COMPRESSIVE CODED APERTURE VIDEO RECONSTRUCTION

Roummel F. Marcia and Rebecca M. Willett

Department of Electrical and Computer Engineering, Duke University
Box 90291, Durham, NC 27708, USA
phone: + (1) 919-323-6734, fax: + (1) 919-660-5293, email: roummel@ee.duke.edu
web: www.ee.duke.edu/roummel

ABSTRACT

This paper concerns compressive sensing methods for overcoming the pixel-limited resolution of digital video imaging systems. Recent developments in coded aperture mask designs have led to the reconstruction of static images from a single, low-resolution, noisy observation image. Our methods apply these coded mask designs to each video frame and use compressive sensing optimization techniques for enhanced resolution digital video recovery. We demonstrate that further improvements can be attained by solving for multiple frames simultaneously, even when the total computation time budget is held fixed.

1. INTRODUCTION

This paper describes a novel technique for increasing the resolution of digital video using a combination of coded aperture sensing and wavelet-based reconstruction algorithms. Many existing techniques for increasing video resolution are based on the concept of superresolution image reconstruction, in which several low resolution, noisy, slightly shifted observations are used to reconstruct an image of the underlying high resolution scene [1]. Other techniques are based upon image upsampling and interpolation methods [2]. These approaches, however, typically either require relatively large numbers of observed pixels or yield unsatisfactory edges and boundaries. In contrast, the pseudorandom coded aperture proposed in this paper facilitates the accurate reconstruction of high resolution video from relatively low resolution observed video frames.

As reported in [3] and reviewed below, coded aperture masks based on recent theoretical work on Toeplitz-structured matrices for compressive sensing can be used to aid reconstruction of high-resolution static images from low-resolution coded observations. In the video setting, however, there is a tradeoff between the sparsity of the vector being reconstructed and computation time, and this tradeoff varies with the number of video frames processed simultaneously. In particular, if each frame were to be processed independently, then the problem size would be relatively small and each iteration of a reconstruction method would require little computation time. In contrast, if a block of sequential frames were processed simultaneously, then the size of the problem would increase linearly with the block size and the computation time required for each iteration would likewise increase, but the solution vector could exploit inter-frame correlations, be significantly more sparse than in the single frame setting, and hence yield more accurate reconstructions.

We explore this tradeoff in detail in this paper by proposing several reconstruction algorithms designed to take advantage of similarities between frames and comparing their performances when the per-frame computation time is held constant. In this setup, the smaller video blocks can be processed using more iterations of an iterative reconstruction method, more closely approaching convergence of the optimization algorithm. We demonstrate that despite this advantage, shorter blocks of video frames yield less accurate reconstructions within the time budget in both high- and low-noise settings.

The authors were partially supported by DARPA Contract No. HR0011-04-C-0111, ONR Grant No. N00014-06-1-0610, and DARPA Contract No. HR0011-06-C-0109.

2. PROBLEM FORMULATION

2.1 Compressive sensing

Nonlinear image reconstruction based upon sparse representations of images has received widespread attention recently with the advent of "compressive sensing". This emerging theory indicates that very high dimensional vectors ($\mathbf{f} \in \mathbb{R}^N$, where $N = n^2$) can be recovered with astounding accuracy from a much smaller dimensional observation ($\mathbf{y}$) when $\mathbf{f}$ has a "sparse" representation in some basis $\mathbf{W}$ (*i.e.,* $\mathbf{f} = \mathbf{W}\theta$ where θ has few non-zero coefficients), subject to a *Restricted Isometry Property* (RIP) [4] condition (described below) on the product of the observation matrix $\mathbf{R}$ and the basis matrix $\mathbf{W}$. We observe $\mathbf{y} = \mathbf{R}\theta + \mathbf{n}$, where $\mathbf{n}$ is white Gaussian noise. The $\ell^2 - \ell^1$ minimization

$$\begin{aligned} \theta^* &= \arg\min_{\theta} \frac{1}{2}\|\mathbf{y} - \mathbf{R}\mathbf{W}\theta\|_2^2 + \tau\|\theta\|_1 \qquad (1) \\ \mathbf{f}^* &= \mathbf{W}\theta^* \end{aligned}$$

will yield a highly accurate estimate of $\mathbf{f}$ with very high probability [5, 6]. The regularization parameter $\tau > 0$ helps to overcome the ill-posedness of the problem. The ℓ^1 penalty term drives small components of θ to zero and helps create sparse solutions.

