Optically compressed image sensing using random aperture coding

A. Stern*[a], Yair Rivenson [b] , Bahram Javidi [c]

[a]Electro Optical Unit, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel;
[b]Department of Electrical and Computer Engineering, Ben Gurion University of the
Negev, Beer-Sheva 84105, Israel
[c]Department of Electrical and Computer Engineering, University of Connecticut,
Storrs, Connecticut 06269-1157

ABSTRACT

The common approach in digital imaging today is to capture as many pixels as possible and later to compress the captured image by digital means. The recently introduced theory of *compressed sensing* provides the mathematical foundation necessary to change the order of these operations, that is, to compress the information before it is captured. In this paper we present an optical implementation of compressed sensing. With this method, a compressed version of an object's image is captured directly. The compression is accomplished by optical means with a single exposure. One implication of this imaging approach is that the effective space-bandwidth-product of the imaging system is larger than that of conventional imaging systems. This implies, for example, that more object pixels may be reconstructed and visualized than the number of pixels of the image sensor.

Keywords: compressive imaging, compressive sensing, compressed sensing, aperture coding, matching pursuit, resolution, stagewise orthogonal matching pursuit.

1. INTRODUCTION

Common digital imaging systems follow the sample-then-compressed framework. According to this framework the imaging system first captures as many pixels as possible. As a result, the captured image is highly redundant. Therefore, the second common step after acquisition is digital compression. The compression is required for storage and communication purposes. Compression techniques exploit the visual redundancy typical to human intelligible images to represent the captured image by less numbers than the number of pixel captured. This way of imaging evokes the question: is it strictly necessary to acquire all the image samples in a pedantic way and then compress them later? Can one capture *optically* capture fewer samples without compromising the quality of the reconstructed image? The answer to this question is positive owing to the recent theory of compressed (or compressive) sensing (CS) theory.[1-5] The basic idea behind CS is that an image can be accurately reconstructed from fewer measurements than the nominal number of pixels if it is compressible by a known transform such as Wavelet or Fourier transform.

The CS theory provides the mathematical background necessary for designing compressive imaging (CI) systems. There are imaging application in which compressing the image before capturing it is beneficial. Some examples of such systems are those in which the acquisition is expensive in terms of hardware (high pixel cost) or acquisition time, or systems that cannot afford digital compression before storage of transmission of the data. The price that is to be paid for implementation of a CS-based imaging system is giving up the convenient structural form of common linear-shift invariant imaging schemes. This implies abandoning conventional imaging design architectures.

Recently, several CI systems were proposed[5-8]. One practical way to implement CS is by capturing random projections of the object and then to apply an appropriate numerical reconstruction algorithm to reconstruct

---

* stern@bgu.ac.il; phone (972)-8-6461840; http://www.ee.bgu.ac.il/~stern/

the visual image. In Ref. 6 a compressed imaging (CI) system is proposed that uses a digital mirror array device to randomly project the image on a single sensor. Successive random exposures are taken by randomly changing the digital mirror array. In Ref. 7 we presented what is, to the best of our knowledge, the first proposed single shot, motion-free CI technique. With this technique the random projection is accomplished by using a randomly coded aperture. In Ref. 8 the CS theory was used for compressed spectral imaging. In Ref. 9 CI is implemented by using a linear sensor scanning the field-of-view by rotational motion. The projections are not random; therefore the compression is less effective. However the imaging architecture is almost similar to conventional ones. In this paper we overview the technique presented in Ref. 7. We further elaborate the technique in Ref. 7 and present new results using a different reconstruction algorithm.

This paper is organized as follows. In section 2 we review the basic concepts behind CS. In Sec. 3 we describe the compressed imaging system proposed in Ref. 7. In Sec. 4 we present reconstructions from simulated compressed images obtained with this compressed imaging system and using a reconstruction technique described in the appendix. Finally, we conclude in section 5, summarizing the main results and discussing future work.

