On the Performance of Compressive Video Streaming for Wireless Multimedia Sensor Networks

Scott Pudlewski and Tommaso Melodia
Wireless Networks and Embedded Systems Laboratory
Department of Electrical Engineering
State University of New York (SUNY) at Buffalo
e-mail: {spudlewski, tmelodia}@eng.buffalo.edu

***Abstract*—This paper investigates the potential of the compressed sensing (CS) paradigm for video streaming in Wireless Multimedia Sensor Networks. The objective is to study performance limits and outline key design principles that will be the basis for cross-layer protocol stacks for efficient transport of compressive video streams. Hence, this paper investigates the effect of key video parameters (i.e., quantization, CS samples per frame, and channel encoding rate) on the received video quality of CS images transmitted through a wireless channels. It is shown that, unlike JPEG-encoded images, CS-encoded images exhibit an *inherent resiliency* to channel errors, caused by the unstructured image representation; this leads to basically *zero* loss in image quality for random channel bit error rates as high as 10^{-4}, and low degradation up to 10^{-3}. Furthermore, it is shown how, unlike traditional wireless imaging systems, forward error correction is not beneficial for wireless transmission of CS images. Instead, an adaptive parity scheme that drops samples in error is proposed and shown to improve image quality. Finally, a low-complexity, adaptive video encoder, is proposed that performs low-complexity motion estimation on sensors, thus greatly reducing the amount of data to be transmitted.**

I. 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 multihop wireless networks of embedded devices is still open and largely unexplored. There are two key shortcomings in systems based on sending predictively encoded video (e.g., MPEG-4 Part 2, H.264/AVC, H.264/SVC) through a layered wireless communication protocol stack, i.e., *encoder complexity* and *low resiliency to channel errors*.

- **Encoder Complexity.** Predictive encoding requires complex processing algorithms, which lead to high energy consumption [1]. Instead, new video encoding paradigms are needed to reverse the traditional balance of complex encoder and simple decoder, which is unsuited for embedded video sensors. Recently developed *distributed video coding* [2] algorithms (aka Wyner- Ziv coding) exploit the source statistics at the decoder, thus shifting the complexity at this end. While promising for WMSNs [1], most practical Wyner-Ziv codecs require end-to-end feedback from the decoder [3], which introduces additional overhead and delay. Furthermore, gains demonstrated by practical distributed video codecs are limited to 2-5 dBs PSNR. Distributed video encoders that do not require end-to-end feedback have been recently proposed [4], but at the expense of a further reduction in performance.
- **Limited Resiliency to Channel Errors**. In existing layered protocol stacks based on the IEEE 802.11 and 802.15.4 standards, 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[1]. This packet loss can lead to the video decoder being unable to decode an independently coded (I) frame, thus leading to 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 perceived video quality should gracefully and proportionally degrade with decreasing channel quality.

We argue, and show through preliminary analysis and experiments, that new cross-layer optimized communication protocol stacks based on the recently proposed compressed sensing (CS) paradigm [5], [6], [7], [8] can offer a convincing solution to the aforementioned problems. However, as will become clearer in the following, this may require a rethinking of traditional wireless streaming functionalities across multiple layers. Compressed sensing (aka "compressive sampling") is a new paradigm that allows the faithful recovery of signals from far fewer measurements than traditional methods based on Nyquist sampling. Hence, CS can offer an alternative to traditional video encoders by enabling imaging systems that sense and compress data simultaneously and much faster, *at very low computational complexity for the encoder*. Image coding and decoding based on CS has been recently explored [9], [10]. So-called single-pixel cameras that can operate efficiently across a much broader spectral range (including

[1]No forward error correction (FEC) is used in either IEEE 802.11 or 802.15.4, and hence a faulty bit corrupts the whole packet.

infrared) than conventional silicon-based cameras have also been proposed [11]. However, transmission of CS images and video streaming in wireless networks, and their statistical traffic characterization, are substantially unexplored.