The observation matrix $\mathbf{R}$ is said to satisfy the RIP of order $3m$ if, for $T \subset \{1, 2, \cdots, N\}$ and $\mathbf{R}_T$, a submatrix obtained by retaining the columns of $\mathbf{R}$ corresponding to the indices in T, there exists a constant $\delta_{3m} \in (0, 1/3)$ such that for all $z \in \mathbb{R}^{|T|}$,

$$(1 - \delta_{3m})\|z\|_2^2 \leq \|\mathbf{R}_T z\|_2^2 \leq (1 + \delta_{3m})\|z\|_2^2 \qquad (2)$$

holds for all subsets T with $|T| \leq 3m$ [4]. An observation matrix $\mathbf{R}$ satisfying RIP with high probability is often referred to as a compressed sensing (CS) matrix. While the RIP cannot be verified for a given $\mathbf{R}$, it has been shown that matrices with entries drawn independently from some probability distributions satisfy the condition with high probability when $k \geq Cm\log(N/m)$ for some constant C, where $m \equiv \|\theta\|_{\ell_0}$ is the number of non-zero elements in the vector θ [4].

2.2 Compressive coded aperture mask designs

Conventional coded aperture imaging masks have not been designed with respect to the compressive sensing framework, but rather to increase the amount of light hitting a detector while allowing for optimal *linear* reconstruction methods. The basic idea is to create a mask pattern which introduces a more complicated point spread function than that associated with a pinhole, and exploit this pattern to reconstruct high-quality image estimates. Seminal work in coded aperture imaging includes the development of Modified Uniformly Redundant Arrays (MURAs) [7], masks for Hadamard transform optics [8], and pseudorandom phase masks [9]. While MURA and Hadamard coded apertures are successful in the context of linear reconstruction, there exists a wide variety of nonlinear reconstruction

methods which can dramatically outperform linear reconstructions when $\mathbf{f}$ has a sparse representation in some basis.

A recent study by the authors [3] addressed the accurate reconstruction of a high resolution static image which has a sparse representation in some basis from a single low resolution observation using *compressive* coded aperture imaging. In the following, the observation $\mathbf{y}$ is assumed to be given by $\mathbf{y} = \mathbf{D}(\mathbf{f} * \mathbf{h}) + \mathbf{n}$, where $\mathbf{D}$ is a downsampling operator and $\mathbf{h}$ is a point-spread function (PSF). The downsampling operator $\mathbf{D}$ corresponds to averaging the intensity values of 16 adjacent pixels forming a square, *i.e.* , using a 4×4 boxcar filter. Let $\mathscr{F}$ be the matrix corresponding to fast Fourier transform, and denote the Fourier transform of $\mathbf{h}$ by $\mathbf{H}$, *i.e.*, $\mathscr{F}\mathbf{h} = \mathbf{H}$. Let $\mathbf{C_H}$ be a diagonal matrix whose diagonal components are the entries in $\mathbf{H}$. We note that multiplication of a vectorized image by $\mathbf{C_H}$ is equivalent to element-wise matrix multiplication of the image by $\mathbf{H}$. Then the observation $\mathbf{y}$ is given by $\mathbf{y} = \mathbf{R}\mathbf{f} + \mathbf{n}$, where $\mathbf{R}$ is the linear operator

$$\mathbf{R} = \mathbf{D}\mathscr{F}^{-1}\mathbf{C_H}\mathscr{F}. \tag{3}$$

In a conventional coded aperture imaging setup, we assume the PSF $\mathbf{h}$ is related to a mask p by

$$\mathbf{h} = p * b, \tag{4}$$

where b is a blur associated with the optics of the imaging system away from the coding element.

In [3], masks were designed such that the corresponding observation matrix $\mathbf{R}$ in (3) satisfies the RIP. Specifically, a method for randomly generating a mask p was developed so that the corresponding matrix product $\mathscr{F}^{-1}\mathbf{C_H}\mathscr{F}$ is block-circulant:

$$\mathscr{F}^{-1}\mathbf{C_H}\mathscr{F} = \begin{pmatrix} A_n & A_{n-1} & \cdots & A_2 & A_1 \\ A_1 & A_n & \cdots & A_3 & A_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ A_{n-1} & A_{n-2} & \cdots & \cdots & A_n \end{pmatrix},$$

where each $A_j \in \mathbb{R}^{n\times n}$ is circulant; *i.e.* , A_j is of the form

$$A_j = \begin{pmatrix} a_n & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_n & \cdots & a_3 & a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & \cdots & a_n \end{pmatrix}.$$

Block-circulant matrices are known to be a compressed sensing matrix, and based upon recent theoretical work on Toeplitz-structured matrices for compressive sensing, the proposed masks are fast and memory-efficient to compute. (See [3, 10] for details).

