16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

COMPRESSIVE VIDEO SAMPLING

Vladimir Stanković and Lina Stanković

Dept of Electronic and Electrical Engineering
University of Strathclyde, Glasgow, UK
phone: +44-141-548-2679
email: {vladimir,lina}.stankovic@eee.strath.ac.uk

Samuel Cheng

Dept of Electrical and Computer Engineering
University of Oklahoma, Tulsa, OK, USA
phone: 1-918-660-3234
email: samuel.cheng@ou.edu

ABSTRACT

Compressive sampling is a novel framework that exploits sparsity of a signal in a transform domain to perform sampling below the Nyquist rate. In this paper, we apply compressive sampling to significantly reduce the sampling rate of video. A practical system is developed that first splits each video frame into non-overlapping blocks of equal size. Compressive sampling is then performed on sparse blocks, determined by predicting sparsity based on previous reference frames which are sampled conventionally. The blocks identified as sparse are reconstructed using the orthogonal matching pursuit algorithm, whereas the remaining blocks are sampled fully. Thus, the acquisition complexity and sampling time are reduced, while exploiting the local sparsity, within the DCT domain, of a video stream. Our simulation results indicate up to 50% saving in acquisition for Y-components of video with very small performance loss compared to traditional sampling.

1. INTRODUCTION

Compressive sampling is a statistical technique for data acquisition and estimation that aims to sample signals sparsely in transform domains. The process of compressive sampling replaces conventional sampling and reconstruction with a more general linear measurement scheme with an optimization procedure to acquire a subset of signals within a source at a rate that is significantly below Nyquist. However, it will work only if the source is sparse in the transform domain of choice. Thus, the challenge is to predict which sources are sparse in a particular transform domain.

Conventionally, after acquisition of a scene, Discrete Cosine Transform (DCT) is performed on the image using values assigned to each pixel. After DCT, many coefficients will be zero or will carry negligible energy; these coefficients are discarded before quantization or/and entropy coding. Hence, though each frame of the image is acquired fully, much of the acquired information is discarded after DCT causing unnecessary burden on the acquisition process. This makes compressive sampling a good candidate for digital image and video applications, where the Nyquist rate is so high that compressing the sheer volume of samples is a problem for transmission or storage.

A number of theoretical contributions have appeared on compressive sampling (see [1, 2, 3]) over the past few years. Yet, only few papers address day-to-day practical situations, for example, analog-to-information converter [4] and one-pixel camera [5]. Recently, we developed practical compressive sampling systems for binary image in [6], by using the fact that natural images are sparse in a transform domain. We observed that the background was a good candidate for compressive sampling, while parts of the image that contain many details were not. Hence, applying compressive sampling to the whole image was ineffective. In order to exploit any sparsity within an image, we split the image into small non-overlapping blocks of equal size. Compressive sampling is then performed only on blocks determined to be sparse, i.e., we exploit *local sparsity* within an image. Note that, in real-time acquisition, it is not possible to test for sparsity of a block before sampling.

In this paper, we build on our previous work [6] considering acquisition of a video stream. We solve the real-time acquisition limitation by testing the sparsity of a scene on a previously acquired frame, denoted as reference frame. Reference frames are sampled fully, and they are used to predict sparsity of the successive non-reference frames. Each block of the reference frame is tested for sparsity using a compressive sampling test. If the block passes the test, it is identified as sparse. All blocks in the successive non-reference frames that spatially correspond to sparse blocks in the previous reference frame will be compressively sampled. Compressively sampled frames are reconstructed at the decoder using the orthogonal matching pursuit (OMP) algorithm [7], which is suboptimal but practical due to its relatively lower complexity compared to other proposed reconstruction methods [1, 2, 8, 9].

However, the OMP algorithm is still not practical when carried out over a very large number of samples, e.g., a whole frame. Having split the frames into smaller blocks, the reconstruction time is automatically reduced, such that our overall system operates in real time. Thus, in addition to exploiting local sparsity, the proposed system benefits from reduced complexity since the OMP algorithm reconstructs a smaller number of coefficients many times (instead of many coefficients in one go), and hence converges faster.

