

Model-Assisted Adaptive Recovery of Compressed Sensing with Imaging Applications

Xiaolin Wu, *Fellow, IEEE*, Weisheng Dong, Xiangjun Zhang, and Guangming Shi, *Senior Member, IEEE*

Abstract—In compressive sensing (CS), a challenge is to find a space in which the signal is sparse and, hence, faithfully recoverable. Since many natural signals such as images have locally varying statistics, the sparse space varies in time/spatial domain. As such, CS recovery should be conducted in locally adaptive signal-dependent spaces to counter the fact that the CS measurements are global and irrespective of signal structures. On the contrary, existing CS reconstruction methods use a fixed set of bases (e.g., wavelets, DCT, and gradient spaces) for the entirety of a signal. To rectify this problem, we propose a new framework for model-guided adaptive recovery of compressive sensing (MARX) and show how a 2-D piecewise autoregressive model can be integrated into the MARX framework to make CS recovery adaptive to spatially varying second order statistics of an image. In addition, MARX offers a mechanism of characterizing and exploiting structured sparsities of natural images, greatly restricting the CS solution space. Simulation results over a wide range of natural images show that the proposed MARX technique can improve the reconstruction quality of existing CS methods by 2–7 dB.

Index Terms—Adaptive modeling, autoregressive process, compressive sensing (CS), inverse problem.

I. INTRODUCTION

THE RECENT development of compressive sensing (CS) theory [1], [2] has stirred quite an amount of excitement in signal processing community. Many applications of CS in image processing and computer vision are being explored. The CS theory reveals, in a pleasant surprise, the possibility of reconstructing a signal from a small number of random measurements, as long as the signal has a sparse representation in some space Ψ . Signal $\mathbf{f} = \{f_n\}_{n=1}^N$ of length N is said to be sparse in space Ψ of base $\{\psi_n\}_{1 \leq n \leq N}$ if transform coefficients $\langle \mathbf{f}, \psi_n \rangle$, $1 \leq n \leq N$, are mostly zero or nearly sparse in space Ψ if a dominant portion of these N coefficients are either zero or

very close to zero. The sparsity of \mathbf{f} in Ψ is quantified by the number of significant (nonzero) coefficients K . The signal can be perfectly recovered from $M = O(K \log(N/K))$ observations with high probability.

Given M CS measurements $\mathbf{y} = \Phi \mathbf{f}$, with Φ producing the random projections, the CS recovery of \mathbf{f} from \mathbf{y} is posed as the following constrained optimization problem:

$$\min_{\mathbf{f}} \|\Psi^T \mathbf{f}\|_{l_1} \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{f}. \quad (1)$$

The ℓ_1 minimization problem of (1) can be solved by linear programming [3]. Few other CS recovery algorithms were recently proposed, i.e., gradient projection sparse reconstruction [4], matching pursuit [5], and iterative thresholding [6].

A much celebrated property of CS is its ability to compactly encode signal \mathbf{f} in total blindness of any structures of \mathbf{f} , i.e., the same random projections Φ can be performed on all signals, regardless of the differences in their characteristics. However, this does not mean that one can escape from the issue of adapting or optimizing the CS recovery process to the specific signal \mathbf{f} on hand. Indeed, a thorny issue in practice is what space Ψ should be chosen to recover a particular \mathbf{f} . For example, in signal compression, while conventional methods strive to encode \mathbf{f} in a transform domain that achieves maximum energy packing of \mathbf{f} , CS methods need to recover \mathbf{f} from $\Phi \mathbf{f}$ in space Ψ in which \mathbf{f} exhibits a high degree of sparsity. Thus, from a system point of view, CS merely transfers the task of signal-dependent code optimization from the encoder to the decoder. Finding sparse space Ψ for optimal CS recovery of signal \mathbf{f} poses as much, if not more, a challenge as finding an adaptive transform to completely decorrelate \mathbf{f} . This is attested by so far disappointing performance of CS-based compression methods, despite the enthusiasm to change the prevailing practice of “oversampling followed by massive dumping” in image acquisition and compression by CS.

