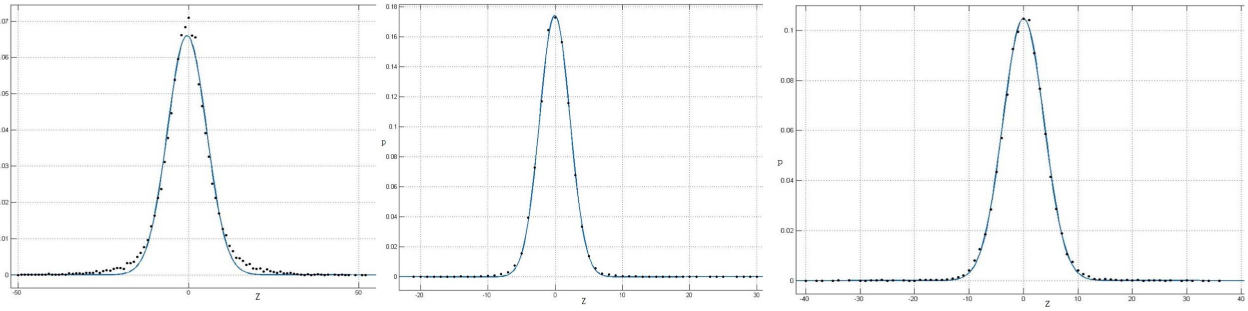


Fig.3. Gaussian fitting for the 3th frame's Z, SNR=10dB

Due to the lack of motion vectors at the encoder side, it's hard to estimate the variance between the video signal and its decoder sider information. To estimate the variance at encoder, firstly we assume the variance satisfy Markov field, i.e. the variance between current frame and its decoder SI is only correlated with the variance between current frame and its encoder side information. Thus at encoder, we take a zero motion prediction on the frame before current frame to get current frame's side information, i.e. the former frame, and calculate the variance between current frame and its encoder SI:

$$D = E(X\text{-}Y)^{2} \qquad (1)$$

X is current frame, Y is X's zero motion prediction, and then the variance between current frame and its decoder side information can be described as:

$$D' = b*D + k \qquad (2)$$

As the correlation noise also contains the noise due to wireless transmission, we have to take this part into account:

$$D' = a*N + b*D + k \qquad (3)$$

N is the channel noise. The left problem is to calculate a, b and k.

Firstly, we assume the channel has no noise, so the noise is only as (2) described. We can simulate the decoder at encoder and transmit 3 frames to calculate b and k. after that we can simulate to transmit 2 frames to calculate a, then we can calculate other variances of other frames, as we know the channel noise N and can calculate D at encoder.

Now let us review how the correlation noise is used in DCAST. At DCAST, the encoder generates video signal's side information to calculate coset coding's quantization step size:

$$q = 2\lceil \max(|X - S'|) + |N_L + N_S| + e \rceil \qquad (4)$$

The X is current frame, S' is the predicted frame at encoder, N_L is the reconstruction noise when coset coding is correct and N_S is the reconstruction noise of the previous frame, both N_L and N_S can be estimated by the channel SNR. In our method, we have to use the variance we calculated to estimate the $max(|X\text{-}S'|)$ in (4).

In our following step, we assume we can get the decoder's SI X' at encoder, then we can calculate $X\text{-}X'$ easily. We let $Z=X\text{-}X'$, so for each frame, we can get a set of Z:

$$Z = \{z_1, z_2, z_3, \ldots\ldots\} \qquad (5)$$

For each z_i and z_j, $i, j=1, 2, 3,\ldots\ldots$ and $i \neq j$, $z_i \neq z_j$. We can get an occurrence probability p_i for each z_i. After various attempts, we find that we can assume p_i and z_i meet the Gaussian distribution, and the fitting results are showed in Fig.3.

As we know the z_i satisfied a certain distribution, we can estimate the distribution arrange of z_i through the concept of confidence interval. In our estimation, we only concern 95% of all possible values that Z will takes, then we can calculate the confidence interval of the average value of Z with a confidence level of 0.95. The confidence interval can be described as:

$$\overline{X} - Z_{\frac{\alpha}{2}} * \sigma, \overline{X} + Z_{\frac{\alpha}{2}} * \sigma) \qquad (6)$$

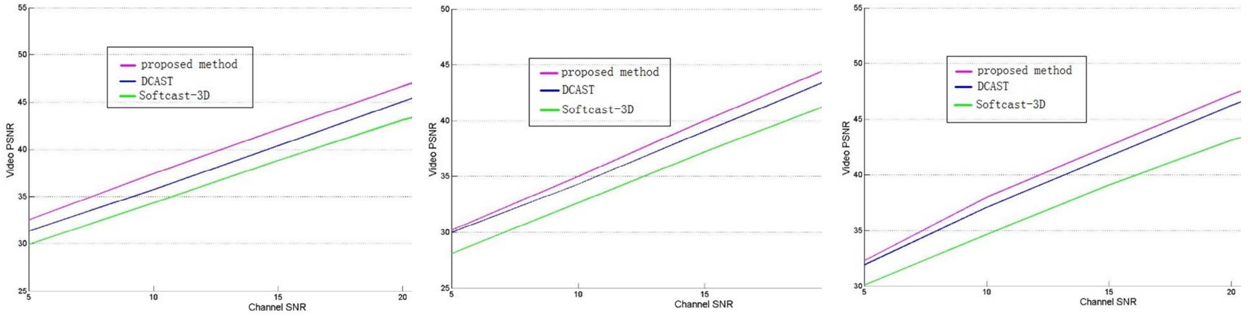
$\overline{X}$ is the mean of a sample, $\alpha=1\text{-}0.95$, $\sigma=D'^{1/2}$, then the length of the confidence interval is $2*Z_{0.025}*\sigma$. To ensure the coset coding is correct, $2*Z_{0.025}*\sigma > max\ (/X\text{-}S'/)$ must be satisfied, then we can let $max(|X\text{-}S'|)=2*Z_{0.025}*\sigma$.