Our simulation results show that is possible to significantly reduce the total number of samples using compressive sampling without sacrificing the reconstruction performance. We also observe that only few frames need to be encoded as reference frames especially for video with almost static background. This paper is organized as follows. Section 2 provides an overview of compressive sampling, Section 3 describes our proposed system, simulation results are presented in Section 4, and we conclude and outline our future work in Section 5.

2. COMPRESSIVE SAMPLING

Compressive sampling or compressed sensing [1, 2] is a novel framework that enables sampling below the Nyquist rate, without (or with small) sacrifice in reconstruction quality. It is based on exploiting sparsity of the signal in some domain. In this section we briefly review compressive sampling closely following notation of [3].

Let $\mathbf{x} = \{x[1], \ldots, x[N]\}$ be a set of N samples of a real-valued, discrete-time random process X. Let $\mathbf{s}$ be the representation of $\mathbf{x}$ in the $\mathbf{\Psi}$ domain, that is:

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s} = \sum_{i=1}^{N} s_i \psi_i, \tag{1}$$

where $\mathbf{s} = [s_1, \ldots, s_N]$ is an N-vector of weighted coefficients $s_i = \langle \mathbf{x}, \psi_i \rangle$, and $\mathbf{\Psi} = [\psi_1 | \psi_2 | \cdots | \psi_N]$ is an $N \times N$ basic matrix with ψ_i being the i-th basic column vector.

Vector $\mathbf{x}$ is considered K-sparse in the domain $\mathbf{\Psi}$, for $K \ll N$, if only K out of N coefficients of $\mathbf{s}$ are non-zero. Sparsity of a signal is used for compression in conventional transform coding, where the whole signal is first acquired (all N samples of $\mathbf{x}$), then the N transform coefficients $\mathbf{s}$ are obtained via $\mathbf{s} = \mathbf{\Psi}^T \mathbf{x}$, and then $N - K$ (or more in the case of lossy compression) coefficients of $\mathbf{s}$ are discarded and the remaining are encoded. Hence severe redundancy is present in the acquisition since large amounts of data are discarded because they carry negligible or no energy.

The main idea of compressive sampling is to remove this "sampling redundancy" by needing only M samples of the signal, where $K < M \ll N$. Let $\mathbf{y}$ be an M-length measurement vector given by: $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, where $\mathbf{\Phi}$ is an $M \times N$ measurement matrix. The above expression can be written in terms of $\mathbf{s}$ as:

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

It has been shown in [1, 2] that signal $\mathbf{x}$ can be recovered losslessly from $M \approx K$ or slightly more measurements (vector $\mathbf{y}$ in (2)) if the measurement matrix $\mathbf{\Phi}$ is properly designed, so that $\mathbf{\Phi}\mathbf{\Psi}$ satisfies the so-called restricted isometry property [2]. This will always be true if $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are incoherent, that is, the vectors of $\mathbf{\Phi}$ cannot sparsely represent basic vectors and vice versa.

It was further shown in [1, 2, 3] that an independent identically distributed (i.i.d.) Gaussian matrix $\mathbf{\Phi}$ satisfies the above property for any (orthonormal) $\mathbf{\Psi}$ with high probability if $M \geq cK \log(N/K)$ for some small constant c. Thus, one can recover N measurements of $\mathbf{x}$ with high probability from only $M \approx cK \log(N/K) < N$ random Gaussian measurements $\mathbf{y}$ under the assumption that $\mathbf{x}$ is K-sparse in some domain $\mathbf{\Psi}$. Note that it is not known in advance which s_i coefficients are zeros, or which $x[i]$ samples are not needed.