Nevertheless, by shifting the burden of code optimization to the decoder, the CS-based compression approach greatly simplifies the pipeline of signal acquisition and encoding. The encoder simply makes a small number of random projections of the signal and quantizes and transmits the projection values. This asymmetric design is highly desired when the data acquisition devices must be simple and operate on limited power budget (e.g., inexpensive resource-deprived sensors) and when high-density sampling can harm the object being captured (e.g., medical imaging). In such scenarios, the signal reconstruction quality is up to the CS decoder, given the number of CS measurements K .

The poor rate-distortion performance of current CS recovery techniques relative to conventional coding techniques is rooted

Manuscript received September 23, 2009; revised November 28, 2010, May 19, 2011 and July 10, 2011; accepted July 11, 2011. Date of publication August 04, 2011; date of current version January 18, 2012. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and the National Natural Science Foundation of China under Grant 60932006, Grant 61033004, Grant 60736043, and Grant 61070138 and in part by the Fundamental Research Funds of the Central Universities of China under Grant K50510020003. This work is presented in part at the 2009 IEEE Data Compression Conference. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Flynn.

X. Wu and X. Zhang are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: xwu@ece.mcmaster.ca; zhangxj@grads.ece.mcmaster.ca).

W. Dong and G. Shi are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: wsdong@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

Digital Object Identifier 10.1109/TIP.2011.2163520

in a major drawback of the problem formulation (1) for CS recovery. Natural signal \mathbf{f} typically has locally varying statistics, and there exists no space Ψ in which all segments of \mathbf{f} exhibit sparsity. The problem is particularly acute for images. For a nonstationary 2-D $N_r \times N_c$ image signal $f(x, y) \in \mathbb{N}^{N_r \times N_c}$, in two different areas \mathcal{A}_i and \mathcal{A}_j of the spatial domain, subimages $f_i(x, y)$ and $f_j(x, y)$ can have very different waveforms (e.g., smooth shade versus strong edge), and hence, they are sparse in different spaces Ψ_i and Ψ_j . Thus, performing CS recovery in fixed space Ψ , such as that of DCT, a wavelet, or total variation (TV), is inefficient and requires more measurements than if adaptive sparse spaces are used. This critique motivates us to seek for a locally adaptive strategy to recover CS-acquired images.

In this paper, we propose a new framework of model-based adaptive recovery of compressive sensing (MARX) to rectify the flaw in the current CS recovery problem formulation (1). The defining feature of MARX, which distinguishes it from other CS recovery techniques, is a locally adaptive sparse signal representation facilitated by a piecewise autoregressive (PAR) model. The PAR model is defined by

$$\mathbf{f} = \mathbf{A}\mathbf{f} + \mathbf{v}. \quad (2)$$

For image processing applications, \mathbf{f} is the vector by stacking all $N = N_r N_c$ pixels of image $f(x, y) \in \mathbb{N}^{N_r \times M_c}$, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a real-valued square matrix with all elements on the main diagonal being zero. Term $\mathbf{v} \in \mathbb{R}^N$ is a random vector that is the excitation of the 2-D autoregressive process. The n th row vector of \mathbf{A} is denoted by \mathbf{a}_n , $a_{nn} \equiv 0$, $1 \leq n \leq N$. In the view that an image is a random Markov field (RMF) of a modest order, each row vector \mathbf{a}_n is sparse, i.e., only a very small portion of the elements of \mathbf{a}_n are nonzero. The nonzero elements of \mathbf{a}_n constitute the 2-D support of regression relation $f_n = \mathbf{a}_n \mathbf{f} + v$ for pixel f_n . The spatial configuration and the order of the regression support for f_n are given by the image waveform at pixel location n .

Respecting the fact that a natural image is a nonstationary RMF, MARX allows the PAR model parameters \mathbf{a}_n to vary in n . As such, via its parameter matrix \mathbf{A} , the PAR model offers a sparse and yet adaptive representation of image signal \mathbf{f} . Therefore, the CS recovery of \mathbf{f} can be formulated as the following problem of ℓ_1 minimization:

$$\min_{\mathbf{A}, \mathbf{f}} \sum_{1 \leq n \leq N} \|\mathbf{a}_n\|_{\ell_1} \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{A} \mathbf{f}. \quad (3)$$

We stress the contrast between the variability of \mathbf{A} in (3) and the predetermined bases of Ψ in (1). Needless to say, the proposed MARX sparsity mechanism can fit image local structures (e.g., edges, textures, and smooth shades) much better than wavelet, curvelet, DCT or whatever predetermined bases of Ψ .

