SOFT MOBILE VIDEO BROADCAST BASED ON SIDE INFORMATION REFINING

Wei Huang, Xiaopeng Fan, Debin Zhao

Dept. of Computer Science and technology, Harbin Institute of Technology, Harbin, China
hwzl360@yahoo.com.cn, {fxp, dbzhao} @hit.edu.cn

ABSTRACT

Video broadcasting is a popular application of wireless network, whose main challenge is to accommodate different users with different channel conditions. Recently, a novel 'D-Cast' approach based on distributed source coding (DSC) is proposed. It can avoid error propagation and still achieve high compression efficiency in inter frame coding by utilizing coset coding and soft broadcast. However, D-CAST is not very efficient because of rough side information. In this work, we present a novel soft mobile video broadcast approach based on side information refinement algorithm (SIR-CAST) to improve the quality of the side information. Moreover, SIR-Cast optimizes the estimate of the quantifying step (Qstep) which is corresponding to the refined side information. Thus, SIR-CAST outperforms D-CAST about 1dB-2dB in video PSNR while maintaining the similar graceful degradation feature as D-CAST.

Index Terms—wireless networks, video multicast, D-CAST, side information refining, coset coding

1. INTRODUCTION

Mobile video broadcasting is a popular application which can transmit video signal to multiple users. Typical wireless video broadcast approaches are based on the DVB-T standard [1]. It combines a layered transmission scheme [2][3] and scalable video coding scheme [4][5]. The video signal is encoded into one base layer (BL) and some enhancement layers (EL) by scalable video coding scheme. The hierarchical modulation (HM) [6] is utilized to make users decode different numbers of layers in transmission. Therefore, both low SNR users and high SNR users can receive video signal which matches their channel conditions. However, the poor compression and transmission efficiency of the layered schemes are not very satisfying. Furthermore, the limited choices of BL and EL rates create cliff effects in video quality as opposed to continuously changing channel condition.

The new proposed approach called Softcast provides a new idea to solve the problem by transmitting the linear transform of the video signal in analog channel. But Softcast is not very efficient in inter frame compression. In a recent improved version of Softcast, the utilization of 3D-DCT partially enables inter frame compression [8]. However, without motion compensation the inter frame redundancy is still not fully exploited in 3D-DCT based softcast.

Recently, a novel wireless video broadcasting approach called D-Cast [9] has been proposed. D-Cast is based on softcast and performs gracefully in video multicast. It not only achieves efficient inter frame compression but also avoids error drifting. D-Cast utilizes coset coding to transmit the coset code [10] of the video signal by raw OFDM and decode the coset code by utilizing the decoder motion estimation techniques of distributed source coding (DSC) [11]. As a result, it reduces the magnitude of the signal and the receiver can still decode the signal with the help of the inter frame prediction (i.e. the side information). D-CAST is not very efficient because of rough side information which is got from a simple motion compensated extrapolation.

In this paper, we propose a new wireless video multicast approach called SIR-Cast. In contrast to D-Cast, SIR-Cast uses a side information refinement (SIR) algorithm [12] to generate the side information. Moreover, SIR-Cast optimizes the estimate of the Qstep which is corresponding to the refined side information. Therefore, we can reconstruct the video frame more accurately with refined side information and optimized Qstep. In experiments, the proposed approach achieves significant gain over D-Cast.

This paper is organized as follows. Section 2 is a brief review of D-Cast. In Section 3, a detailed description of the proposed SIR-Cast is presented. Simulation results are then discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. OVERVIEW OF D-CAST

D-Cast is a wireless video multicast system based on soft broadcast and soft compression. The framework of D-Cast is described in Fig.1. At first, the original image is transformed into DCT domain. Then it partitions the DCT coefficients into several cosets to get the modular reminders for each DCT coefficients. With the encoder side information, Qstep (the step of the coset coding) is calculated. Then the modular reminders are scaled for optimal power allocation. Before