In this paper, we study the potential of compressive video streaming for Wireless Multimedia Sensor Networks by conducting a cross-layer performance evaluation of wireless streaming of CS video on resource constrained devices. Our objective is to study performance limits and outline key design principles that will be the basis for cross-layer protocol stacks designed for efficient transport of compressive video streams over multi-hop wireless networks. Our contributions can be outlined as follows:

- We study the effect of key video parameters (i.e., quantization, CS samples per frame, and channel encoding rate) on the received video quality of CS images transmitted through a wireless channels;
- We show how, unlike JPEG-encoded images, CS-encoded images exhibit an *inherent resiliency* to channel errors, caused by the unstructured image representation; this leads to basically *zero* loss in image quality for random channel bit error rates as high as 10^{-4}, and low degradation up to 10^{-3}. We discuss the profound impact of this finding on wireless protocol design;
- We show how, unlike traditional wireless imaging systems, forward error correction is not beneficial in CS images. Instead, we propose an adaptive parity scheme that drops samples in error, thus improving the quality of the image reconstruction process;
- We propose a low-complexity, adaptive video encoder, optimized for security videos, that performs low-complexity motion estimation on sensors, thus greatly reducing the amount of data to be transmitted.

The remainder of this paper is structured as follows. In Section II, we discuss our system model. In Section III, we discuss wireless transmission of intra-frame encoded CS video. In Section IV we propose an adaptive parity based channel encoding scheme, while in Section V we propose a CS-based inter-frame encoder optimized for security videos. In Section VI we draw the main conclusions and discuss future work.

II. System Model

A. Compressed Sensing Preliminaries

We consider an image signal represented through a vector $\mathbf{x} \in R^N$, where N is the vector length. We assume that there exists an invertible $N \times N$ transform matrix $\mathbf{\Psi}$ such that

$$\mathbf{x} = \mathbf{\Psi s} \tag{1}$$

where $\mathbf{s}$ is a K-sparse vector, i.e., $||s||_0 = K$ with $K < N$, and where $|| \cdot ||_p$ represents p-norm. This means that the image has a sparse representation in some transformed domain, e.g., wavelet. The signal is measured by taking $M < N$ measurements from linear combinations of the element vectors through a linear measurement operator $\mathbf{\Phi}$. Hence,

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi s} = \tilde{\mathbf{\Psi}}\mathbf{s}. \tag{2}$$

We would like to recover $\mathbf{x}$ from measurements in $\mathbf{y}$. However, since $M < N$ the system is underdetermined. Hence, given a solution $\mathbf{s^0}$ to (2), any vector $\mathbf{s}^*$ such that $\mathbf{s}^* = \mathbf{s^0} + \mathbf{n}$, and $\mathbf{n} \in \mathcal{N}(\tilde{\mathbf{\Psi}})$ (where $\mathcal{N}(\tilde{\mathbf{\Psi}})$ represents the null space of $\tilde{\mathbf{\Psi}}$), is also a solution to (3). However, it was proven in [6] that if the measurement matrix $\mathbf{\Phi}$ is sufficiently incoherent with respect to the sparsifying matrix $\mathbf{\Psi}$, and K is smaller than a given threshold (i.e., the sparse representation $\mathbf{s}$ of the original signal $\mathbf{x}$ is "sparse enough"), then the original $\mathbf{s}$ can be recovered by finding the sparsest solution that satisfies (2), i.e., the sparsest solution that "matches" the measurements in $\mathbf{y}$. However, the problem above is in general NP-hard. For matrices $\tilde{\mathbf{\Psi}}$ with sufficiently incoherent columns, whenever this problem has a sufficiently sparse solution, the solution is unique, and it is equal to the solution of the following problem:

$$P_1 : \ minimize \, ||s||_1$$
$$subject \, to : ||\mathbf{y} - \tilde{\mathbf{\Psi}}\mathbf{s}||_2^2 < \epsilon, \tag{3}$$

where ϵ is a small tolerance. Note that problem P_1 is a convex optimization problem. The reconstruction complexity equals $O(M^2 N^{3/2})$ if the problem is solved using interior point methods [12].