2. COMPRESSIVE IMAGING BY RANDOM PROJECTIONS

A block diagram for CS with random projections is shown in Fig. 1. The object f consisting of N pixels is imaged by taking a set, g, of M random projections. We are interested in the case that $M<N$, meaning that the captured image is undersampled in conventional sense. In our discussion we represent two dimensional object f and captured image g in a lexicographic order, that is, in the form of column vectors of sizes N and M, respectively. We assume that f has a sparse representation in some known domain so that it can be composed by a transform Ψ and only K nonzero coefficients of a vector α, that is $f = \Psi\alpha$ where only K ($K<<N$) entries of α are nonzero. We will refer to such an object as K-sparse object. Many natural images are assumed to be sparse or nearly sparse in some domain. For instance, it is commonly assumed for the purpose of image compression that images are nearly sparse in Fourier or some wavelet domain so that (N-K) coefficients are set to be zero. In the measuring step we take M orthogonal random projections Φ of f. Since $M<N$ we get M compressed sensing measurements $g=\Phi\Psi\alpha$.[1-3] Practically, M has to be at least three times larger than K; $M \geq 3K$.[3] The compression operator Ψ has to be incoherent with the measurement operator Φ, that is, their bases are essentially uncorrelated.[1,2] Fortunately, incoherence property holds for many pair of bases. In particular, it holds with high probability for any arbitrary basis of Ψ and the random projections Φ. For random Gaussian measurements, that is for Φ having zero mean identically independent distributed vectors, there are only $M \geq cK \log(N/K) << N$ required measurements, with c a small constant, to fully recover the N-length original image f. For practical cases with which f is not strictly sparse and Φ not necessarily zero-mean Gaussian random projector, M has to be at least three times larger than K; $M \geq 3K$.[3]
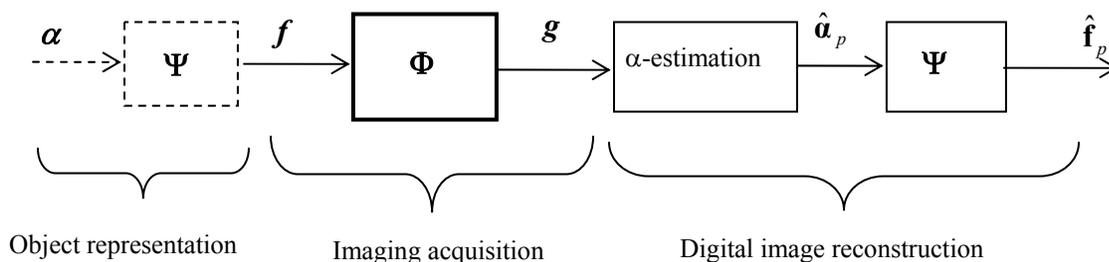

Fig. 1. Imaging scheme of compressed sensing.

In order to reconstruct f we first estimate the coefficients α by solving the following minimizations problem:

$$\hat{\alpha}_p = \min_{\alpha} \| \alpha' \|_p \text{ subject to } \mathbf{g} = \boldsymbol{\Phi}\boldsymbol{\Psi}\alpha = \boldsymbol{\Omega}\alpha \qquad , \qquad (1)$$

where $\boldsymbol{\Omega} \equiv \boldsymbol{\Phi}\boldsymbol{\Psi}$ and $\| . \|_p$ denotes the lp norm defined by $\| \alpha' \|_p = \left(\sum_{i=1}^{N} | \alpha'_p |^p \right)^{1/p}$. Solving (1) we find $\hat{\alpha}_p$ by choosing from all coefficient vectors α' that are related to the measured image by $\mathbf{g} = \boldsymbol{\Phi}\boldsymbol{\Psi}\alpha'$, the one with the minimum p-norm. Sparse solutions for α may be found for p between 0 and 1. With p=0, the l_0 norm operator $\| \alpha' \|_0$ simply counts the number of nonzero entries of α'. In such a case, the reconstruction condition (1) seeks the coefficient vector $\hat{\alpha}_o$ that has the minimum number of nonzero elements such that its corresponding object $\hat{\mathbf{f}}_o = \boldsymbol{\Psi}\hat{\alpha}_o$, after passing through the imaging operator $\boldsymbol{\Phi}$ (Fig. 1), yields the measurement g. In principle, only $M=K+1$ measurements are required to recover the K-sparse signal f with high probability. It can be shown[1,2,10] that the l_0 solution of Prob. (1) yields the sparsest α is f is sufficiently sparse, such that

