A Power-efficient Cloud-based Compressive Sensing Video Communication System

Mengsi Wang, Song Xiao, Lei Quan and Qunwei Li
State Key Lab. of ISN, Xidian University, Xi'an, China
E-mail: xiaosong@mail.xidian.edu.cn Tel: +86-029-88201956

Abstract—**Mobile devices performing video coding and streaming over wireless communication networks are limited in energy supply. In this paper, a power-efficient fast cloud-based Compressive Sensing (CS) video communication system framework is proposed, which shifts the heavy computational burden from mobile devices to cloud thus the operational lifetime of the mobile devices can be prolonged. Firstly, a CS-based video encoder with partial-bidirectional motion inter-frame prediction is proposed to greatly reduce the complexity of video encoding, and then a BER-based adaptive CSEC scheme is proposed to combat the data loss in wireless networks. Meanwhile, cloud platform is employed, which has strong ability of computation and processing to solve the problem of high complexity resulted from CS reconstruction. Simulation results show that the energy consumption of the proposed system can be substantially reduced. Under the condition of the same reconstructed video quality (SSIM=0.9), the proposed scheme only takes about 15% percent power consumption of the state-of-the-art H.264 encoding with conventional FEC-based transmitting scheme.**

I. INTRODUCTION

Nowadays, video processing and transmission has been an important integrated part of mobile devices which allow users to share videos through wireless networks. Since wireless mobile devices are battery-operated and the utility of computing and communication is directly impacted by the lifetime of battery, it is important to conserve the power in video processing. However, current video coding schemes consume a lot of power since they use a high complexity encoder to achieve compression and decrease the number of bits in video streams, and most prior work on power conserving in video processing has focused on designing low power integrated circuits, this raises the question of whether it is possible to decrease the power consumption directly from the scheme of video sampling and encoding?

Fortunately, the emerging field of Compressive Sensing can offer an alternative to traditional video encoders by enabling the image system to sense and compress the data simultaneously at very low computational complexity. It is a novel theory which states that if a signal is sparse in a transform domain, then under a certain condition, it can be precisely reconstructed from only a small set of measurements [1] [2]. Inspired by the advantages of CS theory, researchers have done a lot of work to make it more sufficient to be applied in the area of image and video processing. Lu proposed a block based compressed sensing method in [3], and the computational and memory complexity of image sampling can be reduced. In [4], James E. Fowler et al. proposed a multiscale block Compressed Sensing with smoothed projected Landweber reconstruction algorithm. In [5], Scott Pudlewski et al. studied the effect of key video parameters on the received quality of CS-encoded images, and in [6] they proposed the Compressive Distortion Minimizing Control (C-DMRC) which leverages the estimated received video quality as the basis of the rate control decision.

However, the above researches are mostly focused on developing the front end of CS-based communication system to improve the quality of CS images or video sequences. Since CS reconstruct algorithms are of high complexity, their applicability is severely restricted, especially in mobile devices with limited energy supply. While in this paper, the CS video communication system we proposed is aimed at applying in mobile devices and substantially reducing the power consumption of video communication. In the system we use Cloud-Platform to do CS reconstruction, thus the heavy burden of CS reconstruction can be offloaded and the advantages of the proposed low complexity CS-encoder can be fully utilized. Also a BER-based adaptive CSEC scheme to handle the problem of data loss in realistic wireless channels is proposed.

The rest of this paper is organized as follows: in Section II, we introduce the whole system, and describe the CS-based encoder and the BER-based adaptive CSEC in detail. In Section III, we evaluate the reconstruct quality of the CS-based video, and compare the total energy consumption with that of H.264-encoder using the energy calculation model we proposed.