The generality of the PAR model is ensured by the fact that the Gauss–Markov process of form $X_i = -\sum_{t=1}^k a_t X_{i-t} + n_i$, where n_i 's are independent and identically distributed $\sim \mathcal{N}(0, \sigma^2)$, is the maximum entropy rate stochastic process, if a_t 's are chosen such that $E\{X_i X_{i+t}\} = a_t$, $1 \leq t \leq k$, for all i [7, p. 274]. Thus, the AR model class is versatile and expressive, and the parametric space of \mathbf{A} in (3) is capable of

yielding sparse representation for image waveforms ranging from smooth shades, periodic textures to transients such as edges.

Granted, the proposed MARX objective function is computationally more complex than that in the current CS problem formulation. The former involves joint estimation of image \mathbf{f} and its underlying PAR model \mathbf{A} , rather than estimating \mathbf{f} in fixed space Ψ . The added search space of \mathbf{A} makes the inverse problem of CS recovery severely underdetermined. In the following sections, we will develop algorithm techniques to overcome this difficulty, making the MARX solution feasible and robust. In pursuing maximum confinement of the solution space for (3), we exploit structured sparsities due to fractal behavior (self-similarities) of natural images. We show how the PAR model can be made a convenient machinery to incorporate the structured sparsities into the framework of MARX. The resulting technique not only makes the MARX process computationally tractable but also greatly improves the performance of existing CS recovery algorithms. Extensive experiments on a broad class of natural images, ranging from conventional photographs to biomedical images, establish the superior recovery quality of MARX over other CS methods. The gap in performance can be as much as 2–7 dB, with the advantage of MARX being the most prominent in the recovery of local structural information (fine and subtle image details).

The remainder of this paper is structured as follows: In Section II, we solve the severely underdetermined inverse problem of MARX by imposing constraints stemmed from the Markov property and self-similarities of an image signal. In Section III, we develop a pattern classification technique to characterize and explore structured sparsity in the PAR parameter space and reduce the complexity of MARX. Simulation results are reported and discussed in Section IV. Section V concludes this paper.

II. MARX ALGORITHM BASED ON STRUCTURED SPARSITY

Here, we develop algorithm techniques to solve the MARX problem (3). The power of adaptive PAR model \mathbf{A} lies in its capability of providing, by varying \mathbf{a}_n , different sparse representations for image waveforms in different spatial locations. However, one should fully use statistical knowledge of natural images to structure \mathbf{a}_n 's in ways to confine the solution space of the underdetermined inverse problem (3). Consider the N CS constraints of (3) as

$$\mathbf{y} = \Phi \mathbf{A} \mathbf{f} = \Phi(\mathbf{a}_1 \mathbf{f}, \mathbf{a}_2 \mathbf{f}, \dots, \mathbf{a}_N \mathbf{f})^T \quad (4)$$

each of which involves the PAR prediction of pixel f_n , i.e., $\hat{f}_n = \mathbf{a}_n \mathbf{f}$, $a_{nn} \equiv 0$, $1 \leq n \leq N$. For natural images are RMF, f_n only depends on pixels in a local window, i.e.,

$$\mathbf{w}_n = (f_n, f_{n \odot 1}, f_{n \odot 2}, \dots, f_{n \odot K}). \quad (5)$$

Symbol $n \odot k$ denotes the index of the k th neighbor of pixel f_n in the image domain. The ordering of potentially K effecting pixels $f_{n \odot 1}, f_{n \odot 2}, \dots, f_{n \odot K}$ is given by a fixed 2-D traversal (e.g., raster scan) with respect to pixel location n . We set to zero those coefficients in \mathbf{a}_n that correspond to spatial locations outside of local window \mathbf{w}_n and impose a structure of sparsity

in \mathbf{A} that is intrinsic to the physical problem. Structured sparsity model \mathbf{A} is defined as

$$\mathbf{A} = \{a_{n,j}\}_{N \times N} \text{ s.t. } a_{n,j} = 0 \quad \forall j \in \{1, 2, \dots, N\} \setminus \{n \odot 1, n \odot 2, \dots, n \odot K\}. \quad (6)$$

This means that the signal exhibits not only sparsity in PAR parameter space, i.e., most of the elements of \mathbf{A} are zero, but the zero and nonzero elements are highly structured in the way that, for a given pixel, only parameters associated with its neighbors may be nonzero or significant. This structured sparsity shares the same spirit as [8], but it has different connotations in specifics.