2.3 Video setting

The problem of accurately reconstructing a high resolution video from low resolution observations can be modeled mathematically by solving a sequence of linear systems

$$\mathbf{y}_t = \mathbf{R}_t\mathbf{W}\theta_t + \mathbf{n}_t \tag{5}$$

where $\mathbf{y}_t$ is the observation, $\mathbf{R}_t$ is the observation matrix, $\mathbf{W}\theta_t$ is the reconstruction, and $\mathbf{n}_t$ is zero-mean white Gaussian noise, for each time frame t. For judiciously chosen coded aperture masks p such that the corresponding observation matrix given by (3) satisfies the RIP with high probability, solving a minimization problem akin to (1) will yield highly accurate reconstructions according to compressive sensing theory. However, when the underlying scene is changing slowly relative to the frame rate, it is possible to dramatically improve upon naïvely processing each frame independently.

3. SPARSE REPRESENTATION ALGORITHMS

In this section, we propose various ways of formulating the $\ell^2 - \ell^1$ minimization problem for the compressed sensing problem associated with (5). Our goal is to solve (5) efficiently for a *sequence* of closely related video frame images $\mathbf{f}_t = \mathbf{W}\theta_t$.

Method A. We formulate the reconstruction problem as a sequence of $\ell^2 - \ell^1$ minimization problems associated with compressive sensing:

$$\theta_t^* \;=\; \arg\min_{\theta_t} \frac{1}{2}\|\mathbf{y}_t - \mathbf{R}_t\mathbf{W}\theta_t\|_2^2 + \tau\|\theta_t\|_1. \tag{6}$$

For a scene that changes only slightly from frame to frame (*i.e.*, $\theta_t \approx \theta_{t+1}$), the reconstruction from a previous frame is often a good approximation to the following frame. In Method A, we use the solution (θ_t^*) to (6) at the t^{th} frame as the initial value (θ_{t+1}^0) for the optimization problem for the $(t+1)^{\text{th}}$ frame. Thus, relatively few iterations will be needed for each optimization problem.

Methods B and C. Methods B and C form a family of algorithms that improve upon the Method A approach by solving for multiple frames simultaneously. For solving two frames simultaneously, rather than solving the minimization problem

$$\underset{\theta_t,\theta_{t+1}}{\text{minimize}}\ \frac{1}{2}\left\|\begin{bmatrix}\mathbf{y}_t\\ \mathbf{y}_{t+1}\end{bmatrix} - \widetilde{\mathbf{R}}_{t,2}\begin{bmatrix}\mathbf{W} & 0\\ 0 & \mathbf{W}\end{bmatrix}\begin{bmatrix}\theta_t\\ \theta_{t+1}\end{bmatrix}\right\|_2^2 + \tau\left\|\begin{bmatrix}\theta_t\\ \theta_{t+1}\end{bmatrix}\right\|_1,$$

with

$$\widetilde{\mathbf{R}}_{t,2} = \begin{bmatrix}\mathbf{R}_t & 0\\ 0 & \mathbf{R}_{t+1}\end{bmatrix}, \tag{7}$$

which does not relate the solutions θ_t^* and θ_{t+1}^*, *i.e.*, the objective function is *separable*, we solve the *coupled* optimization problem instead:

$$\underset{\theta_t,\Delta\theta_t}{\text{minimize}}\ \frac{1}{2}\left\|\begin{bmatrix}\mathbf{y}_t\\ \mathbf{y}_{t+1}\end{bmatrix} - \widetilde{\mathbf{R}}_{t,2}\widetilde{\mathbf{W}}_2\begin{bmatrix}\theta_t\\ \Delta\theta_t\end{bmatrix}\right\|_2^2 + \tau\left\|\begin{bmatrix}\theta_t\\ \Delta\theta_t\end{bmatrix}\right\|_1, \tag{8}$$

where $\widetilde{\mathbf{R}}_{t,2}$ is as in (7), and

$$\widetilde{\mathbf{W}}_2 = \begin{bmatrix}\mathbf{W} & 0\\ \mathbf{W} & \mathbf{W}\end{bmatrix}. \tag{9}$$

This formulation not only computes a sparse solution θ_t^* such that $\mathbf{y}_t \approx \mathbf{R}_t\mathbf{W}\theta_t^*$, but it also computes a sparse $\Delta\theta_t^*$ such that

$$\mathbf{y}_{t+1} \approx \mathbf{R}_{t+1}(\mathbf{W}\theta_t^* + \mathbf{W}\Delta\theta_t^*) = \mathbf{R}_{t+1}\mathbf{W}(\theta_t^* + \Delta\theta_t^*).$$

