MOVIF: A LOWER POWER CONSUMPTION LIVE VIDEO MULTICASTING FRAMEWORK OVER AD-HOC NETWORKS WITH TERMINAL COLLABORATION

Ke Liang[1] *Zaoyang Gong*[2] *Changxi Zheng*[3] *Guobin Shen*[4*]

[1] Tsinghua University, Beijing, P.R. China, 100084
[2] Beihang University, Beijing, P.R. China, 100083
[3] Shanghai Jiaotong University, Shanghai, P.R. China, 200240
[4] Microsoft Research Asia, Beijing, P.R. China, 100080

Invited Paper

ABSTRACT

Live video multicasting over wireless ad hoc networks is a tough problem due to the restricted computational ability of the mobile devices and non-predictive networks status. In this paper, we propose MoViF, a lower power consumption multicasting framework for live video multicasting over ad-hoc networks. The proposed MoViF adopts distributed video coding (DVC) scheme and, as a result, enjoys all the benefits of DVC such as lightweight encoder and built-in error resilient capability. Moreover, MoViF manages to apply an elegant strategy to minimize the overall power consumption for all the receivers while improving their decoding quality by dynamically assign the tasks of aid information extraction to some intermediate powerful nodes in the multicast tree. Simulation results demonstrate that the optimal strategy can lower down the overall power consumption comparing to the random strategy.

1. INTRODUCTION

Video transmission over wireless networks has been an appealing application especially in the places that lack of available network infrastructure support. However, there exist many challenges for transmitting video over wireless networks, such as power consumption of terminals, mobility of terminals, and time varying characteristics of the channel (e.g., delay, bit error rate) etc. These challenges have in return imposed stringent requirements on the video coding schemes. Some basic requirements include 1) low computational complexity for both the encoder and the decoder; 2) adaptability to channel bandwidth; and 3) error-resilient ability for channel loss.

To cope with the last two requirement, scalable video coding [13] has been proposed . It has been proved that scalable video coding works effectively when used in conjunction with some simple error handling mechanisms such as forward error correction (FEC) and/or automatic repeat request (ARQ). Unfortunately, scalable video coding significantly increases the computational complexity to the already very complex encoding process. As a result, scalable video coding is not applicable to live video streaming in a mobile environment where the source is a mobile device that has limited processing power.

*The work is done in Microsoft Research Asia, Contact authors's email: jackysh@microsoft.com, liangk03@mails.tsinghua.edu.cn

On the other hand, distributed video coding (DVC), originated from Slepian-Wolf theorem [12] and Wyner-Ziv theorem [14] in the 1970s, suggests that an intraframe encoder - interframe decoder system can achieve similar compression efficiency of the traditional interframe encoder - decoder system. As a result, DVC brings in a simple encoder; what's more, DVC has built-in error-resilient capability in virtue of channel coding techniques. However, DVC schemes typically shifts the computational complexity from the encoder to the decoder, which is also not desirable for multicast applications over ad-hoc networks where the terminals are mobile devices. Fortunately, DVC also provides natural complexity scalability of the decoding processes: *normal decoding* and *enhanced decoding*. Normal decoding, in which the side information is generated by extrapolation, will leads to worse quality but at minimum complexity. Enhanced decoding, in which the more accurate side information is generated by motion compensation interpolation (MCI), will results in better quality at much increased complexity. Note that better quality can also be achieved at slightly increased complexity if the aid information (i.e., motion vectors between the more accurate side information and the previous decoded frame) is available to the normal decoding.

In this paper, we propose MoViF, a lower power consumption live video multicasting framework over ad-hoc networks. In MoViF we adopt the DVC scheme that is based on multilevel coset codes [4] and seek to improve the decoding quality of all the receivers while minimizing the overall power consumption for all the terminals. Besides all the inherited features of a DVC scheme, namely lightweight encoder and built-in error-resilient capability, the proposed MoViF has the feature that the heavy decoding burden is dynamically assigned to some intermediate powerful nodes in the multicasting tree. It is this feature that makes the MoViF a practical scheme. To elaborate a bit, in MoViF, we distinguish all the terminals into *weak nodes* and *strong nodes* which are capable of normal decoding and enhanced decoding, respectively. Furthermore, to reduce the power consumption, only some of the strong nodes are selected at run-time to perform enhanced decoding and the extracted aid information is sending to other nodes to help them improve the decoding quality. These selected nodes are specially called *helper nodes* in the rest of the paper.