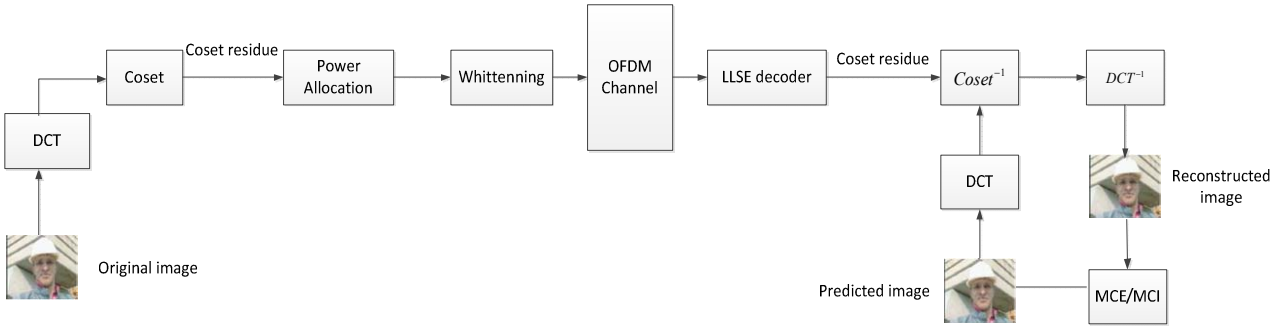


Fig. 1. Framework of D-Cast

soft transmission, D-Cast scales the modular reminders for optimal power allocation. Then the whitening process is applied on the signal and the resulting signal is directly transmitted over the raw OFDM channel like analog transmission. At the receiver, the received signal is raw signal plus channel noise. After denoising, the DCT coefficients of the coset values are estimated by LLSE. Meanwhile we can get the predicted image of the current frame through motion compensated extrapolation (MCE), and the predicted image is transformed into DCT domain. Then with the coset values and the predictors, the coset decoding module recovers the DCT coefficients of the current frame.

2.1. Coset Coding

The quality of the decoded frames of D-Cast is strongly determined by coset coding which is a typical technique used in DSC. D-Cast throw away the main part of each transform coefficient of the current frame X by dividing each X by a Qstep q and get the remainder L as follows.

$$L = X - \left\lfloor \frac{X}{q} + \frac{1}{2} \right\rfloor q \tag{1}$$

L is the coset index and it represents the detail of X in particular.

At decoder, D-Cast performs motion estimation (ME) to get the predicted value of the current frame, which is the so-called motion compensated extrapolation (MCE). Then, the predicted frame is transformed into DCT domain. With the decoded coset value $\hat{L}$ and the decoder side information S (i.e. the predicted DCT coefficients at decoder), the receiver can reconstruct the DCT coefficients.

With the received coset value $\hat{L}$, there are multiple possible reconstructions of X forming a coset C According to (1).

$$C = \left\{ \hat{L}, \hat{L} \pm q, \hat{L} \pm 2q, \hat{L} \pm 3q, \ldots \right\} \tag{2}$$

In the coset C, the nearest one to the decoder side information S is selected as the reconstruction of the DCT coefficient.

$$\hat{X} = \arg\min_{c \in C} |c - S| \tag{3}$$

In D-Cast, it is proved to be correct when q meets the following conditions.

$$q > 2(|X - S'| + |N_L + N_S|) \tag{4}$$

Where N_L is the reconstruction noise when coset decoding is correct and N_S is the reconstruction noise of the previous frame. S' represents encoder side information (i.e. the predicted DCT coefficients at encoder).

As a result, q is

$$q = 2\left\lceil \max(|X - S'|) + |N_L + N_S| + \varepsilon \right\rceil \tag{5}$$

3. SIR-CAST STRATEGY

Obviously the quality of side information is essential and it determines the performance of the D-Cast directly according to (3). D-CAST is not very efficient because of rough side information. Specifically, the decoder side information S is so rough that the decoder may make wrong decisions with a smaller q. In this paper, we refine the side information and find a proper q which is large enough to be decoded accurately. We present a novel approach (SIR-CAST) based on side information refinement algorithm to improve the quality of the side information. Moreover, SIR-Cast optimizes the estimate of q (the quantifying step). SIR-Cast utilizes side information refining algorithm (SIR) [12] to successively refine the decoder side information S, thus providing more accurate side information to get the proper q. In this section, the framework of SIR-Cast will be described in detail.