$$K = \|\alpha\|_0 < \frac{1}{2}\left(1+\frac{1}{\mu\{\boldsymbol{\Omega}\}}\right) \le \frac{1}{2}\left(1+\frac{1}{\mu\{\boldsymbol{\Psi}\}}\right) \qquad (2)$$

where $\mu\{\Omega\}$ is the mutual-coherence defined as the larger absolute normalized inner product between different columns of a matrix Ω:

$$\mu\{\Omega\} \equiv \max_{1 \le i,j \le N, i \ne j} \frac{\left|\left\langle \Omega_i, \Omega_j \right\rangle\right|}{\|\Omega_i\| \cdot \|\Omega_j\|} \qquad (3)$$

Unfortunately, the implementation of the l_0 estimator is unstable and it additionally requires combinatorial enumeration of the $\binom{N}{K}$ possible sparse subspaces, which is prohibitively complex. A more practically approach is estimating f by solving Eq. (1) with p=1 for which traditional linear programming techniques are available,[1-4] such as the Basis Pursuit (BP) algorithm.[1] With condition (2) fulfilled the, linear programming methods for l_1 solution of (1) converge to the desired l_0 solution, that is $\hat{\alpha}_1 = \hat{\alpha}_o$.[1,5,10] Finally, once we find $\hat{\alpha}_0$, the object is reconstructed simply by $\hat{\mathbf{f}}_0 = \boldsymbol{\Psi}\hat{\alpha}_0$.

Another approach for l_1 solution of (1) is via the MP (Matching Pursuit) algorithms, a family of fast greedy algorithms, which were "rediscovered" recently. The new results for MP are comparable with recent results for the Basis Pursuit (BP). The MP algorithms are faster and easier to implement, which makes them an attractive alternative to BP for signal recovery problems. [11]

3. COMPRESSIVE IMAGING USING RANDOM CODED APERTURE

The random projection operator $\boldsymbol{\Phi}$ in Fig.1 can be implemented by employing random aperture coding. Aperture coding was previously used for improving the signal-to-noise ratio, controlling the depth of field and for optical encryption. In Ref. 7 we used aperture coding for accomplishing optical compression. One possible optical setup using such a coded aperture is depicted in Fig. 2. The object is placed at a distance of z_o from the lens. Attached to the lens is a random Gaussian phase screen that randomly encodes the aperture. The scattered light from the random phase screen is collected by a lens with diameter D and focal length f_l. The scattered light reaches an array of CCD detectors, which is located at a distance z_i behind the lens. In Ref. 7 it is shown that if the correlation length, ρ, of the random phase is sufficiently small with

respect to the other dimensions of the imaging system then the imaging operator Φ performs the required random projections. Consequently, Φ and Ψ are incoherent with overwhelming probability[1], as required for CS solution via Eq. 1. It is noted that the compressed image obtained with this system is captured in a single shot. The system is static and no moving or scanning elements are used.
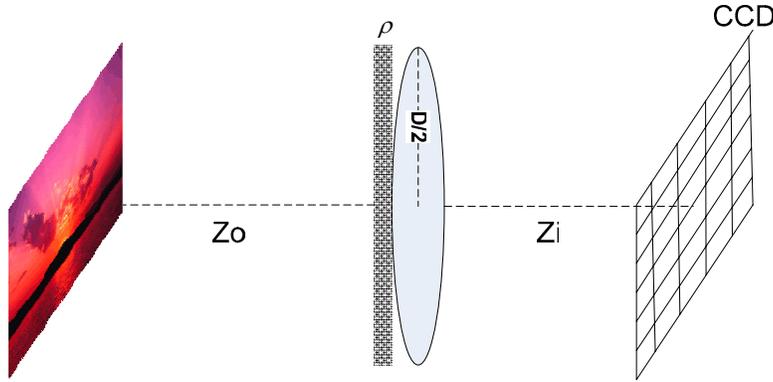