II. SYSTEM MODEL

A. System Overview

The system proposed is illustrated in Fig. 1, which mainly includes video senders, cloud platform and receivers. Video senders use mobile devices with CS-based encoder which will be introduced in Subsection B, to compress the images of a video sequence by taking a small number of linear measurements. Since CS combines sampling and compression

This work is supported by NSFC (No.61372069), National Defense Pre-research Foundation, SRF for ROCS, SEM (JY0600090102), "111" project (No.B08038) and the Fundamental Research Funds for the Central Universities.

together and enables a potentially large reduction of sampling rates, the complexity of CS-based video can be very low. Then the CS-encoded video stream is transmitted through wireless channel to the cloud platform to be reconstructed. The heavy burden of CS reconstruction can be offloaded since the cloud platform can enhance the computing capability of mobile systems and potentially save energy for mobile users [7]. In order to handle the data loss within wireless networks, we use CS erasure code (CSEC) [8] to compensate for the quality of reconstructed video. In part C, we will introduce a CSEC scheme which can adapt to the bit error rate (BER) of realistic wireless channel. Although the H.264 encoder is of high complexity, its decoder is simple and of low complexity. Therefore, the CS-reconstructed video is transcoded to H.264-encoded video streams in the cloud platform and finally transmitted to the receiver. It means if the communication parties have CS-encoder, H.264-decoder and an access to the cloud platform, video information can be changed freely with less consumption of energy. The advantages of the proposed system are fourfold: CS-based encoder is of low complexity, CS-encoded video information is robust to channel erasures, cloud platform has strong computation and processing ability and the H.264 decoder is of low complexity.

B. CS-based Encoder

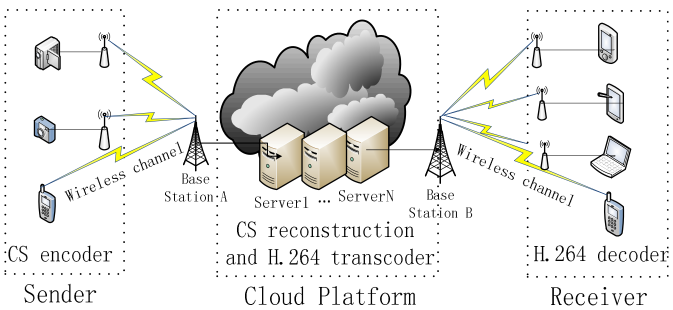

Fig. 1 Power-efficient video transmission system.

In this part, we introduce the video encoder of the above system. First, the encoder receives the raw samples from the camera of mobile devices, let $x_1, x_2, \ldots$ indicate a video sequence, where each $x_i \in \mathbb{R}^N$ is an N-pixel image of it. Then the encoder generates compressed video frames by taking M measurements $y_i = \Phi_i x_i$ from every image, where Φ_i, x_i and y_i denote the measurement matrix, the matrix of image pixels and the image samples respectively.

In our system, we use the scrambled block Hadamard ensemble of size 32×32 as measurement matrix, and the matrix x_i has been randomly reordered and shaped into $32\times(N/32)$ matrix where N is the number of pixels in the image. Then M samples are randomly chosen from x_i and transmitted to the cloud platform.

In our encoder design, we exploit inherent spatial and temporal redundancies in a video sequence by dividing the video frames into two types, i.e. I frame and P frame. I frame is the reference image which is encoded using CS as a static image. P frame is encoded by utilizing the temporal redundancy between adjacent frames. Since the decoder will use the recovered I frame image as reference frame, if there are transmission errors in I frame, the corresponding quality of recovered P frames will degrade rapidly. Thus, different from the GOP structure described in [6], i.e. forward inter-frame motion prediction, we propose a partial-bidirectional inter frame motion prediction structure to improve the reconstruction quality of CS-encoded video. As depicted in Fig. 2, among the P frames, frame 1 and frame 3 are constructed by taking algebraic difference with its closest I frame (i.e. frame 0 and frame 4 respectively), while frame 2 is constructed by taking algebraic difference with the average of the two reference frames, then all of the difference vectors are sampled using CS scheme the same as I frame. Since there is constantly a large amount of redundancy between adjacent frames, the difference of the frames will be much sparser and more compressible than the original image, which means it can be reconstructed with less samples.

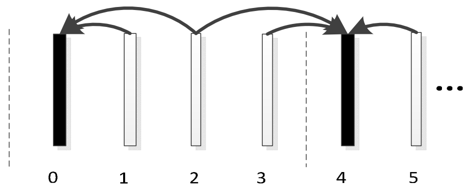

Fig. 2 Frame structure, GOP size is 4.

