Focused Technology Advances Series

A Case for Compressive Video Streaming in Wireless Multimedia Sensor Networks

Tommaso Melodia, Scott Pudlewski, University at Buffalo, SUNY, USA
tmelodia@eng.buffalo.edu, smp25@buffalo.edu

Introduction

Wireless Multimedia Sensor Networks (WMSN) [1] are self-organizing wireless systems of embedded devices deployed to retrieve, distributively process in real-time, store, correlate, and fuse multimedia streams originated from heterogeneous sources. WMSNs will enable new applications including multimedia surveillance, storage and subsequent retrieval of potentially relevant activities, and person locator services.

In recent years, there has been intense research and considerable progress in solving numerous wireless sensor networking challenges. However, the key problem of enabling real-time quality-aware video streaming in large-scale multi-hop wireless networks of embedded devices is still open and largely unexplored. In fact, traditional video streaming systems based on transmitting predictively-encoded video through a layered communication protocol stack suffer from *high complexity at the encoder* and *low resiliency to channel errors*.

- **Encoder Complexity.** Predictive encoding requires complex processing algorithms, leading to high energy consumption [1]. New video encoding paradigms are needed to reverse the traditional balance of complex encoder and simple decoder, which is unsuited for WMSN. Recently developed *distributed video coding* [2] algorithms exploit the source statistics at the decoder, thus shifting the complexity at this end. While promising for WMSNs, most practical Wyner-Ziv codecs require end-to-end feedback from the decoder [3], which introduces overhead and delay. Furthermore, gains of practical distributed video codecs are typically limited to 2-5 dBs PSNR.
- **Limited Resiliency to Channel Errors.** In existing layered protocol stacks based on the IEEE 802.11 and 802.15.4 standards, video frames are split into multiple packets. If even a single bit is flipped due to channel errors, after a cyclic redundancy check, the entire packet is dropped at a final or intermediate receiver. This packet loss can lead to the video decoder being unable to decode an independently coded (I) frame, leading to the loss of the entire sequence of video frames that are dependent on the I frame. Instead, ideally, when one bit is in error, the effect on the reconstructed video should be unperceivable, with minimal overhead. In addition, the video quality should gracefully and proportionally degrade with decreasing channel quality.

Compressive Video Streaming for WMSN

Our preliminary investigation reveals that new cross-layer optimized networking protocols integrated with video encoders based on the recently proposed compressive sensing (CS) paradigm [4], [5] can offer a convincing solution to the aforementioned problems. However, as will become clearer in the following, this will require a careful rethinking of traditional wireless networking functionalities across multiple layers. Compressed sensing (aka “compressive sampling”) is a new paradigm that allows the recovery of signals from far fewer measurements than methods based on Nyquist sampling. In particular, the main result of CS is that a N-dimensional signal can be reconstructed from M noise-like incoherent measurements as if one had observed the $M/log(N)$ most important coefficients in a suitable base [6]. Hence, CS can offer an alternative to traditional video encoders by enabling imaging systems that sense and compress data simultaneously *at very low computational complexity for the encoder*. Image coding based on CS has been recently explored [7], [6]. So-called single-pixel cameras that can operate efficiently across a much broader spectral range (including infrared) than conventional silicon-based cameras have also been studied [8]. However, wireless networking protocols optimized for transmission of CS video, and their statistical traffic characterization, are substantially unexplored areas. In particular, in this position paper we show that CS-based image representation shows an inherent resiliency to random wireless channel errors that should guide and inform protocol design optimized for wireless video streaming in WMSNs.

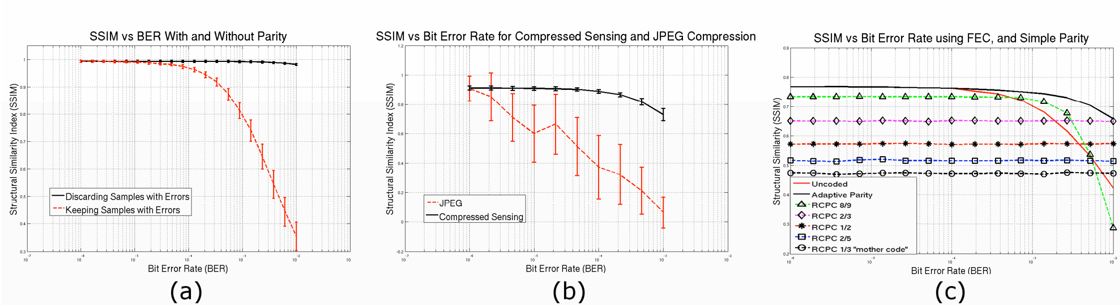

Fig. 1 Structural Similarity (SSIM) Index vs BER for (a) Reconstruction With and Without Incorrect Samples (b) CS vs JPEG images (c) Adaptive Parity

Effect of Channel Errors on CS Video

We conducted a preliminary investigation of the effect of channel errors on wireless networked CS images and video. To assess the impact of channel errors and interference on CS video quality, we evaluated the *Structural Similarity Index* (SSIM) [9] between the original and the encoded image for a standardized set of 25 images. We represented each frame of a quarter common intermediate format (QCIF) video by 8-bit intensity values, i.e., a grayscale bitmap. To satisfy the sparsity requirement of CS theory, the wavelet transform is used as a sparsifying base. The image is sampled using a scrambled block Hadamard ensemble [10], and recreated through GPSR [11]. In CS, the transmitted samples constitute a random, incoherent combination of the original image pixels. This means that, unlike traditional wireless imaging systems, in CS no individual sample is more important for image reconstruction than any other sample. Instead, *the number of correctly received samples* is the only main factor in determining the quality of the received image. Hence, a peculiar characteristic of CS video is its *inherent and fine-grained spatial scalability*. The video quality can be regulated at a much finer granularity than traditional video encoders, by simply varying the number of samples per frame. Also, a small amount of random channel errors does not affect the perceptual quality of the received image *at all*, since, for moderate BERs, the greater sparsity of the "correct" image will offset the error caused by the incorrect bit. This is demonstrated in Fig. 1(a). For any BER lower than 10^{-4}, there is *no noticeable drop in the image quality*. Up to BERs lower than 10^{-3}, the SSIM is above 0.8, still an indicator of good image quality. CS image representation is completely *unstructured*: this fact *makes CS video much more resilient than existing video coding schemes to random channel errors*. This simple fact has obvious, deep, consequences on protocol design for end-to-end wireless transport of CS video.