The rest of the paper is organized as follows: In Section 2, we will review some typical video coding techniques for transmission over wireless applications. In Section 3, we will give an overview of the DVC schemes. We then present the proposed MoViF in

Section 4. In Section 5, we present and analyze the simulation results. Finally, Section 6 concludes the paper.

2. VIDEO CODING TECHNIQUES FOR TARNSMISSION OVER WIRELESS NETWORKS

Many video coding techniques for video transmission over wireless networks have been proposed in the past decade. Among them, the scalable video coding (SVC) is well-known and has been improved in depth. In SVC, video frames are encoded into more than one sub-stream, in terms of importance, quality, etc. In hierarchical layered video coding (HLC) [13], video is compressed into one base layer and multiple enhancement layers where base layer can be decoded independently but the enhancement layers must be decoded upon the corresponding low-level layer. In multiple description video coding (MDC) [11], video is compressed into multiple sub-streams (thus multiple descriptions) with different quality and rate. Each sub-stream can be decoded independently while multiple sub-streams can be jointly decoded and lead to higher fidelity. Progressive fine granularity scalable video coding [6] is similar to HLC, but the granularity of enhancement layer is finer and more flexible.

Scalable video coding can inherently adapt to the bandwidth fluctuation of wireless networks and the heterogeneity of QoS requests in multicast cases because the video can be decoded with a degraded quality if not receiving full sub-streams.

In order to transmit reliably, ARQ and FEC are used. Traditional ARQ method is not very suitable for the wireless networks because of the delay of package it brings. FEC, which uses redundant error-correction bits to combat the bit error of wireless link channel, is an optional substitute.

Some hybrid ARQ/FEC systems have proved that the hybrid ARQ/FEC error control is efficient in wireless environment [17]. However, high flexibility at the encoder results in high computational complexity, and the involved ARQ, FEC or the combination brings in a penalty to the compression efficiency. Therefore it may not be desirable for transmitting video over wireless networks, especially when the sender has low computational ability.

3. OVERVIEW OF DISTRIBUTED VIDEO CODING

Although theoretic principles of DVC was established over 30 years ago, the first practical DVC framework is proposed by Pradhan and Ramchandran in 1999 [9], where the trellis codes is used to partition the source codebook. Wang and Orchard [16] improved the performance by using embedded trellis codes. Since then more sophisticated channel codes were involved, such as turbo codes [3] and LDPC codes [15]to further enhance the coding efficiency.

In a DVC scheme, video sequences are encoded separately at the encoder. Side information is regarded as the output of the to-be-encoded frame through a "correlation channel". To appreciate this concept, let X denotes the to-be-encoded frame. Y denotes the decoded previous frame, which is treated as the side information to X. Z denotes the noise between X and Y. The source codebook is partitioned into cosets of a channel code according to the noise Z. The encoder transmits the coset index to the decoder. The decoder uses the coset index and the side information (Y) to reconstruct the frame.

The typical distributed video encoder consists of two components: one is the conventional intra-frame encoder of H.26x or MPEG, another is a Wyner-Ziv encoder emploiting different channel codes. To improve the coding efficiency of DVC schemes, transform-domain Wyner-Ziv encoder was proposed such that spatial redundancy can be exploited [1]. In a DCT-domain Wyner-Ziv encoder, the to-be-encoded frame is first divided into non-overlapped blocks. A blockwise DCT is then performed and followed by uniform quantization. After that, some low-frequency coefficients are compressed using a trellis code, and the rest coefficients are conventionally entropy coded or simply discarded. A checksum (CRC) of the quantized coefficients is also sent by the encoder to aid motion compensation at the decoder.

4. MOVIF : PROPOSAL FRAMEWORK

As mentioned above, DVC needs to perform very complex decoding if high coding efficiency is desired. However, mobile devices are typically power constrained. As a result, the most natural way is to ask the DVC encoder to send the bit streams to a powerful proxy to transcode the Wyner-Ziv stream to a convention video stream which is then streamed to other mobile receivers [10] [5]. However, this strategy requires a powerful proxy and infrastructure support and is not feasible for mobile ad hoc networks. In this paper, we propose to incorporate multiple powerful nodes in a collaborative way to share the burden of enhanced decoding tasks.