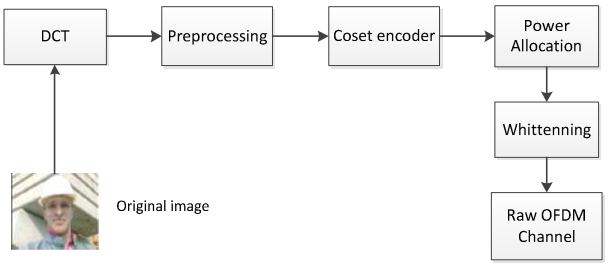


Fig. 2. SIR-Cast server

Fig. 2 depicts the server side of SIR-Cast. SIR-Cast first transforms the original image into DCT domain. For the efficient design of the SIR-Cast approach, some important decisions had to be taken. An integer 4 × 4 block-based DCT will be applied over each frame. According to the position occupied by each DCT coefficient within the 4 × 4 blocks, the DCT coefficients of the entire frame are grouped together, forming the DCT coefficients bands. As a result, there are a total of 16 coefficients bands.

Then the 16 DCT coefficients bands are divided into 4 groups as depicted in Fig. 3. Then coset coding divides 4 DCT bands groups (DCT bands 1, DCT bands 2, DCT bands 3 and DCT bands 4) to get the remainders with 4 different q (q_1, q_2, q_3 and q_4). Q_1, q_2, q_3 and q_4 are decreasing and they are calculated by the step of preprocessing. The preprocessing process will be described in section 3.1.

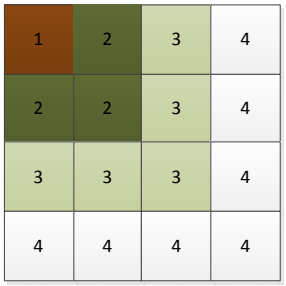


Fig. 3. Partition of DCT coefficients bands

After the step of cost coding, the modular reminders will be scaled for optimal power allocation. Then the whitening process is applied on the signal and the resulting signal is directly transmitted over the raw OFDM channel like analog signal.

The client side of SIR-Cast is depicted in Fig. 4. First, inverse Hadamard transform is applied on the signal received from the raw OFDM channel. The Linear Least Square Estimator (LLSE) is to reconstruct the coset value reminders with minimum distortion. Then the following step is an iterative decoding process. We carry on coset decoding and SIR (refine the side information with the so far decoded bands groups every loop) circularly to get more accurate side information, as a result, we can reconstruct the frame with high quality.

The initial side information is generated from the motion estimation of the previous reconstructed frame. With the initial information (we called it S_1) and the coset values of DCT bands 1, we can recover the DCT coefficients of DCT bands 1. After that, the decoded DCT bands 1 are utilized to refine the side information by the SIR step. Then the refined side information (we called it S_2) will be used by the next decoding which will decode the DCT bands 1 and DCT bands 2. Similarly, S_3 which is generated by SIR with the DCT bands 1 and DCT bands 2 will be used to recover the DCT bands 3 and so on.

As more and more DCT bands groups are decoded, we can get more and more accurate side information by the SIR process which utilizes the so far decoded DCT bands groups to refine the side information. As a result, SIR utilizes all of the 4 decoded DCT bands groups to generate the final side information. Then with the final information and all of the coset values, we can reconstruct the current frame. The detailed SIR process will be described in section 3.2.

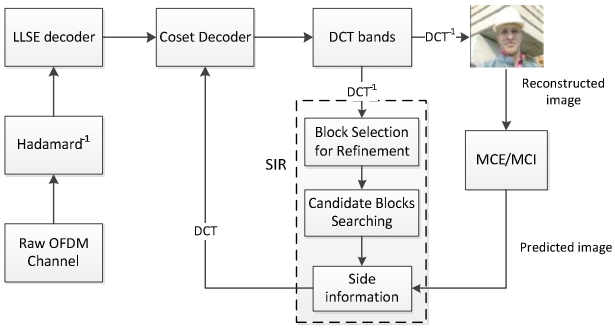


Fig. 4. SIR-Cast client

3.1. Preprocessing

As mentioned before, SIR-Cast divides the DCT coefficients into 4 DCT bands groups. To refine the side information iteratively, the 4 DCT bands groups are encoded with decreasing q. Therefore, it's crucial to find proper q to make right decisions in coset decoding.

