DISTRIBUTED SOFT VIDEO BROADCAST WITH VARIABLE BLOCK SIZE MOTION ESTIMATION

Ailing Zhang, Xiaopeng Fan, Ruiqin Xiong, Debin Zhao
Dept. of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
zal.hit@foxmail.com, fxp@hit.edu.cn, rqxiong@pku.edu.cn, dbzhao@hit.edu.cn

ABSTRACT

In recent years, video broadcast has become a popular application, but the traditional hierarchical design requires the source to pick a bitrate and video resolution for encoding before transmission, it cannot be efficient to accommodate users with different channel quality. The proposed DCAST scheme can solve this problem. DCAST uses the ME and MC technology to generate a predicted frame which helps current frame to do coset coding. However, in the ME process, DCAST uses fixed block size. In this paper, we use the variable block size motion estimation to replace the fixed block size motion estimation, it can effectively reduce the block effect and improve the quality of the reconstructed frame generated by MC and predicted frame. The DCAST with variable block size motion estimation is 0.5dB better than DCAST with fixed block size motion estimation.

Index Terms**—**Video broadcast, DCAST, SoftCast, Motion compensation, Motion homogeneity

1. INTRODUCTION

Wireless video is becoming more and more important in our lives, and the existing wireless video broadcasting technology, a layered transmission scheme[1][2], faces a main challenge that it cannot effectively meet the users with different channel quality. The typical wireless video broadcasting technology requires the source to select a particular bitrate for coding before video transmission, but how to select an appropriate bitrate is a problem. If the source transmits on the high bitrate, it cannot satisfy a user with poor channel, if it transmits on the lower bitrate, it reduces everyone to the performance of the worst receiver. Scalable video coding (SVC) scheme [3] [4] can solve this problem. SVC encodes video signal into a base layer and multiple enhancement layers. In transmission, the layered modulator [5] adds multilayer bits to a wireless video signal. According to their own channel quality users decode different numbers of layers. The more layers user receives, the better the quality of the decoded video will be. Their redundancy is not removed because there is no contact between the bit streams of each layer. SVC reduces the compression and transmission performance. In addition, SVC only provides limited choices of BL and EL rates.

Recently, a novel called SoftCast [6] [7], wireless video multicast method is proposed, which is based on the soft compression and soft transmission. SoftCast introduces a new video codec with new compression and error protection method. The main idea of SoftCast is that it ensures the distances between the transmitted codewords are linearly related to the differences between the pixel values. As a result, the codewords received by a user with a high SNR are close to the transmitted codewords, and therefore decoded pixels that are close to original pixels, the user can obtain an image of high-fidelity. On the other hand, the codewords received by a user with a low SNR are far away from the transmitted codewords, thus decoded pixels are far away from the original pixels, the user will get an image of low-fidelity. Such a coding technique enables the source to use an unfixed bit rate or coding rate to broadcast its package, meanwhile, it allows users to decode a video quality commensurate with its channel quality. SoftCast uses 2D-DCT technology to achieve intra frame compression, uses 3D-DCT technology to realize inter frame compression. Because SoftCast doesn't use motion compensation technology, its compression efficiency is not high. According to the defect in the Softcast, it has formed a new distributed wireless video broadcast method called DCAST [8] [9]. DCAST is based on distributed source coding [10] theory. It also exploits soft broadcast technology. In addition, DCAST adds the motion estimation module. Because the traditional inter frame coding exists error propagation problem, DCAST uses DSC technique to obtain high efficient inter frame compression performance and meanwhile can avoid error propagation problem. DCAST transmits the coset code [11] values of video signal rather than transmits video signal itself by the raw OFDM. Furthermore, in the DCAST scheme, the ME and MV transmission are optimized to improve its performance. DCAST has no frame delay and is applicable in realtime applications. The second part is the detailed introduction of DCAST.

The rest part of the paper is organized as follows: the second section introduces DCAST codec framework, and the third section describes the inter mode decision method used in this paper, the fourth part gives the simulation results, the fifth section concludes this paper.