Thus,

$$\theta_{t+1}^* \approx \theta_t^* + \Delta\theta_t^*,$$

and the initial point for the optimization for the next frame, given by $\theta_{t+1}^0 \equiv \theta_t^* + \Delta\theta_t^*$, is a very accurate estimate of its solution θ_{t+1}^*. Moreover, since $\theta_{t+1}^* \approx \theta_t^*$, then $\Delta\theta_t^* \approx (\theta_{t+1}^* - \theta_t^*)$ is very sparse compared to θ_{t+1}^*, which makes it even better suited to the sparsity-inducing ℓ^1-penalty term in (8).

Methods B and C are different formulations of correlating the coefficients of the current frame with those of the subsequent frames. For solving $k > 2$ frames simultaneously, the optimization problem (8) can be written more generally as

$$\underset{\widetilde{\theta}_{t,k}}{\text{minimize}}\ \frac{1}{2}\left\|\widetilde{\mathbf{y}}_t - \widetilde{\mathbf{R}}_{t,k}\widetilde{\mathbf{W}}_k\widetilde{\theta}_{t,k}\right\|_2^2 + \tau\left\|\widetilde{\theta}_{t,k}\right\|_1, \tag{10}$$

where

$$\widetilde{\mathbf{y}}_{t,k} = \begin{bmatrix}\mathbf{y}_t\\ \mathbf{y}_{t+1}\\ \vdots\\ \mathbf{y}_{t+k-1}\end{bmatrix}, \quad \widetilde{\mathbf{R}}_{t,k} = \begin{bmatrix}\mathbf{R}_t & & & \\ & \mathbf{R}_{t+1} & & \\ & & \ddots & \\ & & & \mathbf{R}_{t+k-1}\end{bmatrix}, \tag{11}$$

and $\widetilde{\theta}_t = [\theta_t\,;\,\Delta\theta_t\,;\,\ldots\,;\,\Delta\theta_{t+k-2}\,]$. Methods B and C differ by how $\widetilde{\mathbf{W}}_k$ is defined and what $\widetilde{\theta}_{t,k}$ represents. To illustrate this difference, we define $\widetilde{\mathbf{W}}_k$ and $\widetilde{\theta}_{t,k}$ for the methods (denoted Methods B4 and C4) for solving $k = 4$ frames simultaneously.

Method B4. The matrix $\widetilde{\mathbf{W}}_k$ for the four-frame system can be defined by extending $\widetilde{\mathbf{W}}_2$ by $\mathbf{W}$ along the first block column and along the block diagonal:

$$\widetilde{\mathbf{W}}_4 = \begin{bmatrix} \mathbf{W} & & & \\ \mathbf{W} & \mathbf{W} & & \\ \mathbf{W} & & \mathbf{W} & \\ \mathbf{W} & & & \mathbf{W} \end{bmatrix}.$$

Then the solution $\widetilde{\theta}_t^*$ of (10) satisfies

$$\begin{aligned} \mathbf{y}_{t+1} &\approx \mathbf{R}_{t+1}\mathbf{W}(\theta_t^* + \Delta\theta_t^*) \\ \mathbf{y}_{t+2} &\approx \mathbf{R}_{t+2}\mathbf{W}(\theta_t^* + \Delta\theta_{t+1}^*) \\ \mathbf{y}_{t+3} &\approx \mathbf{R}_{t+3}\mathbf{W}(\theta_t^* + \Delta\theta_{t+2}^*). \end{aligned}$$

Thus,

$$\begin{aligned} \theta_{t+1}^* &\approx \theta_t^* + \Delta\theta_t^* \\ \theta_{t+2}^* &\approx \theta_t^* + \Delta\theta_{t+1}^* \\ \theta_{t+3}^* &\approx \theta_t^* + \Delta\theta_{t+2}^*. \end{aligned}$$

These highly accurate estimates can then be used to initialize the optimization problem for the $(t+1)^{\text{th}}$ frame. For a slowly moving scene,

$$\begin{aligned} \Delta\theta_{t+1}^* &\approx \theta_{t+2}^* - \theta_t^* \\ \Delta\theta_{t+2}^* &\approx \theta_{t+3}^* - \theta_t^* \end{aligned}$$

are still sparse.