C. BER-based Adaptive CSEC in Realistic Channel

CSEC means compressive oversampling in an incoherent measurement basis, which is proved to be efficient for handling missing data in erasure channels [8]. Based on this foundation, the authors of [5] proposed an adaptive parity scheme that drops samples in error, making the error channel to be erasure channel, thus the CSEC scheme can be applied. However [5] didn't consider the performance of AP scheme under realistic wireless channels. Since errors at the output of realistic wireless channels tend to occur in bursts and the bit error rate (BER) is more variable. In this part, we propose a BER-based adaptive CSEC scheme especially for realistic wireless channels. Adaptive error detecting schemes are used in different channel conditions, and CSEC is combined to compensate for the quality degradation resulted from channel errors. We will first estimate the effect of AP performance with bursty channels using the classical Gilbert-Elliott (GE) model, which can describe many wireless channels accurately. Then compare it with the performance of CRC-4 under same channel conditions.

To model GE channel, a two-state discrete-time binary Markov chain is used. The state of the channel is either "good" or "bad" at a time and the probability of error occurs within the state is assumed to be $P_g = 0$ and $P_b = 1$. P indicates the average BER of the bursty channel, and the burst size is expected to be $b = 8$ bits. The probability of transition from one state to another is computed using the standard relationship $P_{bg} = 1/b$ and $P_{gb} = P/\left(b(1-P)\right)$.

The image of the video sequence is sampled using CS scheme described in subsection B, and quantized with

5bit/sample [5]. Then the video information is encoded by adding 1 parity bit per 8 bits of data and transmitted over GE channel. At the receiver, the detected incorrect samples are discarded. We use the Structural Similarity Index (SSIM) to evaluate the reconstruction quality since it is more accurate to measure error than the widespread PSNR, which has been shown to be inconsistent with human eye perception [9]. The red curve in Fig.3 is the reconstructed SSIMs of CS image with AP scheme. It can be seen when BER is over 10^{-4}, AP scheme is no more efficient to detect errors, thus making the recovered image quality degrade rapidly. When we add the same amount of redundancy by using CRC-4 to detect error, which means encoding the video information by adding 4 CRC bits per 32 bits of data, the reconstructed quality which is shown as the blue curve, is highly improved compared to the AP scheme when BER is higher than 10^{-4}. Additionally, if we reduce the redundancy by adding 4 CRC bits per 40 bits of data, the reconstructed quality which is shown as the purple curve, still outperforms AP scheme. Therefore, the BER-based adaptive CSEC scheme is operated as: when BER is lower than 10^{-4}, AP scheme is used to detect channel errors, while BER is higher than 10^{-4}, CRC-4 is used instead; then according to the feedback of total data loss, CSEC is used in the encoder to compensate for the reconstruction video quality.

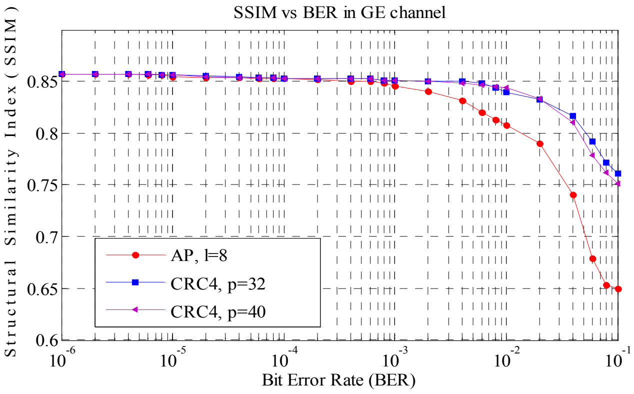

Fig. 3 AP and CRC encoding for variable bit error rates.

III. PERFORMANCE EVALUATION

A. *Evaluation of CS-reconstructed Video*

In this part, we compare the recovery results of the proposed partial-bidirectional motion prediction scheme with that described in [6] using identical reconstruct algorithm at same compression ratio. To recover the CS-encoded video information, we use the minimizing total variation (TV) algorithm, which is more appropriate to recover 2D images and always provides better reconstruction performance. The high computation cost, which is the main disadvantage of TV algorithm, can be offloaded to the cloud platform within our system.

Fig. 4 shows the instantaneous SSIM value comparison of the proposed partial-bidirectional prediction scheme and the forward prediction scheme for the standard test video sequence "Foreman" at the compression ratio of 4:1. From the figure we can see, the quality of reconstructed P frames of CS-PartBi is always much higher than that of CS-Forward, thus the proposed encoder is more desirable.