2. DCAST CODEC FRAMEWORK

DCAST [8] is based on soft compression and soft transmission. It uses DCT to remove the spatial redundancy and distributed source coding to remove the temporal redundancy. DCAST program is divided into server and client. Fig .1 describes the server side of DCAST.

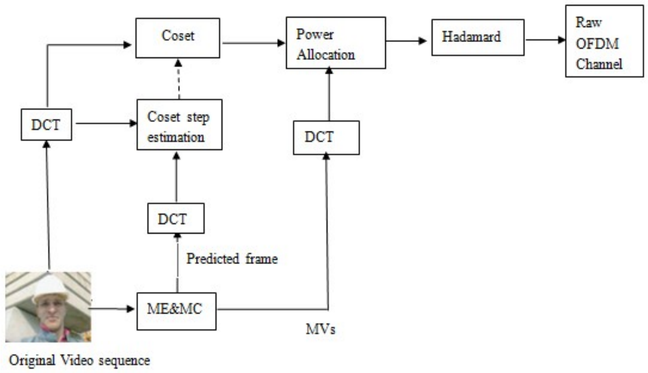


Fig.1. The server of DCAST

On the server side, DCAST first performs DCT on the original image and meanwhile does motion estimation and motion compensation on the original video sequence to generate a predicted frame of the current frame. Then DCAST transforms the MVs into DCT domain and the predicted frame also does DCT. Coset step is estimated by the DCT coefficients of the original frame and the predicted frame. Then, according to the DCT coefficients of the original frame and the step length of the coset coding DCAST performs coset coding on the DCT coefficients of the original frame to get coset values. The coset values and the DCT coefficients of the MVs are scaled for optimal power allocation. Then, DCAST applies Hadamard transformation on the signal to redistrict energy, i.e. each packet has the same importance. In the end, the resulting signal is directly transmitted by the raw OFDM channel.

The client of DCAST is described as Fig.2. The signal received from the OFDM channel contains raw signal and channel noise. DCAST first applies inverse Hadamard transform on the receiving signal. Then, the MMSE module gets coset values and the DCT coefficients of the MVs. Then, DCAST performs inverse DCT on the DCT coefficients of the MVs to get MVs in the spatial domain. According to the reconstructed frame and MVs MC module generates a predicated frame. The predicted frame is transformed into DCT domain, and coset decoding module uses the DCT coefficients of the predicted frame and coset values to recover the DCT coefficients of the current frame. At last, the coefficients obtained by coset decoding module perform inverse DCT, and they are combined with the predicted frame through MMSE to get the final reconstructed frame.

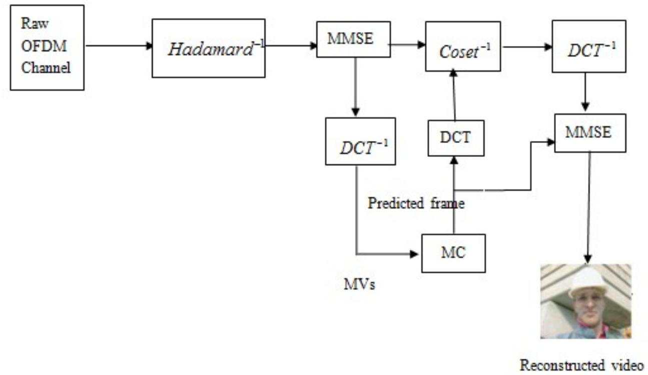


Fig.2. The client of DCAST

The most advanced DCAST scheme, in the motion estimation module, adopts fixed block size. We know that H.264 coding standard contains variable block size motion compensation technology and it improves the coding performance of H.264. If variable block size can be used in the motion estimation of DCAST, it will improve coding performance of DCAST.

3. INTERFRAME MODE DECISION

H.264 video coding standard provides superior coding performance by using some new coding tools, including spatial directional prediction in the intra frame coding, the variable block size and multiple reference frames in inter frame coding, to acquire the efficient coding performance. This paper is intended to add variable block size motion estimation and compensation to DCAST to improve the performance of the codec. H.264 uses rate distortion optimization (RDO) technology [12] [13] to determine the best mode, RDO calculates the RD cost by performing a sequence of encoding operations. If RDO calculation is performed on every possible mode, this will greatly increase the coding complexity. In this paper, fast inter mode decision algorithm based on motion homogeneity [14] is introduced. The algorithm is described as follows:

A. MB Classification Based On Motion Homogeneity

As [14] described, there presents a higher correlation between the motion homogeneity presented in the MB and the optimal mode selected by full mode decision technology. In [14], it proposes three directional motion homogeneity measurement methods, i.e. horizontal direction, vertical direction and quartered direction, which are used for macro block classification.