Method C4. Since consecutive frames are more likely to be more similar, the sparsity of $\Delta\theta_{t+1}^*$ and $\Delta\theta_{t+2}^*$ can be further increased. By letting

$$\widetilde{\mathbf{W}}_4 = \begin{bmatrix} \mathbf{W} & & & \\ \mathbf{W} & \mathbf{W} & & \\ \mathbf{W} & \mathbf{W} & \mathbf{W} & \\ \mathbf{W} & \mathbf{W} & \mathbf{W} & \mathbf{W} \end{bmatrix},$$

the solution $\widetilde{\theta}_t^*$ satisfies

$$\begin{aligned} \mathbf{y}_{t+1} &\approx \mathbf{R}_{t+1}\mathbf{W}(\theta_t^* + \Delta\theta_t^*) \\ \mathbf{y}_{t+2} &\approx \mathbf{R}_{t+2}\mathbf{W}(\theta_t^* + \Delta\theta_t^* + \Delta\theta_{t+1}^*) \\ \mathbf{y}_{t+3} &\approx \mathbf{R}_{t+3}\mathbf{W}(\theta_t^* + \Delta\theta_t^* + \Delta\theta_{t+1}^* + \Delta\theta_{t+2}^*). \end{aligned}$$

Thus,

$$\begin{aligned} \theta_{t+1}^* &\approx \theta_t^* + \Delta\theta_t^* \\ \theta_{t+2}^* &\approx \theta_{t+1}^* + \Delta\theta_{t+1}^* \approx \theta_t^* + \Delta\theta_t^* + \Delta\theta_{t+1}^* \\ \theta_{t+3}^* &\approx \theta_{t+2}^* + \Delta\theta_{t+2}^* \approx \theta_t^* + \Delta\theta_t^* + \Delta\theta_{t+1}^* + \Delta\theta_{t+2}^*. \end{aligned}$$

Note that

$$\begin{aligned} \Delta\theta_{t+1}^* &\approx \theta_{t+2}^* - \theta_{t+1}^* \\ \Delta\theta_{t+2}^* &\approx \theta_{t+3}^* - \theta_{t+2}^*, \end{aligned}$$

which may have greater sparsity than those corresponding to Method B4.

Methods Bk and Ck. For arbitrary $k \geq 2$, the matrix $\widetilde{\mathbf{W}}_k$ and $\widetilde{\theta}_{t,k}$ for Methods B*k* and C*k* are similarly defined as in Methods B4 and C4. We note that for $k = 2$, Methods B2 and C2 are identical since $\widetilde{\mathbf{W}}_2$ and $\widetilde{\theta}_{t,2}$ are defined exactly same for both methods. Thus we denote this method by Method B2/C2.

4. SIMULATION RESULTS

4.1 Sparse representation optimization algorithm

In these experiments, we solve the optimization problems for the various proposed methods using the Gradient Projection for Sparse Reconstruction (GPSR) algorithm of Figueiredo et al. [11]. GPSR is a gradient-based optimization method that is very fast, accurate, and efficient. In addition, GPSR has a *debiasing* phase, where upon solving the $\ell^2 - \ell^1$ minimization problem, it fixes the non-zero pattern of the optimal $\theta_t^{(1)}$ and $\theta_t^{(2)}$ and minimizes the ℓ^2 term of the objective function, resulting in a minimal error in the reconstruction while keeping the number of non-zeros in the wavelet coefficients at a minimum. It has been shown to outperform many of the state-of-the-art codes for solving the $\ell^2 - \ell^1$ minimization problem or its equivalent formulations.