Fig. 2. Single shot compressed imaging scheme. Phase mask with correlation length ρ is attached to a lens with diameter D.

4. SIMULATION RESULTS

We have simulated, using Matlab, images obtained with the CI system shown in Fig. 2. The simulation is carried out by propagating the two-dimensional fields from the object to the image plane according to Fresnel theory. In our simulations we assume that the CCD pixel size is 7.4μm, central wavelength is $\bar{\lambda} = 0.55$ μm, and $z_o=z_i=f_l=140$mm. The random phase mask is assumed to be a random Gaussian phase mask with correlation length of ρ=5.5μm. The lens diameter is D=50mm. These simulation conditions match the random projection requirements listed in the appendix of Ref. 7.

We assume that the object pixel size is 1 mm. Due to computer resources constrains, we limit the object size to be 64x64 pixels. With this object size, the Φ and Ψ matrices are of the order of 4096×4096 elements. Each row in Φ represents a shift variant point spread function of size 4096(=64x64).

In ref. 7 the Matching Pursuit algorithm[i] was used for estimating α in Fig. 1. Here we use an improved version of this algorithm that was recently introduced; the StOMP (Stagewise Orthogonal Matching Pursuit) algorithm.[12] StOMP was specially tailored for random operators Ω, therefore their straight is for solving CS data. In a nutshell, the StOMP algorithm solves the sparse solution problem by calculating a residual from the stage before, backprojects it and determines the dominant entries by thresholding with respect to permitted error. In contrast to the previously developed OMP (Orthogonal Matching Pursuit) algorithm, multiple thresholded entries are permitted. Those entries define indexes of the estimated most significant sparse coefficients. These indexes, together with those estimated in the previous iteration, are used to select a set of columns of Ω that are then used to backproject **g** to obtain the estimated coefficients $\hat{\alpha}^{(s)}$ of iteration steps. The StOMP algorithm is described in more details in the Appendix. In our simulations we used a StOMP implementation based on the SparseLab package.[14]

Figures 3-5 show examples of reconstructed images from simulated compressed images obtained with the above described system. Simulation results of the compressed image and reconstructed image of the "CI" letters shown in Fig. 3(a). The original image in Fig. 3(a) has 64x64 pixels, whereas the captured image in Fig. 3(b) has only 40x40 pixels. It can be seen that due to the random projections, the captured image shown in Fig. 3(b) has absolute no visual meaning. The reconstructed image using the StOMP algorithm is

shown in Fig. 3(c). Note that despite that the captured image in Fig. 3 (b) is represented by only 1550 samples, which are only 36.7% of the original image, perfect reconstruction is obtained. The reconstruction error is MSE$\approx 10^{-6}$.
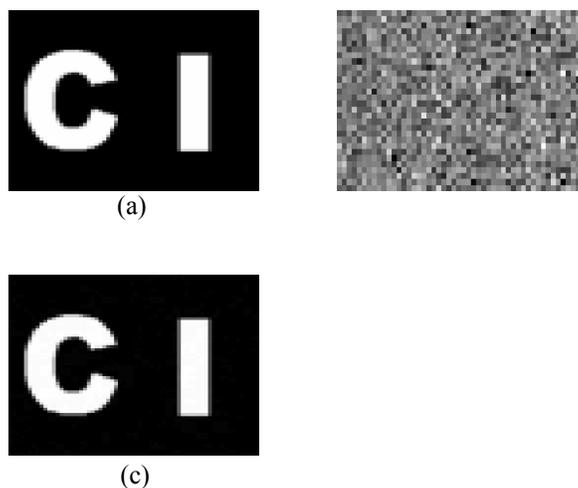

(a)

(c)

Fig. 3. Simulation of CI images. (a) Original image (64x64pixels); (b) Captured image (40X40pixels); (c) Reconstructed image (64x64pixels).