B. Video Model

We represent each frame of the 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. A conventional imaging system or a single-pixel camera [11] can be the base of the imaging scheme. In the latter case, the video source only obtains random samples of the image (i.e., linear combinations of the pixel intensities). In our model, the image can be sampled using a scrambled block Hadamard ensemble [13]

$$\mathbf{y} = \mathbf{H_{32}} \cdot \mathbf{x}, \tag{4}$$

where $\mathbf{y}$ represents image samples (measurements), $\mathbf{H_{32}}$ is the 32×32 Hadamard matrix and $\mathbf{x}$ the matrix of the image pixels. The matrix $\mathbf{x}$ has been randomly reordered and shaped into a $32 \times \frac{N}{32}$ matrix where N is the number of pixels in the image. Then M samples are randomly chosen from $\mathbf{x}$ and transmitted to the receiver. The receiver then uses the M samples transmitted along with the randomization patterns for both randomizing the pixels into $\mathbf{x}$ and choosing the samples out of x to be transmitted (both of which can be decided upon before network setup) and recreates the image solving P_1 in (3) through a suitable algorithm, e.g., GPSR[2] [14], StOMP [15].

III. Transmission of Intra-Frame Encoded Video

In this section, we study the effect of key design parameters on the received video quality of CS images transmitted through a wireless channels; We first consider intra-coded frames,

[2]GPSR is used for image reconstruction in the simulation results presented in this paper.

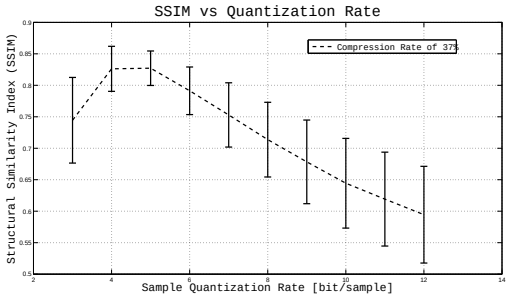

Fig. 1. Structural Similarity (SSIM) Index [16] for Images with a Constant Bit Rate of 37% of the Original Image Size for Varying Quantization Levels

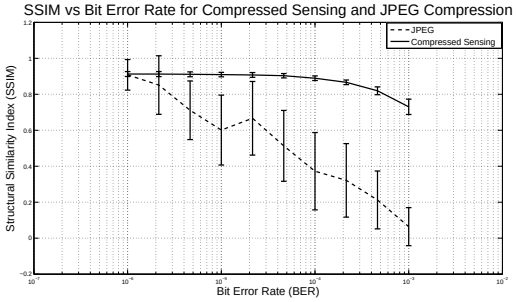

Fig. 2. Structural Similarity (SSIM) vs Bit Error Rate (BER) for compressed sensed images, and images compressed using JPEG

i.e., we temporarily ignore the temporal correlation among different frames. For a given data rate at the transport layer F $[bit/s]$, number of frames per second, and end-to-end bit error rate (BER), there are three main parameters that determine the perceptual quality of the received video frame, i.e., the quantization level of each sample Q, the number of samples per frame M, and the channel encoding rate R.

1) Sample Quantization Rate: The *sample quantization rate* Q $[bit/sample]$ is the number of bits used to quantize each sample. The smaller Q, the lower the amount of information sent per sample, and therefore the greater the number of samples that can be transmitted for a target data rate F, at the expense of greater quantization distortion in each sample. We empirically evaluated the video quality of CS images against the optimal ratio of number of samples M vs quantization rate Q. To do so, we evaluated the *Structural Similarity Index* (SSIM)[3] [16] between the original and the encoded image for a standardized set of 25 images. We kept the total image size constant at 37% of the original image size, i.e., the image size that allows sending N samples (where N corresponds to the number of image pixels) with 3-bit quantization.

Figure 1 shows the average SSIM of the above-mentioned images against sample quantization rate, with 95% confidence intervals. Clearly, *the benefit of more samples outweighs the distortion caused be less accurate samples* down to $5\,\mathrm{bit/sample}$. Intuitively, this is because the recovery algorithm finds *image with the sparsest transform* that minimizes the difference between the samples received and the samples generated from the reconstructed image. This means that even though a small amount of samples (less than one in 10^3) may be corrupted, the reconstructed image is the same or very similar to the image which would have been reconstructed without bit errors.

[3]The SSIM index is preferred to the more widespread PSNR, which has been recently shown to be inconsistent with human eye perception [16]. SSIM is a more accurate measurement of error because the human visual system perceives structural errors in the image more than others. For example, changes in contrast or luminance, although mathematically significant, are very difficult to discern for the human eye. Structural differences such as blurring, however, are very noticeable. SSIM is able to weight these structural differences better to create a measurement closer to what is visually noticeable than traditional measures of image similarity such as mean squared error (MSE) or PSNR.