According to (5), q is determined greatly by the distortion between the current frame X and the side information S. Assume X is the current frame, S is the decoder side information of the current frame. The distortion D could be calculated by the following equation [13]:

$$D = E(X-S)^2 = 2\delta_x^2(1-\rho^{|\Delta|}) \quad (6)$$

Where δ_x^2 is calculated by the variance of the video frame. ρ represents the correlation of the adjacent pixels in video frame. $|\Delta|$ is the distortion of the motion vector. As ρ is less than 1 and $|\Delta|$ is not very large, the equation (6) can be approximated as

$$D = 2\delta_x^2(1-\rho)|\Delta| \tag{7}$$

Assume $X-S$ distribute is Gaussian distribution, so D is the variance of $X-S$. Therefore, $E(|X-S|)$ can be approximated as

$$E(|X-S|) = \frac{\sqrt{D}}{\sqrt{2\pi}} \tag{8}$$

As δ_x^2 and ρ are constants in video frames, $|X-S|$ is determined by $|\Delta|$ according to (7) and (8). Therefore, the equation (5) can be approximated as

$$q = 2\left\lceil a\sqrt{\Delta} + b \right\rceil \tag{9}$$

Where a is a constant which can be calculated form (7) and (8). b represents $|N_L + N_S|$ which so small that it could be ignored. As a result, q which is determined greatly by $|X-S|$ changes as the $|\Delta|$ of the corresponding side information S changes.

As we know, S_4 is refined by DCT bands 1, DCT bands 2 and DCT bands 4. So we can think S_4 is corrected by the 3/4 resolution version of the current frame. Similarly, S_3 is corrected by the 1/2 version of the current frame and S_2 is corrected by the 1/4 version of the current frame. Assume the distortion of the motion vector $|\Delta|$ in S_2 is w, the distortion of the motion vector $|\Delta|$ in S_3 should be w/2 and that in S_4 should be w/3. As a result, the changing trends of q can be approximated as follows according to (9).

$$\begin{aligned} q_3 &= q_2 / \sqrt{2} \\ q_4 &= q_2 / \sqrt{3} \end{aligned} \tag{10}$$

Now the crucial problem is to find a proper q_2. Because S_2 is corrected by DCT bands 1, S_2 is more accurate than the initial decode side information S_1. Therefore, q_2 should be a little smaller than q_1. The experiments prove it works well, when q_2 is set to be $0.9*q_1$.

D-Cast calculates q_1 with the maximum of $|X- S'|$ according to (5). Obviously, it's a little ineffective because we don't have to decode every coefficient accurately. In fact, the residual between S' and X are very different. So SIR-Cast uses the second value to calculate the initial q_1 instead of the maximum of $|X- S'|$. With the 4 different q values, coset encoder divides 4 DCT bands groups (DCT bands 1, DCT bands 2, DCT bands 3 and DCT bands 4) to get the remainders separately.

3.2. SIR algorithm

As depicted in Fig. 4.The SIR algorithm [12] we used in this paper consists in three main steps. The first step defines which blocks are worthwhile to be selected for refinement. The selection is based on the quality of the side information and the current reconstructed frame. In detail, this module should select the blocks for which the side information is rather different from the decoded frame bands.

At first, each 4 × 4 block n is reconstructed. The reconstruction is applied by coping the so far decodes DCT bands from the current frame to last refined side information, creating R_n.

To evaluate the difference between the side information and the decoded frame, we calculate the sum of squared errors ε_n between the reconstructed block R_n and last refined side information collocated block Y_n.

$$\varepsilon_n = \sum_{x=0}^{3}\sum_{y=0}^{3}\left(Y_n(x,y) - R_n(x,y)\right)^2 \tag{11}$$

The (x, y) are the pixel coordinates inside the 4 × 4 block and the reconstructed block R_n is reconstructed by coping the so far decoded DCT bands from the decoded frame to last refined side information. If the sum of squared errors is equal or exceeds a threshold, this block is selected for refinement. In our experiments, the threshold is set to 100.

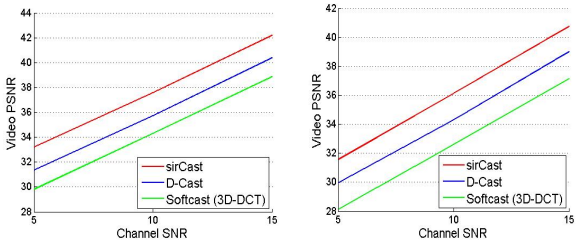
For each selected block that can be refined, the second step determines which the best candidate blocks to become the new side information are. The selected candidate blocks should be similar to the reconstruct block which is selected for refinement.