This inherent resiliency of compressed sensing to random channel bit errors is even more noticeable when compared to traditional compression schemes. Figure 1(b) shows the average SSIM of 25 images transmitted through a wireless channel with varying BER. The quality of CS-encoded images degrades gracefully as the BER increases, and is still very high for BERs as high as 10^{-3}. Instead, JPEG-encoded images very quickly deteriorate. This is visually emphasized in Fig. 4, which shows a frame from a surveillance camera at the University at Buffalo encoded with CS and JPEG and transmitted with end-to-end bit error rates of 10^{-5}, 10^{-4}, and 10^{-3}, respectively. The difference is stunning - the effect of bit errors is much more disruptive for structured data like JPEG-encoded images. The effect on predictively-encoded video is even worse, since even low bit error rates can lead to the loss of I frames, causing the decoder to be unable to decode long sequences of frames that depend on the I frame.

Our preliminary investigation also reveals also that while forward error correction (FEC) is not beneficial for low to moderate values of BER up to 10^{-2}, the perceptual quality of CS images can be improved by dropping errored samples that would contribute to image

Fig. 2 Surveillance Image with CS (above) and JPEG (below) for BER (a) 10^{-5} (b) 10^{-4} (c) 10^{-3}.

reconstruction with incorrect information (Fig. 1(c)). This calls for a data protection strategy based on using even parity on a predefined number of samples, which are all dropped at the receiver or at an intermediate node if the parity check fails. This is particularly beneficial in situations when the BER is still low, but too high to just ignore errors (above 10^{-5}). We have analytically determined the optimal number of samples to be jointly protected, for a given BER and quantization level, based on which the encoder can adaptively regulate the level of protection to track the end-to-end BER. This simple strategy is shown in Fig. 1(c) to considerably improve the received video quality compared to protecting the CS samples FEC with different levels of protection using rate-compatible punctured codes (RCPC) with ¼ mother codes.

Based on the encouraging results of this preliminary investigation, we are currently conducting a cross-layer analysis of the impact of functionalities handled at all layers of the communication protocol stack on the perceived video quality for competing CS-encoded video streams. In addition, we are developing a tool to perform network simulations of CS video applications. The tool will allow generating CS video traces and evaluating the impact of CS video on network protocols over the widely used *ns-2* simulator.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, “A Survey on Wireless Multimedia Sensor Networks,” *Computer Networks (Elsevier)*, vol. 51, no. 4, pp. 921—960, Mar. 2007.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed Video Coding,” Proc. of the IEEE, vol. 93, no. 1, pp. 71—83, January 2005.

[3] A. Aaron, S. Rane, R. Zhang, and B. Girod, “Wyner-Ziv Coding for Video: Applications to Compression and Error Resilience,” in *Proc. of IEEE Data Compression Conf. (DCC)*, Snowbird, UT, March 2003, pp. 93-102.

[4] D. Donoho, “Compressed Sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289—1306, Apr. 2006.

[5] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489—509, Feb. 2006.

[6] J. Romberg, “Imaging via Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14—20, 2008.

[7] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, “Compressive imaging for video representation and coding,” in *Proc. Picture Coding*

Symposium (PCS), April 2006.

[8] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, April 2004.

[10] L. Gan, T. Do, and T. D. Tran, "Fast Compressive Imaging Using Scrambled Block Hadamard Ensemble," *Preprint*, 2008.

[11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal of Selected Topics in Signal Processing,* vol. 1, no. 4, pp. 586–598, 2007.

Tommaso Melodia is an Assistant Professor with the Department of Electrical Engineering at the University at Buffalo, The State University of New York (SUNY), where he directs the Wireless Networks and Embedded Systems Laboratory. He received his Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2007. He had previously received his Laurea (integrated B.S. and M.S.) and Doctorate degrees in Telecommunications Engineering from the University of Rome La Sapienza, Rome, Italy, in 2001 and 2005, respectively. He coauthored a paper that was recognized as the *Fast Breaking Paper in the field of Computer Science* for February 2009 by Thomson ISI Essential Science Indicators. He is an Associate Editor for the Computer Networks (Elsevier) Journal, Transactions on Mobile Computing and Applications (ICST) and for the Journal of Sensors (Hindawi). He serves in the technical program committees of several leading conferences in wireless communications and networking, including IEEE Infocom, ACM Mobicom, and ACM Mobihoc. He was the technical co-chair of the Ad Hoc and Sensor Networks Symposium for IEEE ICC 2009. His current research interests are in modeling and optimization of multi-hop wireless networks, cross-layer design and optimization, wireless multimedia sensor and actor networks, underwater acoustic networks, and cognitive radio networks.

Scott Pudlewski is a PhD student in the Department of Electrical Engineering at the University at Buffalo, The State University of New York (SUNY), in the the Wireless Networks and Embedded Systems Laboratory. He received his BS in electrical engineering at the Rochester Institute of Technology in 2008. He is the recipient of the SAP America Scholarship in 2008. His current research interests are in compressive sensing for wireless multimedia sensor networks and video distortion based networking.