For the reconstruction of Fig. 3(b) we have used the Haar-wavelet transform as our basis for the sparse image representation Ψ. The Haar-wavelet transform, decomposes the image in Fig. 3(a) to a vector α that has only about 880 non-zeroes, so that only approximately 20% of the coefficients are non-zeros ($K/N \cong$ 20%).

The simulation took 2419 seconds to calculate the system's PSF, and 199 seconds to solve the StOMP algorithm on a PC computer with AMD Athlon 64 dual core processor, 3800+, 2GB of RAM, working with Windows XP operating system. In our simulations we found StOMP to be by far the fastest algorithm to solve the SSP, compared to Basis Pursuit (implemented as in the l_1-magic package[14]) and greedy Matching Pursuit algorithm (implemented in Ref. 7).

Figure 4 (a) shows an image of a knife. Figure 4(b) shows the compressed captured image and Fig. 4 (c) shows reconstructed image using the StOMP algorithm. Here again we used Haar-wavelet transforms for Ψ because of the piecewise constant nature of the image. Note that despite the captured image in Fig. 4 (b) being represented by 50% less pixels than the original image, we obtained perfect reconstruction in Fig. 4(c). It can be seen that the complete field of view and full resolution is reconstructed, implying that the entire object space-bandwidth is preserved. The reconstruction error is MSE$\approx 10^{-7}$. This negligible MSE is owing to the fact that the Haar-wavelet transform used as Ψ decomposes the original image to a coefficient vector α that has only K=1031 non-zeroes, that is $K/N \cong 25\%$.
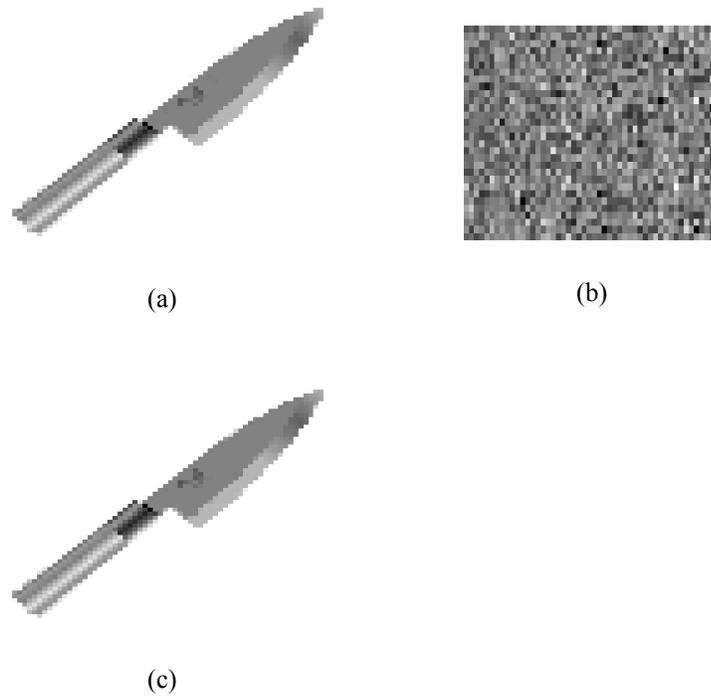
(a)

(b)

(c)

Fig. 4. Simulation for "knife" image. (a) Original image (64x64 pixles). (b) Captured image (45x46 pixels). (c) Reconstructed image (64x64 pixels).