Unfortunately, reconstruction of $\mathbf{x} = \{x[1], \ldots, x[N]\}$ (or equivalently, $\mathbf{s} = [s_1, \ldots, s_N]$) from vector $\mathbf{y}$ of M samples is not trivial. The exact solution [1, 2, 3] is NP-hard and consists of finding the minimum l_0 norm (the number of non-zero elements). However, excellent approximation can be obtained via the l_1 norm minimization given by:

$$\hat{\mathbf{s}} = \arg\min ||s'||_1, \quad \text{such that} \quad \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}' = \mathbf{y}. \tag{3}$$

This convex optimization problem, namely, basis pursuit [1, 2], can be solved using a linear program algorithm of $O(N^3)$ complexity. In contrast to l_0 norm minimization, the l_1 norm minimization usually requires more than $K+1$ measurements. Due to complexity and low speed of linear programming algorithms, faster solutions were proposed at the expense of slightly more measurements, such as matching pursuit, tree matching pursuit [8], orthogonal matching pursuit [7], and group testing [9].

3. PROPOSED SYSTEM

In this section, we describe our system for compressive video sampling using the OMP algorithm [7].

We use an i.i.d. Gaussian measurement matrix for $\mathbf{\Phi}$ and inverse DCT for $\mathbf{\Psi}$ in equation (2). This choice of $\mathbf{\Phi}$ ensures that the restricted isometry property is satisfied. The OMP algorithm is an efficient solution for signal recovery that is easy to implement. It is of $O(MNK)$ complexity, and requires $M \approx 2K \log N$ measurements for error-free recovery of N samples in 99% of time. The algorithm has K iterations, and in each iteration it calculates N inner products between M-length vectors and finds the maximum. Thus, when M and N are large the algorithm is slow and impractical.

Our proposed system is shown in Figure 1. To reduce the acquisition and reconstruction complexity and exploit local sparsity within the frame, each frame is split into B non-overlapping blocks each of size $n \times n = N$ pixels. We define reference and non-reference frames. Each reference frame is sampled fully. After sampling, a compressive sampling test (described below) is carried out to identify which blocks are sparse within the reference frame. The output of the test is binary for each of the B blocks, e.g., true or false.

Reference frames should be inserted regularly in the stream: exactly when reference frames are required can be determined by exploiting decoder feedback. The number of required reference frames depends on the dynamics of the scene, as shown in Section 4.

Let B_s be the number of sparse blocks. Each non-reference frame is sampled in the following way: B_s blocks that spatially correspond to the sparse blocks in the reference frame are compressively sampled; that is, each block is transformed into a $N \times 1$ vector on which Gaussian measurement matrix Φ is applied. The remaining $B - B_s$ blocks are sampled in the conventional way. For each of the B_s selected blocks we acquire $M < N$ measurements. The resulting coefficients can undergo conventional compression in the form of quantization/entropy coding.

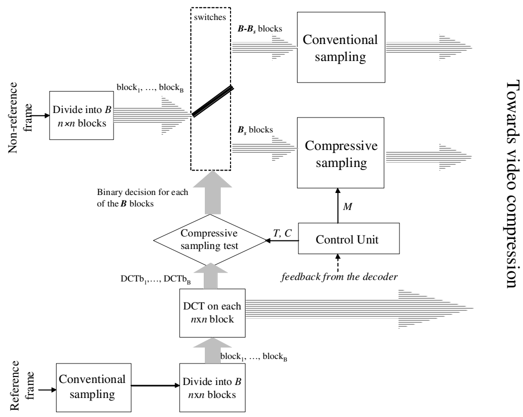

Figure 1: Block diagram of the acquisition process.

Compressive sampling test: Reference frames are sampled fully, and DCT is applied on each of the B blocks. We select B_s sparse blocks in the following manner. Let C be a small positive constant, and T an integer threshold that is representative of the average number of non-significant DCT coefficients over all blocks. If the number of DCT coefficients in the block whose absolute value is less than C is larger than T, the block is selected as a reference for compressive sampling.

During reconstruction of compressively sampled blocks, using the iterative OMP algorithm, all sampled coefficients whose absolute value is less than C are set to zero. Hence, for $C > 0$, the sampling process will always be lossy. Theoretically, as discussed in Section 2, if we have $N-K$ non-significant DCT coefficients, then at least $M = K+1$ samples are needed for signal reconstruction. Therefore, $T < N-K$.