The structure of our sparsity is actually richer than the aforementioned local configuration of nonzero parameters and clustering of zero parameters. A more important and beneficial structured sparsity of \mathbf{A} presents itself if one considers fractal property of natural images. Namely, image \mathbf{f} often exhibits a similar waveform structure, i.e., having similar second-order statistics, in different localities and scales. To characterize the 2-D image waveform and noting that PAR parameters \mathbf{a}_i and \mathbf{a}_j are the same if $\mathbf{w}_i = \mathbf{w}_j + c\mathbf{1}$, we associate each pixel f_n with a mean-removed feature vector, i.e.,

$$\tilde{\mathbf{w}}_n = \mathbf{w}_n - \mu_n \mathbf{1}, \quad \mu_n = \frac{1}{K+1} \mathbf{1}^T \mathbf{w}_n. \quad (7)$$

Observing that f_i and f_j are generated by an underlying PAR model of the same or similar parameters if the two corresponding feature vectors $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{w}}_j$ are close to each other, we can strengthen the MARX formulation of (3) to

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{f}} \quad & \sum_{1 \leq n \leq N} \|\mathbf{a}_n\|_{\ell_1} \\ \text{subject to} \quad & |\mathbf{a}_n \mathbf{f} - f_{n_i}| \leq \kappa \|\tilde{\mathbf{w}}_n - \tilde{\mathbf{w}}_{n_i}\| \\ & 1 \leq n \leq N, \quad n_i \in \{i | \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_n\| < \tau\} \\ & \mathbf{y} = \Phi \mathbf{A} \mathbf{f} \end{aligned} \quad (8)$$

where κ is a normalizing factor, and τ is a threshold to select samples having similar 2-D waveform as $\tilde{\mathbf{w}}_n$ to learn \mathbf{a}_n . This technique of enforcing similar autoregressive functions on similar patches of the image bears some resemblance to nonlocal image modeling [9].

Even with added constraints of structured sparsity, directly solving (8) is still difficult and numerically less stable. We propose an iterative approach of constructing PAR model \mathbf{A} and recovering \mathbf{f} from \mathbf{y} alternately in the MARX framework. Specifically, we estimate the PAR model parameters in \mathbf{A} with respect to an initial estimated image $\mathbf{f}^{(0)}$, i.e., solving the constrained ℓ_1 minimization problem (8) given $\mathbf{f} = \mathbf{f}^{(0)}$. The resulting estimated PAR model $\mathbf{A}^{(0)}$ is then used to improve $\mathbf{f}^{(0)}$ to $\mathbf{f}^{(1)}$, and in turn, $\mathbf{f}^{(1)}$ is used to improve $\mathbf{A}^{(0)}$ to $\mathbf{A}^{(1)}$, and so forth. In iteration t , given $\mathbf{A}^{(t)}$ computed by solving (8), the next estimated image $\mathbf{f}^{(t+1)}$ is obtained by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{f}^{(t+1)}} \quad & \left\{ \sum_{1 \leq n \leq N} \left\| \mathbf{f}^{(t+1)} - \mathbf{A}^{(t)} \mathbf{f}^{(t+1)} \right\|_{\ell_1} \right\} \\ \text{subject to} \quad & \mathbf{y} = \Phi \mathbf{f}^{(t+1)}. \end{aligned} \quad (9)$$

To show that formulation (9) is essentially a CS recovery problem, we let $\mathbf{f}^{(t+1)} = \mathbf{A}^{(t)} \mathbf{f}^{(t)} + \mathbf{x}^{(t+1)}$ and $\mathbf{z}^{(t+1)} = \mathbf{y} - \Phi \mathbf{A}^{(t)} \mathbf{f}^{(t)}$, and rewrite (9) as

$$\begin{aligned} \min_{\mathbf{x}^{(t+1)}} \quad & \left\{ \sum_{1 \leq n \leq N} \left\| \mathbf{x}^{(t+1)} \right\|_{\ell_1} \right\} \\ \text{subject to} \quad & \mathbf{z}^{(t+1)} = \Phi \mathbf{x}^{(t+1)}. \end{aligned} \quad (10)$$

Since $\mathbf{x}^{(t+1)} = \mathbf{f}^{(t+1)} - \mathbf{A}^{(t)} \mathbf{f}^{(t)}$ is a sparse signal in the image domain, (10) is the conventional formulation of the CS problem.