First, this algorithm performs motion estimation on 4×4 block size. Then, a normalized MV field is formed at 4×4 block level, which is used for calculating the motion homogeneity of each macro block. For every 4×4 block in current frame c, the normalized MV is calculated by formula (1), where c is the frame number and r0 is the reference frame number. In this paper, we choose the previous frame as the reference frame, i.e. c-r0 = 1.

P-frames: $\mathrm{NMV}_{m,n} = \frac{MV_{m,n}}{c-r0} = MV_{m,n}$ (1)

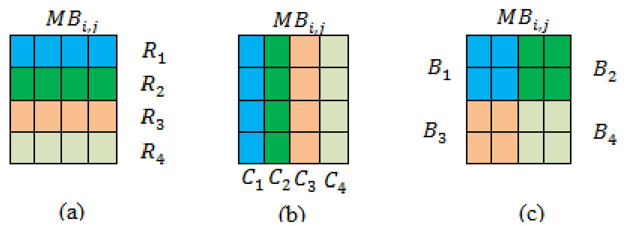


Fig.3. Partitions used in calculating motion homogeneity measures

For one macro block, there are three divided manner. For example, a macro block $\mathrm{M}_{i,j}$ covers 16 blocks and its normalized MVs are described as $NMV_{m,n} = \{mvx_{m,n}, mvy_{m,n}\}, m \in [4i, 4i+3], n \in [4j, 4j+3]$. As shown in Fig.3. Each partition P_k may represent each row R_k in Fig.3 (a), or each column C_k in Fig.3 (b), or each 8×8 block B_k in Fig.3(c). The mean deviation of each partition P_k is defined as follows:

$$\mathrm{MD}(\mathrm{P_k}) = \frac{1}{4}\sum\nolimits_{\mathrm{P_k}}\{|\mathrm{mvx_{m,n}} - \frac{1}{4}\sum\nolimits_{\mathrm{P_k}} \mathrm{mvx_{m,n}}| + ||\mathrm{mvy_{m,n}} - \text{14Pkmvym,n}|\}$$

The motion homogeneity of each MB in the horizontal, vertical, quartered direction is defined as follows:

$$\mathrm{H_{i,j}} = \frac{1}{4}\sum_{1}^{4}\mathrm{MD}(\mathrm{R_k}), \mathrm{V_{i,j}} = \frac{1}{4}\sum_{1}^{4}\mathrm{MD}(\mathrm{C_k}),$$

$$\mathrm{Q_{i,j}} = \frac{1}{4}\sum\nolimits_{1}^{4}\mathrm{MD}(\mathrm{B_k}) \quad (2)$$

According to three directional motion homogeneity measures calculated by formula (2) each MB is classified into one of the following 5 kinds.

Class A： The motion of each direction is completely homogeneous

$$H_{i,j} < T_S \text{ and } V_{i,j} < T_S \text{ and } Q_{i,j} < T_S$$

Class B：The motion is complex and there is no obvious motion homogeneity in any direction

$$H_{i,j} > T_c \text{ and } V_{i,j} > T_c \text{ and } Q_{i,j} > T_c$$

Class C：Motion in horizontal direction is more likely to be homogeneous

$$H_{i,j} < V_{i,j} \text{ and } H_{i,j} < Q_{i,j}$$

Class D：Motion in vertical direction is more likely to be homogeneous

$$V_{i,j} < H_{i,j} \text{ and } V_{i,j} < Q_{i,j}$$

Class E：Motion in the horizontal or vertical direction is more likely to be homogeneous, when one of the following two conditions is satisfied

$$Q_{i,j} < H_{i,j} \text{ and } Q_{i,j} < V_{i,j}$$

$$V_{i,j} = H_{i,j} \text{ or } H_{i,j} = Q_{i,j} \text{ or } V_{i,j} = Q_{i,j}$$

The threshold T_S is set 0.1 and T_C is set 0.5. The value of T_S and T_C is set as shown in [14].