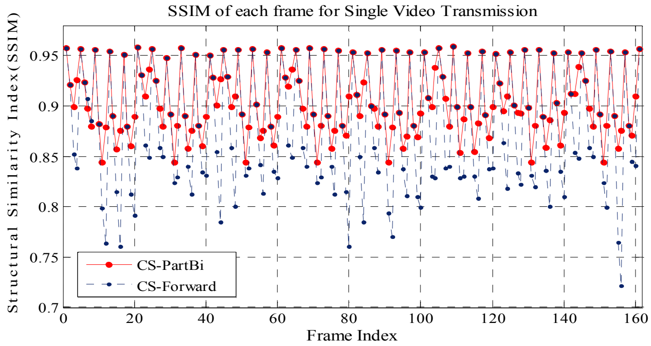

Fig. 4 SSIM of each frame for video transmission

B. *Energy Comparison with H.264*

In this part, we compare the energy consumption of the proposed CS-based encoder with that of H.264-encoder, which represents the state of the art among current video compressing techniques. Though H.264-encoder has excellent compression efficiency, it is at the sacrifice of very high computational power for encoding and high sensitivity to transmission errors [10]. The compared energy consists of two parts, i.e. P_E and P_T , which represent the power consumption for encoding and transmission respectively. The transmission energy is included because the H.264-encoded information also needs transmitting to the server when the two communication parties are far away that networks like Bluetooth can't be used.

As for the power consumption for encoding, we calculate it by running the software on computer and then scaling the results for the clock speeds of the mobile devices. The encoding of the two encoders was implemented using Microsoft Visual Studio 2010 on an Intel i3-2120 3.3GHz processor running operation system Windows 7 Ultimate SP1. The H.264-encoder we used is x264[11], which has been widely used in many popular applications like ffdshow, ffmpeg and MEncoder. It is important to note that although H.264 on mobile devices is implemented on dedicated chips, the proposed CS-based encoder has the overwhelming superiority of power consumption if it is also made into dedicated chips since it is significantly much simpler.

We take both the low BER and high BER channel conditions into consideration. When BER is low, we use RS (63, 57) as channel coding scheme for H.264-encoded video information, and AP scheme to detect errors of CS-encoded video information. When BER is relatively high, RS (255, 239) and CRC schemes are used for the two encoders respectively. We compare the energy consumption of the two encoders at the same reconstruction quality. In Table I we show the results of H.264-encoded and CS-encoded video sequence "Foreman" under different channel conditions when the target reconstruction quality is SSIM=0.9 ($PSNR \approx 34.6dB$), which indicates visually acceptable reconstructed video quality. We recorded the processor load of the computer when encoding. Also, we recorded the total time of encoding 600 frames of the video sequence "Foreman" for H.264-encoder, and 24000 frames of the same video sequence for CS-based encoder. Then for each encoder, we calculated the average CPU utilization per frame as

$$CPU utilization = \frac{Processor Load}{fps} = \frac{Processor Load}{Frame Amount / Total Time} \quad (1)$$

And we get the CPU utilization per frame of mobile device according to the clock speed ratio. That is,

$$CPU\, utilization_{mobile} = CPU\, utilization \times \frac{Clock\, Speed_{platform}}{Clock\, Speed_{mobile}} \quad (2)$$

Then, the energy consumption for encoding can be calculated as

$$P_E = P_c \times CPU\, utilization_{mobile}, \quad (3)$$

where P_c indicates the power mobile system consumes for computing.

As for the energy consumption for transmission, it can be estimated as

$$P_T = P_t \times \frac{D}{B}, \quad (4)$$

where P_t is the power that mobile system consumes for transmitting data, D is the data volume and B is the network transmission bandwidth.

We select IEEE 802.11b as the medium access control (MAC) layer protocol, and the data rate can be reached at $B = 5Mbps$. The advantage of the above model is that simulation on a certain mobile device is not needed, and the energy consumption can be calculated as long as we know some parameters specific to the mobile system. For example, Intel Xscale processor, which is widely used in many PDAs and smart phones, has the following values: $P_c \approx 0.9W$, $ClockSpeed \approx 400MHz$, $P_t \approx 0.8W$. So with the data in Table I and formula (1) (2) (3) we calculated the encoding energy consumption of the both encoders: $P_{E\_h.264} = 271.43mJ$ and $P_{E\_cs} = 6.76mJ$. When BER is 10^{-5}, using (4) we calculated the transmission energy consumption of the two encoders: $P_{T\_H.264} = 1.53mJ$ and $P_{T\_cs} = 34.75\,mJ$ .So we can get the total energy consumption of the two encoders: $P_{H.264} = 272.93mJ$ and $P_{cs} = 41.51mJ$. Thus the total energy consumption of the CS-based encoder takes about 15.2% of the H.264-encoder. And similarly when BER is 10^{-2} , the result is about 15%.