Initial estimated image $\mathbf{f}^{(0)}$ can be obtained by a nonadaptive CS recovery technique, for instance, the TV method [4], i.e.,

$$\min_n \sum_n \left| \nabla f_n^{(0)} \right| \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{f} \quad (11)$$

where ∇ is the 2-D Laplacian operator. Other CS recovery methods can be also used to produce $\mathbf{f}^{(0)}$, such as wavelet-based method, DCT-based method, and matching pursuit. This iterative procedure is also applied to solve simplified versions of the MARX problem in (15) and (18).

The impact of approximation error vector $\mathbf{e}_t = \mathbf{f} - \mathbf{f}^{(t)}$ on the solution of (8), i.e., $t = 0, 1, \dots$, is greatly reduced by imposing CS constraints $\mathbf{y} = \Phi \mathbf{A} \mathbf{f}^{(t)}$ in (8). Since natural image signals are not strictly sparse with many small but nonzero coefficients in the recovery space, error \mathbf{e}_t is caused by setting those small nonzero coefficients to zero [8]. The discarded coefficients represent insignificant high-frequency components of \mathbf{f} , and thus, approximation error \mathbf{e}_t has zero mean. Therefore

$$\begin{aligned} \mathbf{y} &= \Phi \mathbf{A} \mathbf{f}^{(t)} \\ &= \Phi \mathbf{A} \mathbf{f} - \Phi \mathbf{A} \mathbf{e}_t \\ &\approx \Phi \mathbf{A} \mathbf{f} \end{aligned} \quad (12)$$

meaning that the use of $\mathbf{y} = \Phi \mathbf{A} \mathbf{f}^{(t)}$ has almost the same effect as though true \mathbf{f} was known.

Since the accuracy of initial estimate $\mathbf{f}^{(0)}$ increases in the number of CS measurements M , the value of κ in (8) should be adjusted for M and decreases in subsequent iterations as the estimation precision improves. As to the termination criterion for the iterative MARX algorithm, we monitor the successive model fit errors $\delta_t = \|\mathbf{f}^{(t+1)} - \mathbf{A}^{(t)} \mathbf{f}^{(t+1)}\|$, which are the by-product of solving (9). The MARX algorithm terminates when $\delta_{t-1} - \delta_t < \epsilon$, where ϵ is a threshold, or the algorithm can terminate after a given number of iterations. In our experiments, we observe that two to three iterations suffice to produce good results.

In (8), the N sets of PAR parameters (all N row vectors of \mathbf{A}) are jointly estimated. The complexity of the MARX algorithm can be greatly reduced without materially affecting the performance by individually estimating $\mathbf{a}_n^{(t)}$, i.e., $n = 1, 2, \dots, N$, in iteration t of the estimation process. Estimating \mathbf{A} one row vector at a time reduces the number of unknowns by N folds. Denote by

$$\tilde{\mathbf{w}}_n^{(t)} = (f_n^{(t)}, f_{n \odot 1}^{(t)}, \dots, f_{n \odot K}^{(t)}) - \mu_n \mathbf{1} \quad (13)$$



Fig. 1. Sample images. (Top row) Mic. Image1; Mic. Image 2; Girl; MRA-Vessel; MRA-Brain. (Middle row) Monarch; Parrot; Barbara; Bike; Boat. (Bottom row) Hat; House; Leaves; Lena; Peppers.

the mean-removed feature vector of the t th estimated pixel $f_n^{(t)}$, and let

$$\mathbf{A}_n^{(t)} = \begin{pmatrix} \mathbf{a}_1^{(t-1)} \\ \mathbf{a}_2^{(t-1)} \\ \vdots \\ \mathbf{a}_n^{(t)} \\ \vdots \\ \mathbf{a}_N^{(t-1)} \end{pmatrix}. \quad (14)$$