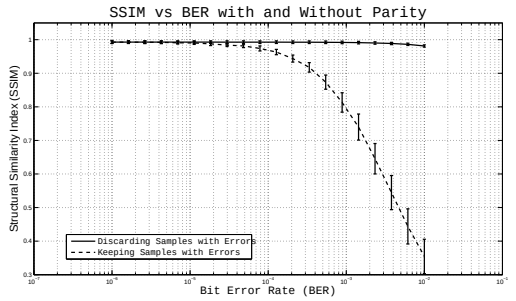

Fig. 3. Compressed Sensed Images Reconstructed With and Without Incorrect Samples

2) Samples per Frame: The number of samples N needed to reconstruct the image to a predefined quality level is dependent on the sparsity of the transmitted image. The greater the number of transmitted samples compared to the sparsity of the image, the better the image quality of the received frame. Depending on the desired video quality at the receiver, the maximum number of samples per frame can be selected to achieve that quality. We will discuss this further in Section V.

3) Effect of Channel Errors: 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. Also, a small amount of random channel errors does not affect the perceptual quality of the received image *at all*, since, for moderate bit error rates, the greater sparsity of the "correct" image will offset the error caused by the incorrect bit. This is demonstrated in Fig. 3. 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, which is 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 has important consequences and *provides a strong motivation for studying* compressive wireless video streaming in WMSNs.

This inherent resiliency of compressed sensing to random

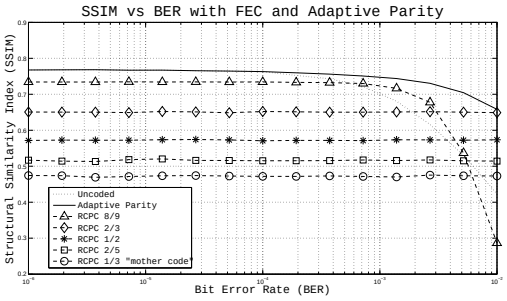

Fig. 4. Adaptive Parity vs RCPC Encoding for Variable Bit Error rates

channel bit errors is even more noticeable when compared to traditional image compression schemes. Figure 2 shows the average SSIM of 25 images transmitted through a binary symmetric channel with varying crossover probability. 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 Figs. 5 and 6. Figure 5 shows the received image *Lena* encoded with CS and transmitted with bit error rates of 10^{-5}, 10^{-4}, and 10^{-3}. Figure 6 shows the same image, but encoded with JPEG. The difference is stunning - the effect of channel errors is disruptive for structured data like JPEG-encoded images. The reader will easily realize that the effect of channel errors on predictively-encoded video is even more disruptive, 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.

To determine the *channel encoding rate*, we first must determine the channel coding strategy appropriate for compressed-sensed imaging data transmitted over a multi-hop wireless network. One of the biggest advantages of compressed sensing is that the transmitted samples constitute a random, incoherent combination of the original data. This means that no single sample is any more important than any other sample. Instead, only *the number* of correctly received samples is the main factor in determining the quality of the received image. Also, following the same logic as for the quantization parameter selection, a small amount of errors will not considerably affect the perceptual quality of the received image, since, for a moderate error rate, the greater sparsity of the correct image will offset the error caused by the incorrect bit. This is demonstrated in Figs 3 and 4. In Fig. 3, the same set of images were reconstructed both with and without corrupted samples after being transmitted through a binary symmetric channel. Clearly, the image quality considerably improves when the corrupted samples are dropped.

IV. Adaptive Parity-based Channel Coding

As discussed in the previous section, for a fixed number of bits per frame, the perceptual quality of video streams can be further improved by dropping errored samples that would contribute to image reconstruction with incorrect information. This can be obtained by 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. To determine the amount of samples to be jointly encoded , the amount of correctly received packets is modeled as

$$C = \left(\frac{Q \cdot b}{Q \cdot b + 1}\right)(1 - BER)^{Q \cdot b + 1}, \tag{5}$$