Figure 5 shows results of compressed sensing of complex object image. Figures 5(b) to 5(f) present reconstructed images from compressed image of sizes 2500, 2500, 3000, 35000 and 3800 pixels, which are 48.9%, 61.1%, 73.4%, 85.6% and 92.7% of the nominal (64x64=4096 pixels), respectively. Unlike Figs. 3(a) and 4(a), the gun image is not piecewise constant, and therefore it cannot be compressed efficiently by Haar-wavelet transform. For the reconstructions in Fig. 5 we used the CDF (Cohen-Daubechies-Feaveau) 9/7 wavelet,[16] which we found empirically to be the best among several wavelet transforms Ψ we considered. CDF 9/7 wavelet is well known for its popularity in the JPEG2000 standard. We see that reconstruction from compressed images having 48.9%, 61.1%, 73.4%, less samples than nominal [Figs. 5(b)-(c)] appear blurred and noisy. Images reconstructed from less compressed images, having only 26.6%, 14.4% and 7.3% less samples than nominal [Figs. 5(c)-(d)], are much sharper. The noisy appearance is explained by the fact that unlike the "knife" image (Fig. 4), in which many of its wavelet coefficients are zero, less coefficients of Fig. 5(a) are absolute zero. Many other coefficients have a small value (after the transform), and are being discarded by the StOMP false detection rate thresholding, creating the "noisy" look of the image.
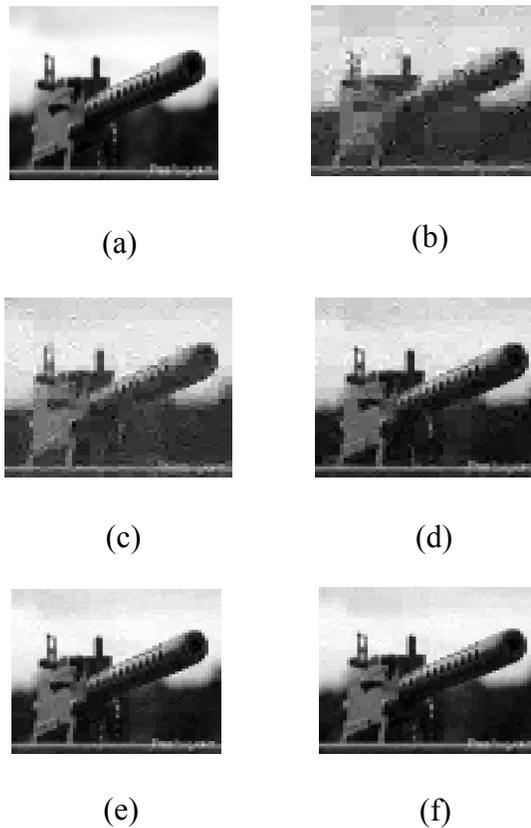
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5. "Machine gun" image, sampled at different size of detector array, using CDF9/7 as Ψ. (a) original (4096 pixels), (b) Reconstruction using 2000 samples, (c) 2500 samples, (d) 3000 samples, (e) 3500 samples, (f) 3800 samples.

When comparing the above obtained results with typical results obtained with digital compression, one needs to keep in mind that the reference images used here are much smaller than those generally considered in digital compression examples (64x4 pixels here versus 256x256 or 512x512 pixels in digital compression). Therefore, the typical images considered in digital compression examples are much more redundant and much more compressible. Consequently, larger compression rates can be obtained for given reconstruction quality. For original images of size 64x64 pixels, as considered in Figs. 3-5, the percentage of compression coefficients (K/N) required for a given reconstruction quality is much larger than for cameramen images, having at least 256x256 pixels as considered in digital compression examples. In other words, there is much less redundancy in figures having 64x64 pixels than in typical images that are much larger. Therefore it is expected that for common images, much larger than those demonstrated here, much larger optical compression can be achieved.