Then, the original sparse estimation problem of (8) can be broken into N ℓ_1 minimization problems of much smaller size and approximately solve

$$\begin{aligned} & \min_{\mathbf{a}_n^{(t)}} \left\| \mathbf{a}_n^{(t)} \right\|_{\ell_1} \\ & \text{subject to } \left| \mathbf{a}_n^{(t)} \mathbf{f} - f_i^{(t)} \right| \leq \kappa \left\| \tilde{\mathbf{w}}_n^{(t)} - \tilde{\mathbf{w}}_i^{(t)} \right\|, \\ & \quad i \in \left\{ i : \left\| \tilde{\mathbf{w}}_i^{(t)} - \tilde{\mathbf{w}}_n^{(t)} \right\| < \tau \right\} \\ & \mathbf{y} = \Phi \mathbf{A}_n^{(t)} \mathbf{f}^{(t)} \\ & n = 1, 2, \dots, N. \end{aligned} \quad (15)$$

III. MARX WITH PATTERN CLASSIFICATION

In the previous section, the MARX algorithm employs N distinct PAR models, one for each pixel f_n , $1 \leq n \leq N$. The intent is to allow the maximum degree of freedom in modeling. However, for natural images, the second-order statistics may spatially change, but the change is smooth. Furthermore, the change may be periodic so that a waveform can repeat itself in different locations. Therefore, we can induce an even stronger structure in sparse matrix \mathbf{A} by grouping the N pixels into $C \ll N$ classes of waveforms. The classifier $Q: \mathbb{R}^{K+1} \rightarrow \{1, 2, \dots, C\}$ is designed by a C -clustering of the N feature vectors $\tilde{\mathbf{w}}_n$, $1 \leq n \leq$

N . Each of the resulting C representative waveforms is fit to a PAR model, and as such, image \mathbf{f} can be described in the MARX framework by C PAR models, one per classified waveform.

For our purpose of classification, the difference metric between two mean-removed feature vectors $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{w}}_j$ is defined as

$$D_{i,j} = \sqrt{\left\| \tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_j \right\|_2^2 + \eta d^2(i,j)} \quad (16)$$

where $\left\| \tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_j \right\|_2^2$ is the squared Euclidean distance between two feature vectors $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{w}}_j$, and $d^2(i,j)$ is the squared geometric distance between pixels i and j . Weight η is used to adjust the relative importance of the waveform difference and geometric distance. The inclusion of $d(i,j)$ in $D_{i,j}$ is to improve the performance of MARX by incorporating the prior knowledge that natural images tend to be piecewise stationary, and the second-order statistics only smoothly change. In particular, this technique translates spatial coherence of edge pixels (e.g., orientation consistency) into structured sparsity in \mathbf{A} .

Let $q_n = Q(\tilde{\mathbf{w}}_n) \in \{1, 2, \dots, C\}$ be the index of the feature class that $\tilde{\mathbf{w}}_n$ falls into and $\hat{\mathbf{a}}_c$ be the model parameter vector for class c . To keep the compact notation for PAR model $\mathbf{f} = \mathbf{A}\mathbf{f} + \mathbf{v}$, we introduce permutation π_n on the elements of $\hat{\mathbf{a}}_c$ such that $f_n = \pi_n(\hat{\mathbf{a}}_c)\mathbf{f} + v$. Now, the coefficient matrix of the C PAR models becomes

$$\hat{\mathbf{A}} = \begin{pmatrix} \pi_1(\hat{\mathbf{a}}_{q_1}) \\ \pi_2(\hat{\mathbf{a}}_{q_2}) \\ \vdots \\ \pi_n(\hat{\mathbf{a}}_{q_n}) \\ \vdots \\ \pi_N(\hat{\mathbf{a}}_{q_N}) \end{pmatrix} \quad (17)$$

in which there are only $C \ll N$ distinct row vectors up to the permutation as the same model $\hat{\mathbf{a}}_c$ acts on all pixels $n \in$