TABLE I
PARAMETERS OF H.264-ENCODED AND CS-ENCODED VIDEO SEQUENCE

BER	10^{-5}		10^{-2}	
SSIM	0.90		0.90	
encoder	H.264	CS	H.264	CS
Channel coding	RS(63,57)	AP	RS(255,239)	CRC
Frame amount	600	24000	600	24000
Total time	85.5s	86.2s	85.9s	81.2s
fps	7.0	278.4	6.98	295.6
Processor load	24.99%	24.92%	24.99%	24.92%
CPU utilization/frame	3.57%	0.089%	3.58%	0.084%
Data transmitted/frame	1.22KB	27.8KB	1.17KB	27.8KB

From the results we can see in spite of its higher transmitted data, the CS-based encoder within the proposed multimedia video communication system can achieve substantially reduction of energy consumption compared with that of H.264 encoder at the same reconstruction quality. Furthermore, if quantization efficiency is improved and appropriate coding scheme is used after quantization, the transmitted CS-encoded data can be further reduced. We do not cover it in this paper and leave it as future work.

IV. CONCLUSIONS

In this paper, a power-efficient Compressive Sensing video communication system for cloud-based mobile network is proposed to reduce the energy consumption of mobile devices. The BER-based adaptive CSEC scheme is proposed to combat data loss of realistic wireless channel. The simulation results show that in terms of the same reconstruction quality (SSIM=0.9), the total energy consumption of video processing and transmission using the proposed system can be greatly reduced and only takes about 15% of the H.264-encoder.

REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust Uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information theory, IEEE Trans.*, vol.52, no.2, pp. 489-509, Feb.2006.

[2] D. Donoho, C.D., "Compressed Sensing," *Information theory, IEEE Trans.* , vol.52, no.4, pp. 1289-1306, Apr.2006.

[3] L. Gan, "Block compressed sensing of natural images," *In Proceedings of the 15th International Conference on Digital Signal Processing,* Washington D.C., USA, pp.403-406, Jul. 2007.

[4] J. E. Fowler, Sungkwang Mun and Eric W.Tramel, "Multiscale Block Compressed Sensing with Smoothed Projected Landweber Reconstruction," *In proceedings of 19th European Signal Processing Conference (EUSIPCO),* Aug 29-Sep 2, 2011.

[5] Pudlewski, S. and Melodia,T., "On the Performance of Compressive Video Streaming for Wireless Multimedia Sensor Networks," *2010 IEEE International Conference on Communications (ICC)*, pp.1-5, May23-27, 2010.

[6] Pudlewski, S., Prasanna, A. and Mwlodia, T., "Compressed Sensing Enabled Video Streaming for Wireless Multimedia Sensor Networks," *Mobile Computing, IEEE Trans.,* vol.11, pp.1060-1072, 2012.

[7] Kumar, K. and Yung-Hsiang Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?," *Computer,* vol. 43, pp. 51- 56, Apr. 2010.

[8] Chakraborty, S., Zahedi, S., Younghun Kim, Srivastava, M.B., "Compressive Oversampling for Robust Data Transmission in Sensor Networks,"INFOCOM, *2010 Proceedings IEEE,* pp.1-9, March 2010.

[9] Z.Wang, A.Bovik, H. Sheikh and E.Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image processing, IEEE Trans.,* vol.13, no.4, pp.600-612, Apr. 2004.

[10] J.Ostermann, J.Bormans, P.List, D.Marpe, M.Narroschke, F.Pereira, T.Stockhammar, and T.Wedi, "Video coding with H.264/AVC: Tools, performance, and complexity," *IEEE Circuits and System Magazine,* vol. 4, no.1, pp. 7-28, April 2004.

[11]X264, Available: http://www.videolan.org/developers/x264.html