4.1. DVC scheme adopted in MoViF

The Wyner-Ziv coder in MoViF is based on our work presented in [7]. The paradigm is shown in Fig.1. We divide the video sequence into two partitions: *Key frames*, which are encoded by conventional H.263+ encoder, and *Wyner-Ziv frames*, which are encoded by Coset Encoder using multilevel coset codes. We choose the quantized DCT coefficients of the to-be-encoded frame as the source, the side information is quantized DCT coefficients of the previous frame. DCT operation increases the complexity as compared with pixel-domain DVC, but the complexity is affordable and the gain is significant. The side information can be improved by MCI at the decoder.

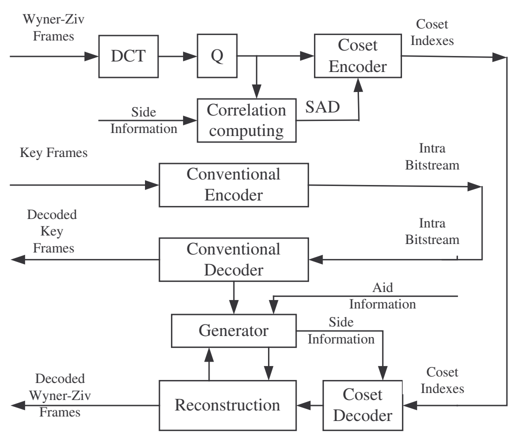

Fig. 1. Framework of MoViF Codec

At the encoder, if the frame is encoded as a *Wyner-Ziv frame*, the main encoding process is akin to that in [10]. The difference is that the source and the side information are both quantized DCT coefficients. We use Lagrangian cost function, J, to optimize the rate distortion performance : $J = D + \lambda R$, where λ is a nonnegative real number, R is the rate, and D is the expected distortion.

Because low-frequency DCT coefficients contain most of information, we transmit m low-frequency DCT coefficients to the decoder. To minimize J, an appropriate m should be selected according to the SAD (sum of absolutely difference) of the two blocks. The larger SAD of two blocks, the larger m should be selected.

At the decoder, the coset decoder decodes the low frequency coefficients of the blocks using the coset indexes without error, the high frequency coefficients of the blocks are stuffed by the co-located coefficients in the side information. In our implement, the quantized coefficients of the previous frame are used directly as the side information. Note that if the side information is not corrected by MCI, then the IDCT operation can be saved and the computational complexity will be lower than those in [1] and [10] where the side information is generated by extrapolation.

4.2. Multicast tree formation in MoViF

As previously mentioned, for a DVC scheme, in order to obtain high coding efficiency, the decoder needs to perform MCI to provide a more accurate side information. Unlike other proposal that adopts a powerful proxy, we seek to leverage some powerful mobile terminals in a collaborative way so that the overall processing burden is minimized.

In MoViF, we first distinguish the mobile terminals into weak nodes and strong nodes, where strong nodes are capable of MCI. The strong node will help other nodes (both strong nodes and weak nodes) by transmission the aid information (basically the motion vectors between the side information generated by MCI and the previous frame). There strong nodes are specially called *helper nodes*. The assignment of MCI to help nodes is dynamically performed.

Obviously, there are two possibilities in forming the multicast tree. The first method is to construct two separate multicast trees: one for normal bit stream and one for the aid information. The second method is to simply construct a shared multicast for both the normal bit stream and the aid information. For easy synchronization, we adopt the second method. The detailed tree construction procedure is as follows: first of all, identify all the strong nodes; secondly, select helper nodes from the strong nodes and build a multicast tree among helper nodes as the backbone; and finally, connect the remaining weak nodes to the backbone. In order to balance the workload among helper nodes, we perform dynamical assignment of MCI of different frames to different helpers. The whole process is elaborated below.