TABLE I
PSNR (IN DECIBELS) RESULTS FOR DIFFERENT CS RECOVERY METHODS

Image	Method	Number of measurements				
		$M = 0.1N$	$M = 0.2N$	$M = 0.3N$	$M = 0.4N$	$M = 0.75N$
Micrograph image 1	TV	20.07	22.72	24.78	26.65	34.28
	[10]	20.00	22.63	24.64	26.55	33.98
	[11]	19.83	22.43	24.55	26.41	34.01
	MARX	21.18	24.72	27.45	29.63	36.62
	MARX-PC	22.00	25.86	28.53	30.70	37.30
Micrograph image 2	TV	19.10	21.25	23.19	25.08	32.81
	[10]	19.05	21.09	23.09	24.99	32.82
	[11]	18.87	21.02	23.04	25.17	33.31
	MARX	19.72	22.84	24.99	27.00	34.56
	MARX-PC	20.48	24.33	26.73	28.76	35.98
Girl	TV	29.60	31.28	32.57	33.75	39.64
	[10]	29.25	31.13	32.44	33.62	39.55
	[11]	29.20	31.05	32.31	33.52	39.58
	MARX	29.50	31.10	32.31	33.45	38.77
	MARX-PC	30.13	31.70	32.84	33.92	39.06
MAR-Vessels	TV	21.92	26.06	29.79	33.19	45.19
	[10]	21.84	25.84	29.23	32.59	44.87
	[11]	21.76	26.08	29.88	33.72	45.84
	MARX	24.63	30.84	34.67	37.55	46.85
	MARX-PC	26.73	32.98	36.29	39.23	47.25
MRA-Brain	TV	20.05	23.39	26.25	29.08	41.94
	[10]	19.98	23.43	26.28	28.97	42.23
	[11]	19.99	23.51	26.48	29.49	44.02
	MARX	20.89	25.08	28.09	30.69	41.47
	MARX-PC	22.28	26.44	29.41	31.86	41.98
Monarch	TV	24.21	28.78	31.93	34.84	43.45
	[10]	24.08	28.55	31.93	34.69	43.53
	[11]	24.31	29.10	32.43	35.32	43.88
	MARX	24.70	29.18	32.15	34.88	43.63
	MARX-PC	27.01	31.17	34.03	36.53	44.51
Parrot	TV	27.65	31.84	34.76	37.00	44.95
	[10]	27.93	32.00	34.80	37.12	44.95
	[11]	28.23	32.32	35.27	37.54	45.17
	MARX	29.00	32.88	35.46	37.68	44.96
	MARX-PC	30.82	34.22	36.51	38.49	45.19
Barbara	TV	22.79	24.78	26.72	28.87	38.22
	[10]	22.64	24.51	26.48	28.79	38.06
	[11]	22.58	24.70	26.81	29.25	39.08
	MARX	22.66	27.06	30.81	33.63	40.85
	MARX-PC	24.11	30.24	33.76	36.00	42.45
Bike	TV	20.00	22.78	25.12	27.27	36.32
	[10]	20.10	22.58	24.96	27.11	36.28
	[11]	19.90	22.76	25.23	27.55	37.09
	MARX	20.47	24.55	27.73	30.45	41.02
	MARX-PC	21.67	25.84	29.10	31.93	41.57
Boats	TV	25.58	29.01	31.52	33.77	41.51
	[10]	25.46	28.80	31.35	33.61	41.71
	[11]	25.51	28.89	31.71	34.21	42.37
	MARX	25.77	29.59	32.25	34.61	42.38
	MARX-PC	27.88	32.12	34.75	37.07	43.93
Hat	TV	25.97	29.59	32.30	34.68	43.23
	[10]	25.87	29.48	32.24	34.60	43.32
	[11]	25.58	29.31	31.91	34.34	43.05
	MARX	27.94	32.60	35.80	38.39	46.52
	MARX-PC	29.06	33.64	36.67	39.08	46.91
House	TV	30.50	33.63	35.54	37.20	44.21
	[10]	30.40	33.40	35.46	37.16	44.25
	[11]	30.83	33.69	35.66	37.18	44.39
	MARX	29.88	33.35	35.04	36.74	43.57
	MARX-PC	32.83	35.27	36.87	38.65	45.24
Leaves	TV	19.28	23.51	27.22	30.85	43.37
	[10]	18.91	23.54	27.37	30.41	43.53
	[11]	19.30	23.78	27.82	31.33	44.85
	MARX	20.99	26.61	30.71	34.39	44.59
	MARX-PC	24.02	30.60	34.24	37.20	46.12
Lena	TV	26.48	29.63	32.32	34.74	43.54
	[10]	26.30	29.64	32.36	34.69	43.66
	[11]	26.41	29.81	32.64	35.03	44.06
	MARX	27.23	31.27	34.13	36.72	44.81
	MARX-PC	29.20	33.17	36.00	38.16	45.43
Peppers	TV	30.73	33.34	34.83	35.99	40.54
	[10]	30.50	33.17	34.74	35.93	40.51
	[11]	30.69	33.32	34.77	35.96	40.42
	MARX	29.66	32.33	33.88	35.08	39.92
	MARX-PC	31.84	33.97	35.13	36.18	40.90
Average	TV	24.26	27.44	29.92	32.20	40.89
	[10]	24.15	27.32	29.82	32.05	40.88
	[11]	24.20	27.45	30.04	32.40	41.41
	MARX	24.95	28.93	31.70	34.06	42.03
	MARX-PC	26.67	30.77	33.39	35.58	42.92