Where C is the estimated amount of correctly received samples, b is the number of jointly encoded samples, and Q is the quantization rate per sample. To determine the optimal value of b for a given BER, (5) can be differentiated, set equal to zero and solved for b. If the end-to-end BER can be estimated by the transmitting node, the optimal channel encoding rate can then be chosen and used to encode the samples. The received video quality using the parity scheme described was compared to different levels of channel protection using rate compatible punctured codes (RCPC). Specifically, we use the $\frac{1}{4}$ mother codes discussed in [17]. Briefly, a $\frac{1}{4}$ convolutional code is punctured to decrease the amount of redundancy needed for the encoding process. These codes are punctured progressively so that every *higher rate* code is a subset of the lower rate codes. For example, any bits that are punctured in the $\frac{4}{15}$ code must also be punctured in the $\frac{1}{3}$ code, the $\frac{4}{9}$ code, and so on down to the highest rate code, in this case the $\frac{8}{9}$ code. Because of this setup, the receiver can decode the entire family of codes with the same decoder. This allows the transmitter to choose the most suitable code for the given data. Clearly, as these codes are punctured to reduce the redundancy, the effectiveness of the codes decreases as far as the ability to correct bit errors. Therefore we are trading bit error rate for transmission rate.

Figure 4 shows the adaptive parity scheme compared to RCPC codes. Clearly, for all reasonable bit error rates, the adaptive parity scheme outperforms all levels of RCPC codes. The parity scheme performs better for all levels of BER, and it is also much simpler to implement than more powerful forward error correction (FEC) schemes. The parity scheme performs better because, even though the FEC schemes show stronger error correction capabilities, the additional overhead does not make up for the video quality increase compared to just dropping the samples which have errors.

V. Inter-frame Encoded Compressed Video Streaming

In this section, we present a method for inter-frame encoding. While the proposed method is general, it works particularly well for security videos. Security videos are a special case of video in which we can assume that the camera is not moving, but only the objects within the field of view (FOV) of the camera are moving. Because of this, there will often be a large amount of redundancy from one frame of the video to the next. One way to exploit this redundancy within the framework of compressed sensing is by taking the algebraic difference between two frames, encoding this

(a) (b) (c)

Fig. 5. Lena encoded using compressed sensing with (a) 10^{-5} BER (b) 10^{-4} BER (c) 10^{-3} BER.

(a) (b) (c)

Fig. 6. Lena encoded with JPEG with (a) 10^{-5} BER (b) 10^{-4} BER (c) c10^{-3} BER.

difference, recreating an image representing this difference and combining it with the reference frame at the receiver. If the image being encoded and the reference image are very similar (i.e. have a very high correlation coefficient), then this difference image will be sparser than either of the original images, and can therefore be transmitted at the same quality using fewer samples than the original image. Note that while the reference image is sparse only in the transformed domain (e.g., wavelet), the difference image is sparse in the original domain. Even though there are proposed algorithms that obtain higher received video quality [18], [19], [20], these methods all involve having access to at least some of the frames at the transmitter. Because the proposed method works directly on the samples and does not require knowledge of this original image, it is suitable for use with a single pixel camera where the original image is not available.

The procedure for encoding the video is based on the amount of correlation between frames, as measured by the *correlation coefficient*. In order to minimize the propagation of errors between frames, a *reference frame* is used to both compare the current frame, and to base a difference frame from for a slow moving video sequence. The correlation coefficient α is calculated between the reference frame and the current frame being evaluated. The sparsity β of the recreated difference frame can be calculated by a linear function of α defined by (6)

$$\beta = -A \cdot \alpha + B \tag{6}$$

where A and B can be estimated by regression techniques. Since the amount of samples required to correctly decode the image is based on the sparsity of the received image, the type of compression used can be directly based on α. Two threshold levels are set, θ_{high} and $\theta_{moderate}$, which determine what type of encoding should be used.