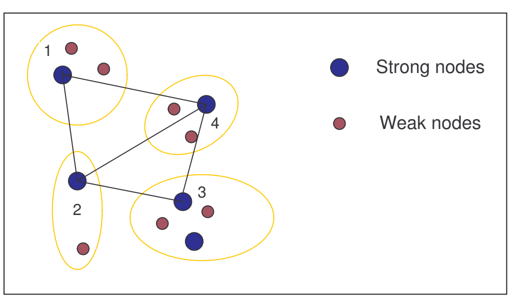

Fig. 2. Node clusters, where head of each clusters are helper nodes

In Fig. 2, there are five strong nodes and we identify four of them as helper nodes because either of the two strong nodes in cluster three can cover the whole cluster. To avoid duplicated MCI operation, only one of them is selected to join the backbone. To exploit the parallelism and achieve load balance among helper nodes, we may construct multiple multicast trees originating from different helper nodes.

Now there are also two possible ways to assign the MCI task for each incoming Wyner-Ziv frame to helper nodes (and the corresponding multicast trees). For a random selection strategy, we can randomly select the idle (not doing MCI) helper node. However, to lower down the overall power consumption, an ideal strategy should select the helper node that has minimum number of packet forwarding (in other word, minimum intermediate nodes) which is the most important cost factor in transmission over ad-hoc networks. Back to the example shown in Fig. 2, we can construct four multicast trees (Fig. 3), one for each helper node. When the four strong nodes are all idle, we select the second or the fourth node as the helper node because less packet forwarding would incur. If they have the same processing power, they will have the same priority to do MCI. Otherwise, the more powerful node has a higher priority. The general algorithm is as follows.

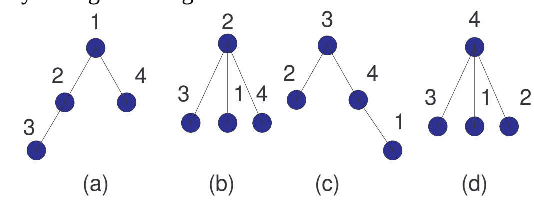

Fig. 3. Multiple multicast trees for aid information transmission: the sources are *helper nodes*

Finding a multicast tree with minimum intermediate nodes is the set cover problem (called maximum leaf spanning tree problem in mathematic field) which is NP-complete [8]. Thus we use a greedy method to construct the multicast tree (originate at a specific helper node): in the first stage, link to as many other helper nodes as it can.[1] In the second stage, for each child, count the number of links to other remaining helper nodes, and select the one with largest number of links as intermediate nodes. The process is performed recursively until all the helper nodes are connected to the tree. After generating the multicast tree for each helper node, the priority of assigning MCI of an incoming Wyner-Ziv frame is determined according to the number of intermediate nodes in the multicast tree: the fewer intermediate nodes, the higher priority the helper node has.

5. SIMULATION RESULTS

The rate distortion performance and error-resilient ability of DVC are reported in the related works [5] [2]. In this paper, we simulate the average power consumption of the overall receivers using different strategy of selecting the helper node. In our simulation, there are seven helper nodes among all the receivers. Maximum hops between any two helper nodes is set to two.

The topology of helper nodes $(n_0, n_2, \cdots, n_6)$ in our simulation is shown in Fig. 4. To demonstrate the heterogeneity among mobile terminals, we assume the seven helper nodes have different processing power and they are listed in a descending order as: $n_3, n_1, n_5, n_0, n_6, n_4, n_2$. We use the time of doing MCI to quantify the processing power of each node. The MCI time for helper node n_0 through n_6 is: $4.4t$, $3t$, $7.3t$, $2t$, $6.3t$, $4t$, $5.7t$, where t means the time interval between two Wyner-Ziv frames. Totally 50 Wyner-Ziv frames are used in our experiments.

[1] We allow multiple hops, but in the path there should be no other helper nodes.