The choice of values for M, T, and C depends on the video sequence and the size of the blocks. These parameters, as well as the number and position of the reference frames in the stream, can be adjusted using feedback from the decoder. Indeed, during reconstruction, the OMP algorithm can fail (if M or T is too low) even if the block has passed the compressive sampling test. Often, it is possible for the decoder to detect the OMP algorithm failure, for example, by comparing the result with neighboring blocks/frames, or if all the DCT coefficients are zero, or if division by zero in the OMP algorithm appears. In this case, the decoder can send feedback to the encoder to adjust its sampling parameters or to encode a reference frame.

Note that, any transform can be applied instead of DCT (e.g., wavelet transform) under the condition that the frame is sparse in the transform domain.

4. RESULTS

In this section we report our experimental results, namely the effect of compressive sampling with different combination of parameters T, M, C, on reconstructed video PSNR and perceptual quality. The system described in Section 2 is applied to Y-components of two video sequences: the QCIF Akiyo sequence and CIF Stefan sequence. In all our simulations, a block size of $N = 32 \times 32 = 1024$ pixels is used. This block size was observed to provide a good trade-off between compressive sampling efficiency, reconstruction complexity, and decoding time.

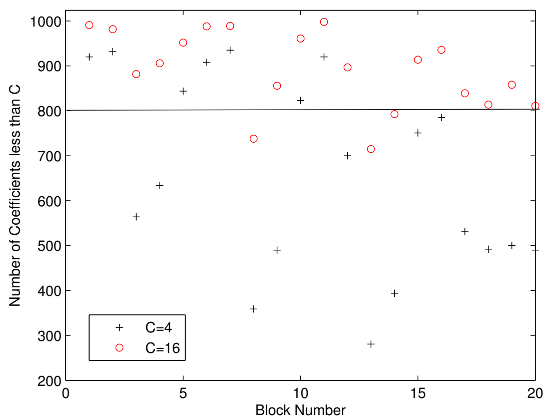

Figure 2: Distributions of the number of coefficients less than C over B=20 blocks, each of 32×32 pixels for the first frame of Akiyo sequence.

Figures 2 and 3 show the number of DCT coefficients less than C for the first frame of the Akiyo and Stefan sequences, respectively, for two different values of C. It can be seen from the figures that if C is large enough, many blocks have more than 80% of its DCT coefficients $< C$, and thus can be regarded as sparse. This determines our choice of T (horizontal line in the figures). The figures show that not all blocks are sparse, and hence the compressive sampling test described in Section 3 is necessary to select blocks on which compressive sampling can be applied to reduce sampling rate.

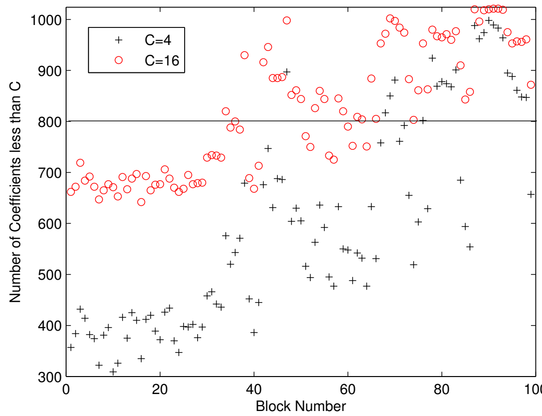

Figure 3: Distributions of the number of coefficients less than C over B=99 blocks, each of 32×32 pixels for the first frame of Stefan sequence.

Figure 4 shows results for the 217^{th} frame of the Akiyo sequence, with the first frame used as a reference. The results are shown as PSNR vs the percentage of the collected samples for four different values of M. We fix T to 800, and show PSNR for C ranging from 2 to 30 with a step size of 2 to obtain different sampling rates. It can be seen from the figure that M=400 is the best choice for a large range of sampling rates. Indeed, for M=400, only about 50% of samples are needed to obtain PSNR of 30 dB. For higher sampling rates (above 85%) M=600 or 800 are the best. This is expected since for a fixed C, larger number of samples M leads to a better quality. Similar results were obtained for different frames. In our next simulation, we fixed M to 420 and vary T.