4.2 Numerical simulation

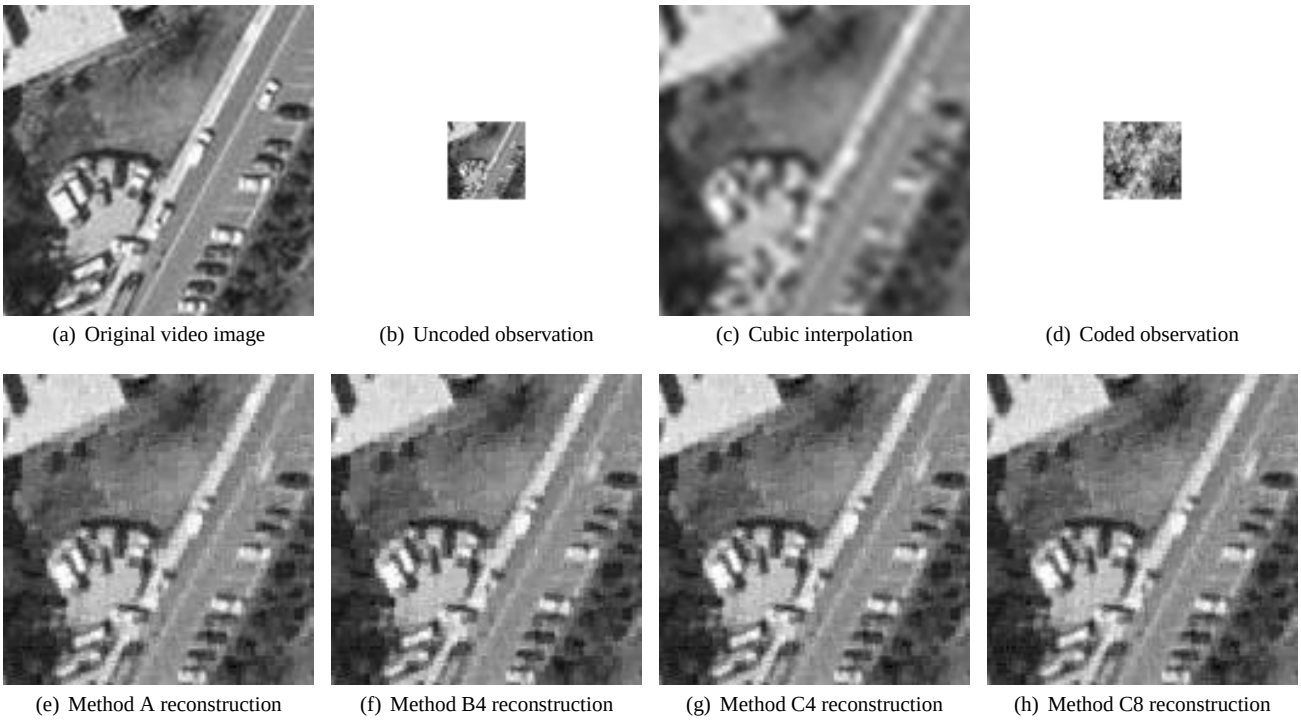
We compare the performance of several methods described in the previous section. In particular, we consider methods solving several different number of frames simultaneously: Methods A (1 frame), B2/C2 (2 frames), B4 and C4 (4 frames), C8 (8 frames) and C12 (12 frames). The video consists of 50 256×256 gray-scale frames depicting the movement of a vehicle on a street at the Duke University campus (see Fig.1(a)). In the video, the vehicle moves with the campus background remaining stationary. To simulate a low-resolution video image, each frame is downsampled by four in each dimension. Thus, the observed scene is 64×64 in pixel resolution (see Figs.1(b) and 1(d)). Each observation y_t in (5) is obtained using a single mask p_t in (4). Using different masks ensures that the optimization problems (6) and (10) are well-posed. We limit our experiments to four unique masks and rotate through these masks (*i.e.*, $p_{t+4} = p_t, p_{t+5} = p_{t+1}$, etc.) to collect the coded observations. More masks could further increase the well-posedness of the problem, but would be difficult to implement in practical optical systems without sophisticated electro-mechanical devices.

The maximum number of iterations for each method are chosen so that the aggregate time per frame for each method is approximately the same. In our first numerical experiment, 5 seconds of GPSR calculations are allotted per frame. In a Linux workstation with dual core Intel Xeon processors running at 3.0 GHz, Method A took approximately 110 GPSR iterations (55 main iterations and 55 debiasing iterations); Method B2/C2 took 48 GPSR iterations (24 main iterations and 24 debiasing); both Methods B4 and C4 took 22 (11 main and 11 debiasing); Method C8 took 9 (4 main and 5 debiasing); and Method C12 took 4 (2 main and 2 debiasing). These numbers are consistent with the amount of computational work needed in GPSR, where the multiplication by the observation matrix $\mathbf{R}$ and the basis matrix $\mathbf{W}$ take up the bulk of the computation: Method B2/C2 requires twice as much matrix-vector multiplications as Method A, and Methods B4 and C4 requires four times as much. The startup cost for GPSR is higher for the larger frame systems (Methods C8 and C12), and the computational scaling is, thus, skewed for these methods.

Numerical experiments with noise variances $\sigma_{\mathbf{n}} = 0.1$ and 1.0 were run, *i.e.*, $|\mathbf{n}_t| \leq \sigma_{\mathbf{n}}$ in (5). The results are shown in Figs.2(a) and 2(b). We repeat the experiment with $\sigma_{\mathbf{n}} = 1.0$ but allotting 20 seconds of GPSR calculations per frame. These results are shown in Fig. 2(c).