- $\alpha < \theta_{moderate}$. When the correlation between a frame and the reference frame is low, the frame being considered is compressed using the standard intra-frame process described in the previous section. Also, the frame being considered is marked as the current *reference frame*.
- $\theta_{moderate} < \alpha < \theta_{high}$. When there is moderate correlation between the frame being considered and the reference frame, a *difference frame* is calculated between the frame being considered and the most recent reference frame. The difference frame is generated by finding the difference between the samples of the current frame and the samples of the reference frame. Then, the first L samples are transmitted, where L is calculated by a linear function of the correlation coefficient.
  These samples are then decoded at the receiver, thus recreating the actual difference frame. That difference frame can be added to the reference frame stored at the receiver in order to reconstruct the frame at the source. This method allows us to exploit the correlation between similar frames while keeping the complexity at the source low.
  The three main advantages of this method are:
  - The reference frame is *inherently sparse*, i.e., it is sparse in the domain of random samples, not in the transformed domain. Hence, it can be decoded without using an additional sparsifying transform. This can greatly speed up the performance of the decoder;
  - The reference frame can be quantized using less bits than a reference frame. This is because there is less

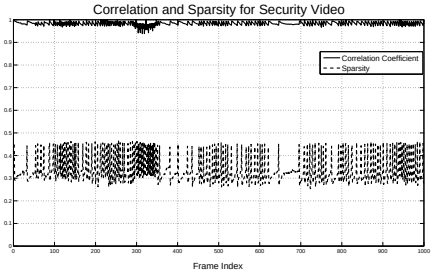

Fig. 7. Correlation and Sparsity for a Security Video

variance within the samples of the reference frame than in a typical frame;
  - Because the difference frame can be sparser than the transform of a normal frame, less samples need to be transmitted and decoded for the same quality;
- $\alpha > \theta_{high}$. If the correlation coefficient is very high, this is an indication that there is no action is the frame. In this case, there is no need to transmit anything and the decoder can simple copy the reference frame directly (i.e. the difference frame is zero).

This is demonstrated in Fig. 7. In this figure, we show the amount of correlation between the two frames being considered, the amount of sparsity as measured at the decoder and the quality (SSIM) of the decoded frame as compared to the original uncompressed frame. This simulation was created using a surveillance camera in a mall. The "spikes" in the sparsity value indicate place where the intra-frame compression was used. In the sections without those spikes, clearly there is a negative linear relationship between the two. Since the correlation can be measured using only information at the video source, this allows the source node to estimate the amount of sparsity at the receiver.

The mean value for the SSIM for this particular scenario was 0.780, with a compression ratio of 4.54. Though this is clearly not as much compression as can be obtained from MPEG-2 or other modern codecs, this protocol is only exploiting the *spacial correlation* between between frames, without considering the possibility of using motion vectors or huffman coding as is used in most standard encoders. Further, the algorithm is very simple to implement at the video source, involving much less source complexity than most standard encoders.

VI. Conclusions

We have investigated the potential of the compressed sensing (CS) paradigm for video streaming in WMSNs. We have shown that, unlike JPEG-encoded images, CS-encoded images exhibit an *inherent resiliency* to channel errors, caused by the unstructured image representation; this leads to basically *zero* loss in image quality for random channel bit error rates as high as 10^{-4}, and low degradation up to 10^{-3}. Furthermore, we have shown that, unlike traditional wireless imaging systems, forward error correction is not beneficial for wireless transmission of CS images. Instead, we proposed an adaptive parity scheme that drops samples in error thus improving the quality of the reconstructed image. Finally, we have proposed a low-complexity adaptive video encoder that performs motion estimation on the video sensors, thus considerably reducing the amount of data to be transmitted. Our future work will be focused on designing cross-layer optimized communication protocols for CS-based WMSN based on the principles outlined in this preliminary investigation.

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] T. Sheng, G. Hua, H. Guo, J. Zhou, and C. W. Chen, "Rate allocation for transform domain Wyner-Ziv video coding without feedback," in *ACM Intl. Conf. on Multimedia*, October 2008.

[5] D. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] 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.

[7] ——, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[8] E. Candes and T. Tao, "Near-optimal Signal Recovery from Random Projections and Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[9] 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.

[10] J. Romberg, "Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.

[11] 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.

[12] I. E. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 1994.

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

[14] 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.

[15] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Preprint*, 2007.

[16] 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.

[17] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *Communications, IEEE Transactions on*, vol. 36, no. 4, pp. 389–400, Apr 1988.

[18] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive Video Sampling," in *In Proc. of the European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, August 2008.

[19] J. Park and M. Wakin, "A Multiscale Framework for Compressive Sensing of Video," in *In Proc. of the Picture Coding Symposium (PCS)*, Chicago, Illinois, May 2009.

[20] R. Marcia and R. Willett, "Compressive coded aperture video reconstruction," in *In Proc. of the European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, August 2008.