B. Candidate Intermodes Decision

According to the above MB classification results, for each MB, the candidate modes are tested in RDO, candidate modes of each class are summarized in the Table.1.

Table.1. Candidate inter modes for each MB class

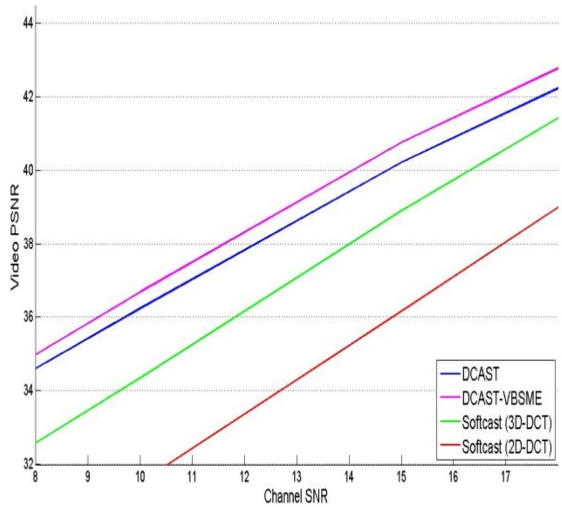
MB Class	Candidate Intermodes
A	16×16
B	16×16,16×8,8×16,8×8
C	16×16,16×8
D	16×16,8×16
E	16×16,16×8,8×16

We introduce fast inter frame mode decision algorithm in the ME module of DCAST. For flat area, big segmentation will be selected, for detailed area, small segmentation will be selected. Compared with the fixed block size motion estimation, DCAST with variable block size motion estimation can effectively reduce the block effect. Therefore, it can improve the quality of predicted frame.

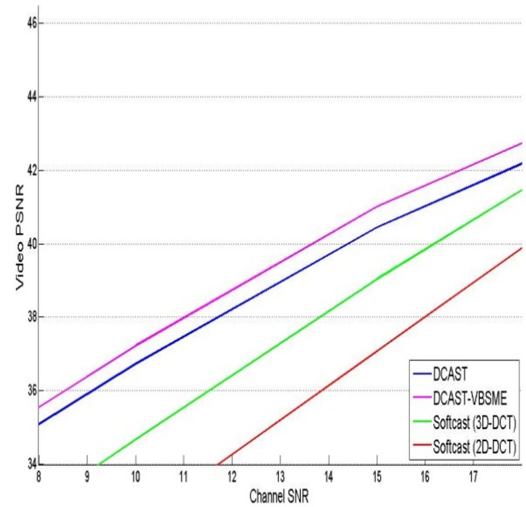
4. EXPERIMENTAL RESULTS

In the experiment, we test the performance of the DCAST with variable block size motion estimation in the video broadcast, we compare improved DCAST with DCAST [8], SoftCast [6] [7] and analog transmission. Analog transmission is to transmit video pixels directly over the channel. The experiment is to broadcast video to multiple users with different SNR.