4.3 Results and analysis

We first analyze the results for $\sigma_{\mathbf{n}} = 0.1$. (The results for $\sigma_{\mathbf{n}} = 1.0$ are similar.) As a point of comparison, we used cubic interpolation (see Fig.1(c)) to the uncoded observation (Fig.1(b)) to determine whether improvements were made using compressive coded aperture. The MSE values for the cubic interpolation at each frame are relatively constant, with the final MSE value (50^{th} frame) being 0.1428. The MSE values for the proposed methods for each frame is shown in Fig.2. By the fourth frame, all methods already



(a) Original video image (b) Uncoded observation (c) Cubic interpolation (d) Coded observation

(e) Method A reconstruction (f) Method B4 reconstruction (g) Method C4 reconstruction (h) Method C8 reconstruction

Figure 1: Images at frame $t = 50$ from the numerical experiments with $\sigma_{\mathbf{n}} = 1.0$ and 20 sec/frame. (a) Original frame image. (b) Low-resolution observation with no coding: original image downsampled by a factor of 16. (c) Reconstruction using cubic interpolation (MSE = 0.1428). (d) Observed coded image. (e) Reconstruction using Method A (MSE = 0.1076). (f) Reconstruction using Method B4 (MSE = 0.0999). (g) Reconstruction using Method C4 (MSE = 0.0955). (h) Reconstruction using Method C8 (MSE = 0.0937).

have MSE values less than the final MSE value of the cubic interpolation, which indicates that reconstruction methods that use coded aperture improve on that which only uses uncoded cubic interpolation. Unlike those for the cubic interpolation, the MSE values for each method monotonically decrease, which means that the accuracy of the reconstruction improves with the progression of the video. Also, the steep drops in MSE values for the proposed methods in the first several frames are evidence that using the solution for one frame as the initial value for optimization in the next frame improves overall performance.

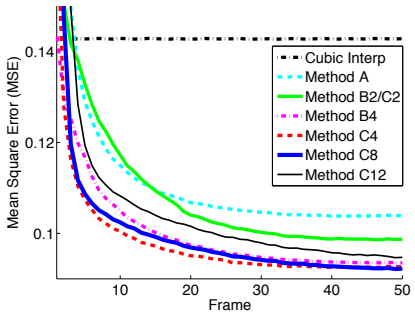
Figure 2 shows the convergence history of the various proposed methods, and Table 1 lists the MSE values for each method at the final (50^{th}) frame. The behavior of the MSE values for the 1-frame approach Method A and the 2-frame approach Method B2/C2 are comparable in the first 20 frames, after which Method B2/C2 begins to outperform Method A. The 4-frame approaches Methods B4 and C4 outperform Methods A and B2/C2 from the beginning. The steeper drop in MSE values for Methods B4 and C4 in the first several frames is consistent with the observation that the initial values from these methods are more accurate than those from Methods A and B2/C2. For both values of $\sigma_{\mathbf{n}}$, Method C4 performs better than B4. The improvement of Method C4's performance over Method B4's may be attributed to the greater sparsity of the Method C4 solution for each optimization problem. The performances of Methods C4 and the 8-frame approach Method C8 are comparable, but Method C8 appears to be more robust to noise. The differences in performances for each method, especially among Methods B4, C4 and C8, are slightly more pronounced for the simulations with larger noise variance $\sigma_{\mathbf{n}} = 1.0$ (see Table 1). The relatively poor performance of the 12-frame approach Method C12 may seem surprising at first glance; however, it can easily be attributed to the number of GPSR iterations possible within the computation time budget. In our particular setup, only two GPSR main iterations and two debiasing iterations are possible in the 12-frame case, and a large fraction of the computation time budget is used for memory allocation and variable initialization.

For the numerical experiments allowing 20 seconds per video frame (four times the amount in the previous numerical experiment), the same general performance for each method holds with one notable exception: the performance of Method C12 vastly improves and is even comparable to the best performing method in the previous experiment (Method C8). In this setup, Method C12 is no longer limited to the restrictive 2-main and 2-debiasing iterations-per-frame constraint. Rather, significant improvements can be now achieved at each time frame since Method C12 is allowed many more GPSR iterations per optimization problem (in this case, 13 GPSR main iterations and 13 debiasing iterations). Thus, solving for numerous frames simultaneously is clearly advantageous provided that the number of iterations within the time allotted for each optimization problem is not too small.