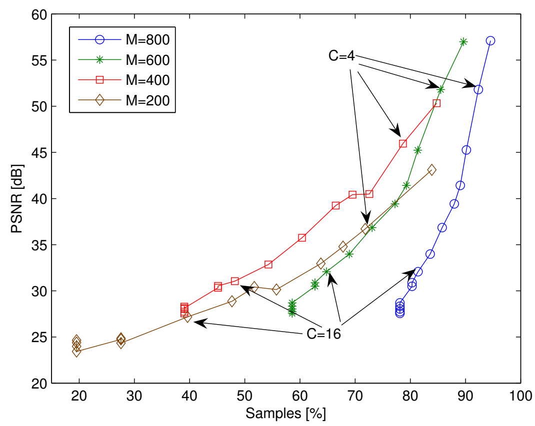

Figure 4: PSNR vs. percentage of the acquired samples for the 217^{th} frame of Akiyo. The first frame was used for sparsity prediction. $T = 800$.

Figure 5 shows PSNR between the original frame and the reconstruction as a function of the percentage of the collected samples for three different values of T. Again, C ranges from 2 to 30 with a step size of 2. It can be seen from the figure that T=800 is the best choice. Good reconstruction quality is achieved with 50% of the samples only.

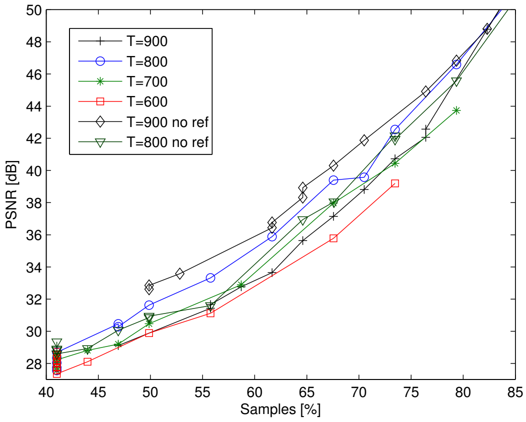

Figure 5: PSNR vs. percentage of the acquired samples for the 217^{th} frame of Akiyo. Either the first frame was used for sparsity prediction or the sampled 217^{th} frame itself ('no ref' curves). $M = 420$.

As a benchmark, we include in the figure results ('no ref' curve) for the ideal case when the compressive sampling test is carried out on the 217^{th} frame itself. It can be seen, that negligible performance loss (about 1 dB) is incurred in our practical case (with $T = 800$) when the first frame is used as reference compared to the best benchmark case ($T = 900$). Hence, for the Akiyo sequence, we can conclude that it is sufficient to fully sample only the first frame and successive frames can be compressively sampled using the compressive sampling test decision on blocks of the first reference frame.

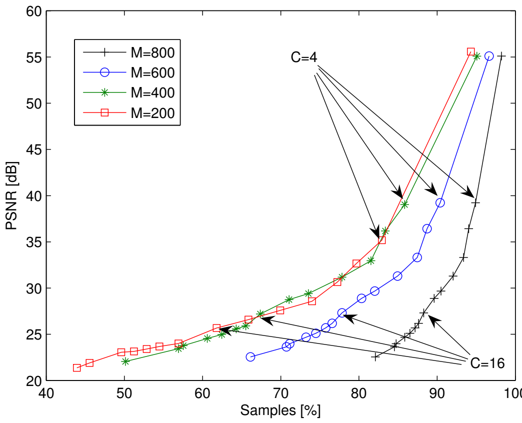

Figure 6: PSNR vs. percentage of the acquired samples for the 100^{th} frame of Stefan. The first frame was used for sparsity prediction. $T = 800$.

Results for the 100^{th} frame of the Stefan sequence are shown in Figures 6 and 7. The first frame was used as a reference. From Figure 6 we can observe that $M = 200, 400$ give the best performance for $T = 800$. In Figure 7, we set $M = 420$ and vary T. We observe that roughly 75% of samples are needed with the proposed system for T=700, whereas in the ideal case ('no ref' curve) with T=800 less than 70% of samples will be required to obtain PSNR greater than 30 dB. Note that when $T = 900$ it is not possible to obtain sampling rate below 70% because only few blocks pass the compressive sampling test.