In DCAST, the noise $|N_S+N_L|$ were estimated based on the assumption that the channel SNR is 5dB. This assumption did not consider the change of the channel condition. As we can estimate the distortion that caused by the noise, then we can use it to calculate the step size to make it more accurate.

4. EXPERIMENTAL RESULTS

In experiments, we evaluate the performance of our method, and we compare it with DCAST and 3D-DCT based Softcast (Softcast3D). Softcast3D is another wireless video broadcasting approach, and it partially introduce 3D-DCT to achieve inter frame compression.

All the three methods, our method, DCAST, Sofecast-3D, encode the video into packets by soft compression. The video test sequences are 'foreman_qcif.yuv', 'mother-daughter_qcif.yuv' and 'salesman_qcif.yuv'. The video



(a)foreman_qcif (b)mother-daughter.qcif (c)salesman_qcif

Fig.4. Performance comparison

(a) Original frame (b)Softcast-3D (c)DCAST (d) proposed method

Fig.6.Visual quality comparison, the 3th frame of foreman_qcif.yuv,SNR=10dB

frame rate is 30HZ. The GOP structure is 'IPPP'. The channel bandwidth is equal to the video bandwidth (i.e. the number of video pixels per second) .The experiments is video multicast to users with diverse SNR.

The video packets are transmitted to OFDM. The OFDM signal is transmitted over AWNG channel. The receiver passes the signal to the OFDM module to perform CFO corrections, channel estimation and correction, and phase tracking. Then it inverts the operations of the transmitter and forwards the soft information to video decoding layer.

The results are given in Fig.4. DCAST is 1-4dB better than softcast3D. Our method is 0.5-1.5dB better than DCAST. The visual quality comparison is given in Fig.5. The channel SNR is set to be 10dB. It is clear that our method has better visual quality than DCAST and sofecast3D.

5. CONCLUSIONS

In this paper, we propose a new method to estimate the variance of correlation noise between current frame and its decoder side information. As the encoder lacks the motion vector in DVC, the variance is hard to estimate. We applied our method in DCAST, the experimental results show that the method improve the performance and meanwhile reduce the encoder complexity significantly.

6. ACKNOWLEDGEMENT

This work was supported in part by the Major State Basic Research Development Program of China's 973 Program under Grant 2009CB320905, the National Science Foundational of China (NSFC) under grants 61272386 and 61100095, and the Program for New Century Excellent Talents in University (NCET) of China (NXET-11-0797).

6. REFERENCE

[1] X.Fan, F.Wu, and D.Zhao, "Distributed soft video broadcast (DCAST) with explicit motion," Data Compression Conference (DCC), pp.199 - 208, 2012.

[2] S.Pradhan and K.Ramchandran, "Distributed source coding using syndromes (DISCUS): de-sign and construction," IEEE Trans. Inform. Theory, vol. IT-49, pp. 626–643, 2003.

[3] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," in Proc. IEEE Data Compression Conf., 1999, pp. 158–167.

[4] X. Fan, F. Wu, and D. Zhao, "D-Cast: DSC based Soft Mobile Video Broadcast," in ACM International Conference on Mobile and Ubiquitous Multimedia (MUM), Beijing, China, December 2011.

[5] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding

performance," in Proc. IEEE ICASSP , Toulouse, France, May 2006, vol. 5, pp. II-525–II-528.

[6] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in Proc. 5th EURASIP, Slovak Republic, Jul. 2005, pp.176–181.

[7] P. Meyer, R. P. Westerlaken, R. K. Gunnewiek, and R. L. La-gendijk, "Distributed source coding of video with non-stationary side-information," Proc. SPIE, vol. 5960, pp. 857–866, Jul. 2005.

[8] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in Proc. 36th Asilomar Conf. Signals, Syst. Comput., vol. 1. Nov. 2002,pp. 240–244.

[9] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," inProc. Annu. Allerton Conf. Commun., Contr. Comput. , vol. 40, no. 1. 2002, pp. 586–595.

[10] M. Dalai, R. Leonardi, and F. Pereira, "Improving turbo codec integra-tion in pixel-domain distributed video coding," in Proc. ICASSP , vol. 2.2006, pp. 537–540.

[11] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 9, pp. 1177–1190, Sep. 2008.

[12] X. Fan, O. C. Au, and N. M. Cheung, "Adaptive correlation estimation for general Wyner-Ziv video coding," in Proc. ICIP, Nov. 2009, pp.1409–1412.

[13] X. Huang and S. Forchhammer, "Improved virtual channel noise model for transform domain Wyner-Ziv video coding," in Proc. ICASSP , Apr. 2009, pp. 921–924.