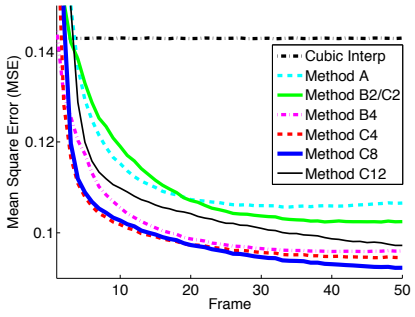
Qualitatively, the reconstructions from the proposed methods (Figs.1(e)-1(h)) more accurately recover details of the original video (Fig.1(a)) than that from cubic interpolation (see Fig.1(c)). For example, street and parking lines are more defined, the separation between vehicles are more pronounced, and the edges of the building in the upper left corner are crisper. The videos from the numerical simulations are available for download at `http://www.ee.duke.edu/nislab/eusipco2008`.

5. CONCLUSION

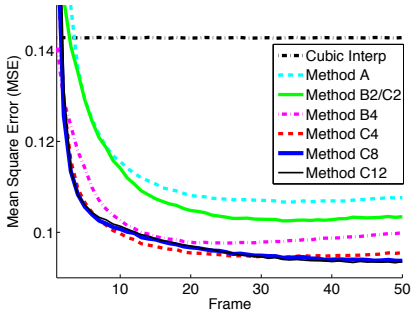
In this paper, we present compressive sensing methods for overcoming the pixel-limited resolution of digital video imaging systems. Our methods apply coded mask designs for each video frame and use sparse representation optimization techniques for signal recovery. The methods and experiments presented in this paper demonstrate that pseudorandom coded masks can yield accurate high reso-



(a) $\sigma_{\mathbf{n}} = 0.1$ and 5 sec/frame time budget



(b) $\sigma_{\mathbf{n}} = 1.0$ and 5 sec/frame time budget



(c) $\sigma_{\mathbf{n}} = 1.0$ and 20 sec/frame time budget

Figure 2: MSE values for proposed methods for observations with different noise variances $\sigma_{\mathbf{n}}$ and time budgets.

lution digital video from a sequence of low resolution coded frames. In particular, when the amount of processing time allotted per frame is held constant, the accuracy generally increases with the number of frames processed simultaneously in a block processing algorithm which exploits similarities between subsequent frames. The improvement in accuracy only abates when the size of the problem is such that only a very small number of reconstruction iterations can be run within the allotted time. All the approaches presented in this paper outperform frame-wise upsampling or interpolation.

The results presented in this paper have an alternative interpretation. If the desired accuracy is held fixed, then the amount of processing time required to achieve that accuracy is in general *smaller* when the block size (the number of frames processed simultaneously) is *larger*. This somewhat counterintuitive result demonstrates the importance of exploiting inter-frame correlations, even when it means increasing the size of the optimization problem.

	5 sec/frame		20 sec/frame
	$\sigma_{\mathbf{n}} = 0.1$	$\sigma_{\mathbf{n}} = 1.0$	$\sigma_{\mathbf{n}} = 1.0$
Cubic Interp	0.1428	0.1428	0.1428
Method A	0.1039	0.1066	0.1076
Method B2/C2	0.0987	0.1025	0.1034
Method B4	0.0935	0.0960	0.0999
Method C4	0.0926	0.0945	0.0955
Method C8	0.0922	0.0923	0.0937
Method C12	0.0947	0.0972	0.0936

Table 1: MSE values for the final (50^{th}) frame for each method in the numerical experiments.

REFERENCES

[1] R. C. Hardie, K. J. Barnard, and E.E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1633, 1997.

[2] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," in *Handbook of medical imaging*, pp. 393–420. Academic Press, Inc., Orlando, FL, USA, 2000.

[3] R. F. Marcia and R. M. Willett, "Compressive coded aperture superresolution image reconstruction," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2008.

[4] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 15, no. 12, pp. 4203–4215, 2005.

[5] E. J. Candès and T. Tao, "The Dantzig selector: statistical estimation when *p* is much larger than *n*," *Annals of Statistics*, 2005, To appear.

[6] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[7] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. Opt.*, vol. 28, 1989.

[8] N. J. A. Sloane and M. Harwit, "Masks for Hadamard transform optics, and weighing designs," *Appl. Opt.*, vol. 15, no. 1, pp. 107, 1976.

[9] A. Ashok and M. A. Neifeld, "Pseudorandom phase masks for superresolution imaging from subpixel shifting," *Appl. Opt.*, vol. 46, no. 12, pp. 2256–2268, 2007.

[10] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. of Stat. Sig. Proc. Workshop*, 2007.

[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: Special Issue on Convex Optimization Methods for Signal Processing*, To appear.