5. SUMMARY AND DISCUSSION

In this work we presented a method for compressive imaging using aperture coding. We overviewed and further elaborated the CI approach recently introduced in Ref. 7. The CI system randomly projects the object field in the image plane with the help of random phase mask. The random phase mask can be viewed as a random scrambler of rays. The compressed image is captured with a single exposure without using

moving elements. Here we presented more accurate simulations of the captured images than in Ref. 7. We also used a more advanced restoration algorithm. Simulations have shown that for synthetic images, exact reconstructions can be obtained from compressed images that have approximately 65% less pixels than the original image. In other words, we obtained optical compression of ~35% with absolute no loss of resolution or field of view. For non synthetic images more samples are required; images having approximately 85% of nominal samples yield satisfactory reconstructions.

It is important to point out that due to computational limitations our results were obtained for small object images, having 64x64 pixels. For larger images we expect better optical compression ratios. The reason is as follows. Empirical studies show that in order to have good reconstructions with CS algorithms[3] the number of captured samples need to be three to five times the number of nonzero coefficients, i.e., $M=3K\div5K$. On the other hand we know from digital compression practice that for regular size images compression rates of 15-25 yield satisfactory reconstructions; that is $K/N \approx 4\%\div6.7\%$. Putting these two facts together infer that compressed optical imaging with compression ratios approximately 15-30% can be expected. However such compression ratio can only be expected for regular sized images. In this work we have obtained poorer optical compression ratios because we used small objects that have much larger K/N ratios and because CS generally works less effectively with a relatively small number of captured samples M.

The compressed imaging technique discussed in this work may be further improved by optimizing the imaging setup and the reconstruction technique. The optical setup shown in Fig. 1 may be further optimized considering different layouts than in Fig. 2. Depending on the type of the sparsity of the object, the reconstruction may be optimized by post processing and by multi-scale compressed sensing.[2,3] The reconstruction algorithm may be accelerated by employing the structure of $\mathbf{\Psi}$,[4] which is beneficial if very large images are considered.

As a final note, we believe that the concept presented in this paper may be extended effectively for three-dimensional imaging because three-dimensional images are highly compressible.[17,18]

6. APPENDIX A- DESCRIPTION OF THE STOMP ALGORITHM[12]

StOMP operates in S stages, building up a sequence of approximations $\alpha_0, \alpha_1, ...$ by removing detected structure from a sequence of residual vectors $r_0, r_1, ...$ Figure 6 gives a diagrammatic representation.
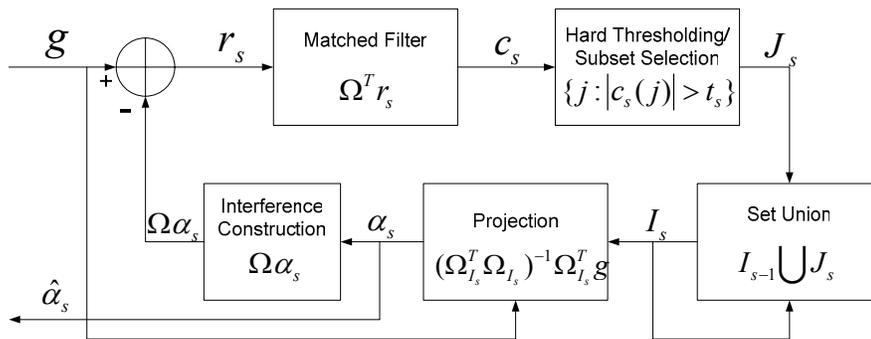

Fig.6 Block diagram of StOMP algorithm (after Ref. 10)