After the candidate blocks searching, we get some blocks which is similar to the reconstructed block. To create the new side information, the candidate blocks collected by the second step are used by the third step to create the improved side information coefficients. Then the refined side information will be used to decode the next DCT bands groups.

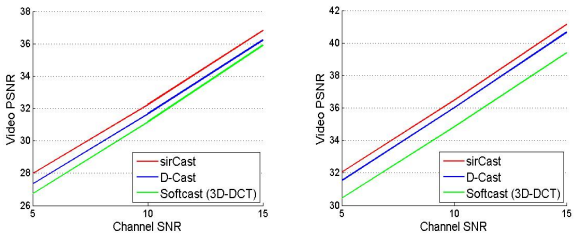
4. EXPRIMENTS

In experiments, we implement another two existing methods for comparison in video multicast. We compare the performance of SIR-CAST with that of D-Cast [9] and 3D-Softcast [8].

All of them are based soft broadcast and soft compression. The video 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.



(a)Foreman_qcif (b)News_qcif



(c)Bus_qcif (c)Soccer_qcif

Fig. 5. Performance comparison

(a) Original frame (b) Softcast(3D-DCT)

(c) D-Cast (d) SIR-Cast

Fig. 6. Visual quality comparison, the 15th frame of foreman_qcif.yuv, AWGN channel, SNR=5dB

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.5. Our SIR-CAST is 2dB-4dB better than softcast3D, and is 1dB-2dB better than D-Cast.

Fig.6 depicts the visual quality comparison. The channel SNR is set to be 5dB.It is clear that SIR-CAST has better visual quality than D-Cast and softcast3D.

5. CONCLUSIONS

In this work, we present a novel approach (SIR-CAST) based on side information refinement algorithm to improve the quality of the side information. Moreover, SIR-Cast optimizes the estimate of the quantifying step. As a result, SIR-Cast achieves better performance than the state-of-art multicast approach D-Cast.

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 Foundation of China (NSFC) under grants 61272386 and 61100095, and the Program for New Century Excellent Talents in University (NCET) of China (NCET-11-0797)

7. REFERENCES

[1] "Digital Video Broadcasting (DVB)," Website, 2009, http://www.etsi.org/deliver/etsien/300700300799/300744/01.06.0160/en300744v010601p.pdf.

[2] N. Shacham, "Multipoint communication by hierarchically encoded data," in INFOCOM '92.Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE, may 1992, pp. 2107 –2114 vol.3.

[3] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in Conference proceedings on Applications, technologies, architectures, and protocols for computer commu-nications, ser. SIGCOMM '96. New York, NY, USA: ACM, 1996, pp. 117–130. [Online]. Available: http://doi.acm.org/10.1145/248156.248168

[4] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 11, no. 3, pp. 332 –344, mar 2001.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h.264/avc standard," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 9, pp. 1103 –1120, sept. 2007.

[6] K. Ramchandran, A. Ortega, K. Uz, and M. Vetterli, "Multiresolution broadcast for digital hdtv using joint source-channel coding," in Communications, 1992. ICC '92, Conference record, SUPERCOMM/ICC '92, Discovering a New World of Communications., IEEE International Conference on, jun 1992, pp. 556 –560 vol.1.

[7] S. Jakubczak, H. Rahul, and D. Katabi, “One-Size-Fits-All Wireless Video,” in Proc. Eighth ACM SIGCOMM HotNets Workshop, New York City, NY, October 2009.
[8] S. Jakubczak and D. Katabi, “A cross-layer design for scalable mobile video,” in Proceedings of the 17th annual international conference on Mobile computing and networking, ser. Mobi-Com ’11. New York, NY, USA: ACM, 2011, pp. 289–300. [Online]. Available: http://doi.acm.org/10.11-45/2030613.2030646
[9] 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.
[10] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” in Proc. IEEE Data Compression Conf., 1999, pp. 158–167.
[11] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” IEEE Trans. Inform. Theory, vol. IT-49, pp. 626–643, 2003.
[12] R. Martins, C. Brites, J. Ascenso, and F. Pereira, “Refining side information for improved transform domain wyner-ziv video coding,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, no. 9, pp. 1327 –1341, sept. 2009.
[13] X. Fan, O.C. Au, N.M. Cheung, and J. Zhou, “Successive refinement based Wyner-Ziv video compression,” Signal Processing: Image Communication, vol. 25, no. 1, pp. 47-63, 2010.