Figure 8 shows results for the 4^{th} frame and the 297^{th} frame, using the first frame as reference. It can be seen that for the 4^{th} frame, roughly 70% of samples are needed to achieve PSNR of 30dB. On the other hand, if we also maintain the first frame as a reference for sampling the 297^{th} frame, more than 90% of samples would be necessary. In this case, results could be improved by inserting another reference frame before the 297^{th} frame.

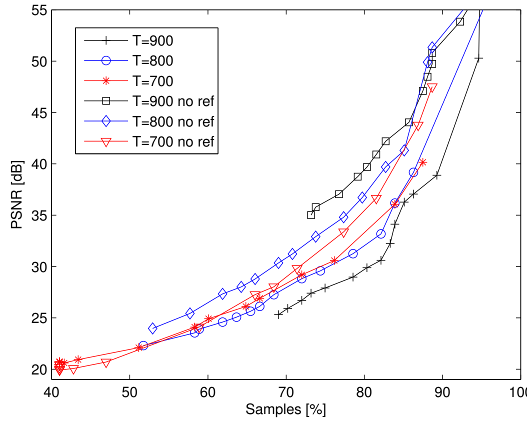

Figure 7: PSNR vs. percentage of the acquired samples for the 100^{th} frame of Stefan. Results denoted as 'no ref' show the ideal case when compressive sampling test is done on the sampled 100^{th} frame itself. $M = 420$.

It is obvious from the above figures that compressive sampling is more efficient with the Akiyo sequence than with the Stefan sequence. This is expected due to increased dynamics of the scene in the Stefan sequence. Thus for the Stefan sequence reference frames will have to be inserted more frequently.

Figure 9 shows the 217^{th} frame of Akiyo compressively sampled with $M = 200$, 400, 600, and 800 with the first frame used as a reference. With $T = 800$ and $C = 8$, there were $B_s = 10$ compressively sampled blocks, i.e., 50% of the reference frame was deemed sparse. It is clear that $M = 200$ samples per block (60% of samples are acquired in the whole frame) is too low in this setting for many applications (reconstructed PSNR is 31.94 dB), whereas $M = 400$ (70%), $M = 600$ (79%), and $M = 800$ (89%), resulting in PSNR of 38.22 dB, 41.41 dB, and 41.42 dB, respectively, would usually be more than enough.

Results for the 100^{th} frame of Stefan are shown in Figure 10 for $T = 700$, $C = 4$, and M =200, 400, 600, and 800. The number of compressively sampled blocks was $B_s = 27$, i.e., 27% of the first frame (reference) was deemed sparse. When $M = 200$ samples were taken for each of the selected blocks, PSNR was 26.06 dB. With $M = 400$ and 600 samples, however, PSNR increased to 36.34 dB and 37.58 dB,

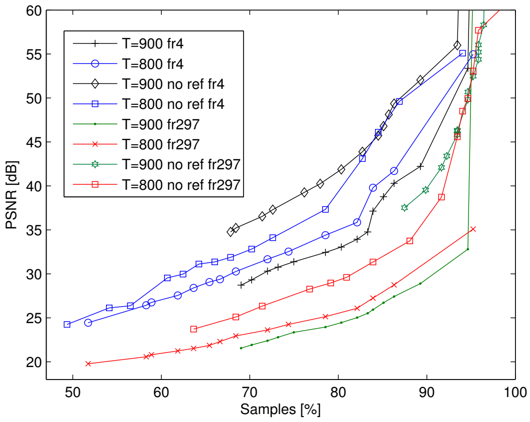

Figure 8: PSNR vs. percentage of the acquired samples for the 4^{th} and 297^{th} frame of Stefan. Results denoted as 'no ref' show the ideal case when compressive sampling test is done on the sampled frame itself (4^{th} or 297^{th}). $M = 420$.