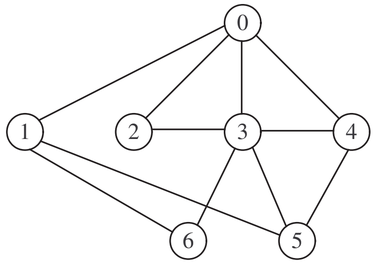

Fig. 4. Topology of helper nodes in the experiments

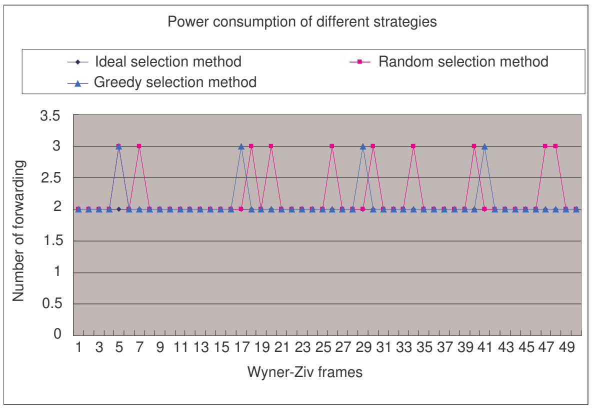

Fig. 5. Power consumption of different strategies

As shown in Fig. 5, the random selection strategy needs more packet forwarding, thus consumes more power. The proposed greedy method can achieve a similar performance to the ideal method and is much more efficient than the random selection strategy.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed a lower power consumption multicasting framework for live video multicasting over ad-hoc networks - MoViF. To enhance the overall quality of all the receivers, we choose some helper nodes from strong nodes to generate and transmit aid information to the others. To minimize the average power consumption of the overall receivers, the average number of packet forwarding should be minimized. Finding a multicast tree with minimum intermediate nodes is NP-complete, therefore, we propose a greedy method by selecting the helper node which is idle and has maximin links to other helper nodes.

The proposed MoViF is far from complete and needs to be enriched. For example, the tradeoff between the delay and the power consumption should be concerned in our future work.

7. REFERENCES

[1] A. Aaron, S. Rane, E. Setton, and B. Girod. Transform-domain wynercziv codec for video. In *SPIE Visual Communications and Image Processing Conf.*, 2004.

[2] A. Aaron, S. Rane, R. Zhang, and B. Girod. Wyner-ziv coding for video: Applications to compression and error resilience. In *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2003.

[3] A. Aaron, R. Zhang, , and B. Girod. Wynercziv coding of motion video. In *Asilomar Conference. Signals and Systems*, 2002.

[4] G. D. Forney, J. Trott, and M. D. S.-Y. Chung. Sphere-bound-achieving coset codes and multilevel coset codes. *IEEE Transactions on Information Theory*, 46:820–850, May 2000.

[5] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. Distributed video coding. In *Proceedings of the IEEE, Special Issue on Video Coding and Delivery*, volume 93, pages 71–83, Jan. 2005.

[6] W. Li. Overview of fine granularity scalability in mpeg-4 video standard. *IEEE Transaction on circuits and systems for video technology*, 11, Mar. 2001.

[7] K. Liang, L. F. Sun, and Y. Z. Zhong. Adaptive distributed video coding for video applications in ad-hoc networks. In *Accepted by PCM05*.

[8] H. Lim and C. Kim. Multicast tree construction and flooding in wireless ad hoc networks. In *Proceedings of the 3rd ACM internationalworkshop on Modeling, analysis and simulation of wireless and mobilesystems table of contents*, 2000.

[9] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (discus): Design and construction. In *Proceedings of the Data Compression Conference (DCC99)*, 1999.

[10] R. Puri and K. Ramchandran. Prism: a new robust video coding architecture based on distributed compression principles. In *Allerton Conference. Communication, Control, and Computing*, 2002.

[11] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T.Orchard, and R. Puri. Multiple description video coding using motion-compensated temporal prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 12:193–204, Mar. 2002.

[12] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19:471–480, July 1973.

[13] D. Wu, Y. W. Thomas, and Y. Q. Zhang. Scalable video coding and transport over broad-band wireless networks. In *Proc. of IEEE*, Jan. 2001.

[14] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, Jan. 1976.

[15] Q. Xu and Z. Xiong. Layered wyner-ziv video coding. In *Proc. VCIP'04: Special Session on Multimedia Technologies for Embedded Systems*, San Jose, CA, 2004.

[16] X.Wang and M. Orchard. Design of trellis codes for source coding with side information at the decoder. In *Proceedings of the Data Compression Conference (DCC01)*, 2001.

[17] F. Zhai, Y. Eisenberg, T. N. Pappas, R.Berry, and A. K. Katsaggelos. Rate-distortion optimized hybrid error control for real-time packetized video transmission. In *Proceeding IEEE International Conference Communication*, 2004.