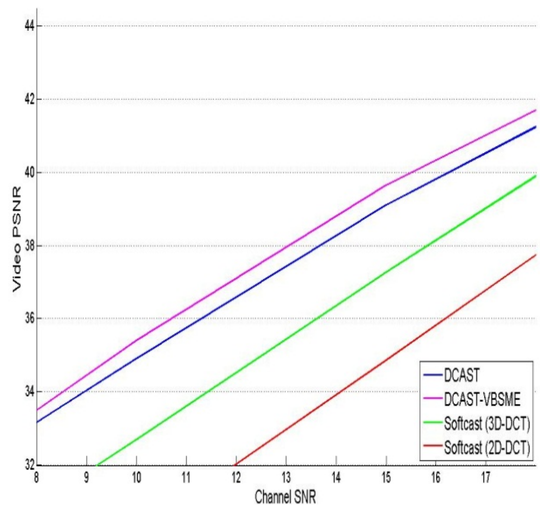
In experiments, the frame number of QCIF and CIF video sequence is 32, the test sequences are "foreman_qcif.yuv", "foreman_cif.yuv", "news_qcif.yuv"and "salesman_qcif.yuv", the video frame rate is 30Hz, the users' SNR range are 5-25, the search range of motion estimation is 32, mode decision method in motion estimation process is third part descriptive method, the GOP structure is IPPP, the reference frame of P frame is the previous frame. The experimental results are given in Fig.4. SoftCast based on the 2D-DCT [6] is 4-6 dB higher than the analog transmission, SoftCast-3D [7] is 2- 3dB higher than SoftCast-2D [6], DCAST is 0.8-2.2dB higher than SoftCast-3D [7], DCAST with variable block size motion estimation is 0.5dB higher than DCAST [8] with fixed block size motion estimation. The mode decision is fast mode decision algorithm, so improved DCAST does not increase too much computational complexity and is applicable for realtime video multicast like DCAST [8]. The visual quality comparison is given in Fig.5. Improved DCAST has better visual quality than DCAST and SoftCast.



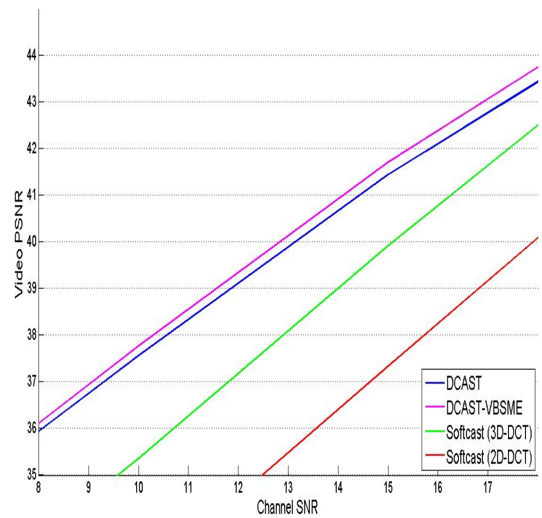
(a)Foreman_qcif



(b)Salesman_qcif



(c) news_qcif



(d) Foreman_cif

Fig.4. Performance comparison in AWGN channel

(a) Original frame (b)SoftCast-3D (c)DCAST (d)DCAST-VBSME

Fig.5. Visual quality comparison, the 6th frame of foreman_qcif.yuv, SNR=10dB

5. CONCLUSION

In this paper, we introduce an efficient inter mode decision method in the ME module of DCAST. On the coding side, complexity does not add too much and the predicted frame of the encoding side and decoding side is more accurate, thereby the reconstructed frame generated by MC and predicted frame is better, i.e. DCAST with variable block size motion estimation outperforms DCAST with fixed block size motion estimation.

6. ACKNOWLEDGEMNT

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] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.MobiCom ’11. New York, NY, USA: ACM, 2011, pp. 289 – 300. [Online]. Available: http://doi.acm.org/10.1145/2030613.2030646.

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

[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): de-sign and construction,” IEEE Trans. Inform. Theory, vol. IT-49, pp. 626 – 643, 2003.

[11] ——, “Distributed source coding using syndromes (DISCUS): design and construction,” in Proc. IEEE Data Compression Conf., 1999, pp. 158 – 167.

[12] I. Choi, J. Lee, and B. Jeon, “Fast coding mode selection with rate-distor-tion optimization for MPEG-4 Part-10 AVC/H.264,” IEEE Trans. Cir-cuits Syst. Video Technol. , vol. 16, no. 12, pp. 1557 – 1561, Dec. 2006.

[13] P. Yin, H. C. Tourapis, A. M. Tourapis, and J. Boyce, “Fast mode decision and motion estimation for JVT/H.264,” in Proc. IEEE ICIP’03,Sep. 2003, vol. 3, pp. 853 – 856.

[14] Z.Liu, L.Q. Shen, and Z.Y. Zhang, “An Efficient Intermode Decision Algorithm Based on Motion Homogeneity for H.264/AVC,” IEEE Trans. Cir-cuits Syst. Vide.