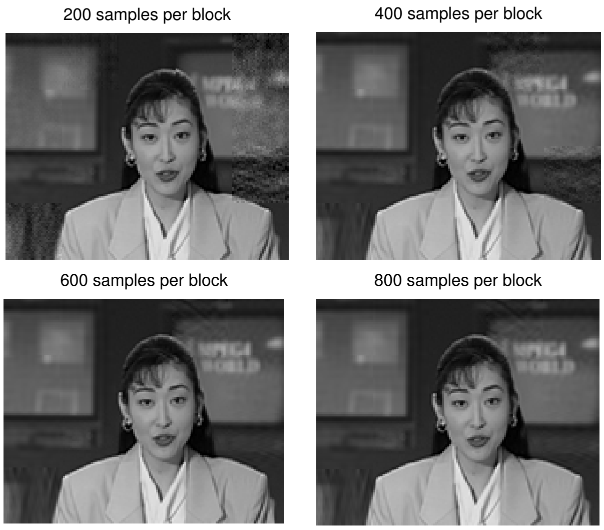

Figure 9: The compressively sampled 217^{th} frame of Akiyo with OMP reconstruction. $T = 800$.

respectively.

In Figures 9 and 10, whenever the decoder detected the error due to the OMP algorithm failure, the whole block was replaced with the spatially corresponding reference block.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a system based on the novel concept of compressive sampling to achieve real-time, low complexity acquisition of video. Each frame of the video is split into a number of smaller non-overlapping blocks of equal size to reduce the complexity of compressive sampling algorithms and exploit the varying sparsity across blocks within a frame. Compressive sampling is performed only on those blocks that satisfy our proposed simple sparsity test, while the remaining blocks are sampled fully. Full sampled reference frames are used to predict sparsity of the blocks within successive frames. Experimental results show great potential for compressive sampling for video acquisition, with up to 50% savings in acquisition with good reconstruction quality.

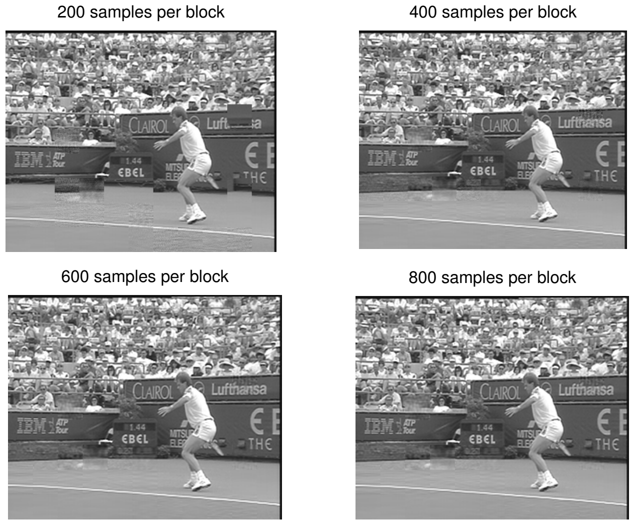

Figure 10: The compressively sampled 100^{th} frame of Stefan with OMP reconstruction. $T = 700$.

The main future challenge is to develop a new compression scheme that should follow compressive acquisition.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory,* vol. 52, pp. 1289-1306, April 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory,* vol. 52, pp. 489-509, Feb. 2006.

[3] R.G. Baraniuk, "Compressive Sensing [Lecture Notes]," *IEEE Signal Processing Magazine,* vol. 24, pp. 118–121, July 2007.

[4] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R.G. Baraniuk, "Analog-to-information conversion via random demodulation," *IEEE Dallas Circuits and Systems Workshop,* Oct. 2006.

[5] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R.G. Baraniuk, "Compressive imaging for video representation and coding," *PCS06,* Beijing, China, April 2006.

[6] V. Stanković, L. Stanković, and S. Cheng, "Compressive sampling of binary images," *CISP08 Congress on Image and Signal Processing*, Sanya, Hainan, China, May 2008.

[7] J. Tropp and A.C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," preprint 2005.

[8] M.F. Duarte, M.B. Wakin, and R.G. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," *Proc. SPARS05 Signal Processing with Adaptive Sparse Structure Representation,* Rennes, France, Nov. 2005.

[9] G. Cormode and S. Muthukrishnan, "Towards an algorithmic theory of compressed sensing," *DIMACS Tech. Report* 2005-40, 2005.