StOMP starts with initial 'solution' $\alpha_0 = 0$ and initial residual $r_0 = g$. The stage counter, s, starts at s = 1. The algorithm also maintains a sequence of estimates $I_0, ..., I_s$ of the locations of the non zeros in α_0. The s-th stage applies matched filtering to the current residual, getting a vector of residual correlations

$$c_s = \Omega^T r_{s-1}$$

$$(4)$$

which is assumed that conatins a small number of significant non zeroes in a vector disturbed by Gaussian noise in each entry. The procedure next performs hard thresholding to find the significant non zeroes; the thresholds, are specially chosen based on the assumption of Gaussianity. Thresholding yields a small set J_s of "large" coordinates:

$$J_s = \{j : |c_s(j)| > t_s \sigma_s\}$$

$$(5)$$

where σ_s is a formal noise level and t_s is a threshold parameter. We merge the subset of newly selected coordinates with the previous support estimate, thereby updating the estimate:

$$I_s = I_{s-1} \bigcup J_s$$

$$(6)$$

We then project the vector y on the columns of Ω belonging to the enlarged support. Letting Ω_I denote the $nx|I|$ matrix with columns chosen using index set I, we have the new approximation α_s supported in I_s with coefficients given by:

$$(\alpha_s)_{I_s} = (\Omega_{I_s}^T \Omega_{I_s})^{-1} \Omega_{I_s}^T g$$

$$(7)$$

The updated residual is

$$r_s = g - \Omega \alpha_s$$

$$(8)$$

We check a stopping condition and, if it is not yet time to stop, we set $s := s + 1$ and go to the next stage of the procedure. If it is time to stop, we set $\hat{\alpha}_s = \alpha_s$ as the final output of the procedure.

REFERNCES

[1] D. L. Donoho, "Compressed sensing", IEEE Transactions on Information Theory , vol 52(4), 1289-1306, April 2006.
[2] E. J. Candes, J. Romberg and T. Tao, "Robust uncertainty Principles: Exact signal reconstruction from highly incomplete frequency information", IEEE Transactions on Information Theory , 52(2), 489- 509 (2006).
[3] Y. Tsaig and D. L. Donoho, "Extensions to compressed sensing", Signal Processing, 86, 549-571 (2006).
[4] R. G. Baraniuk, " Compressive sensing", IEEE signal processing Magazine (lecture notes), 24, 118-121 (July 2007).
[5] M. Elad, "Optimized Projections for Compressed-Sensing", IEEE Trans. on Signal Processing, 22(12), 5695-5702 (Dec 2007).
[6] D. Takhar, J. N. Laska, M. B. Wakin. M. F. Durate, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression", Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose, California (2006).
[7] A. Stern and B. Javidi, "Random projections imaging with extended space-bandwidth product", IEEE/OSA journal of Display Technology, 3(3), 315-320 (Sept. 2007).
[8] M. E. Gehm, R. John, D. J. Bardy, E. M. Willett, and T. J. Schultz, "Single-shot compressive imaging with a dual-disperser architecture", Optic Express, 15(21), 14013-14027 (2007).
[9] A. Stern, "Compressed imaging system with linear sensors", Optics Letters, 32(21), 3077-3079 (Nov. 2007).[
[10] D.L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via l1 minimization", Proc. Nat. Aca. Sci., 100, 2197-2202 (2002).

[11] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A Wide-Angle View at Iterated Shrinkage Algorithms" SPIE (Wavelet XII), San-Diego CA, August, 26-29, 2007.

[12] S. Mallat, and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", IEEE Trans. Sig. Proc., 41 (12), 3397-3415 (1993).

[13] D.L. Donoho, Y. Tsaig, I. Drori, and J.L. Starck, "Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit" , preprint (March 2006)

[14] [SparseLab software package, http://sparselab.stanford.edu/]

[15] [l1-magic software package, http://www.acm.caltech.edu/l1magic/]

[16] I. Daubechies, [Ten Lectures on wavelets], SIAM: Society for Industrial and Applied Mathematics, (June 1992).

[17] S. K. Yeom, A. Stern, and B. Javidi, "Compression of 3D color integral images," Optics Express, 12, 1632-1642 (2004).

[18] E. Elharar, A. Stern, O. Hadar and B. Javidi, "A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform" , IEEE/OSA journal of Display Technology, 3(3), 315-320 (Sept. 2007).