$\{i|Q(\tilde{\mathbf{w}}_i) = c\}$. Feature classifier Q imposes strong signal-dependent structures in sparse model parameter matrix \mathbf{A} and reduces the MARX objective function to

$$\begin{aligned} & \min_{\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_C, \mathbf{f}} \sum_{1 \leq c \leq C} \|\hat{\mathbf{a}}_c\|_{\ell_1} \\ & \text{subject to } |\pi_n(\hat{\mathbf{a}}_{q_n})\mathbf{f} - f_{n_i}| \leq \kappa \|\tilde{\mathbf{w}}_n - \tilde{\mathbf{w}}_{n_i}\| \\ & \quad 1 \leq n \leq N, n_i \in \{i | \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_n\| < \tau\} \\ & \quad \mathbf{y} = \Phi \hat{\mathbf{A}} \mathbf{f}. \end{aligned} \quad (18)$$

The iterative MARX algorithm developed in the previous section can be straightforwardly extended to solving the estimation problem of (18). We only need to replace $\mathbf{A}^{(t)}$ by $\hat{\mathbf{A}}^{(t)}$, which is the parameter array of the PAR models corresponding to the C feature classes that are updated in iteration t .

Similar to the development in (14) and (19), we can estimate $\hat{\mathbf{a}}_c$ one at a time for $c = 1, 2, \dots, C$, reducing the size of the optimization problem by C folds. In iteration t of the estimation process, each $\hat{\mathbf{a}}_c$ is computed as follows:

$$\begin{aligned} & \min \|\hat{\mathbf{a}}_c^{(t)}\|_{\ell_1} \\ & \text{subject to } |\pi_i(\hat{\mathbf{a}}_c^{(t)})\mathbf{f} - f_i^{(t)}| \leq \kappa \|\mathbf{m}_c^{(t)} - \tilde{\mathbf{w}}_i^{(t)}\|, \\ & \quad i \in \{i | Q(\tilde{\mathbf{w}}_i^{(t)}) = c\} \\ & \quad \mathbf{y} = \Phi \hat{\mathbf{A}}_c^{(t)} \mathbf{f}^{(t)} \\ & \quad c = 1, 2, \dots, C \end{aligned} \quad (19)$$

where $\mathbf{m}_c = E\{\tilde{\mathbf{w}}_i | Q(\tilde{\mathbf{w}}_i) = c\}$ is the centroid of all mean-removed feature vectors that are quantized to class c .

IV. EMPIRICAL RESULTS AND REMARKS

Here, we report the experimental results of the proposed MARX technique and discuss our findings. In our implementation, the parameters are set as follows: The order of PAR model K is set to 9 (3×3 window); the number of CS measurements M is given as the percentage of the total number of pixels; instead of specifying a value of τ in (8) and (18), we use the J^{th} ($J = 50$) closest matched patches; η in (16) is empirically set to 0.1; the algorithm termination threshold ϵ is set to 4. In fact, we found that two iterations suffice to generate good recovery results.

To highlight the importance of spatial adaptability of a CS recovery method and verify the efficacy of MARX in this regard, we conduct a comparative study between MARX and the conventional CS method in a “sparse” space of choice. A fairly general set of test images was used in our comparative study, including photographic images commonly found in the literature (e.g., Lena, House, Barb, and Monarch) and many biomedical images. Fifteen of these test images are shown in Fig. 1. To evaluate the benefits of the pattern classification, we report the results of the two proposed methods, i.e., the MARX method without pattern classification (denoted by MARX) and the MARX with pattern classification (denoted by MARX-PC).

Table I lists the peak signal-to-noise ratio (PSNR) values versus the number of CS measurements (presented as the percentage of the total number of pixels N) for different CS recovery methods. Our experiments are set up to benefit the

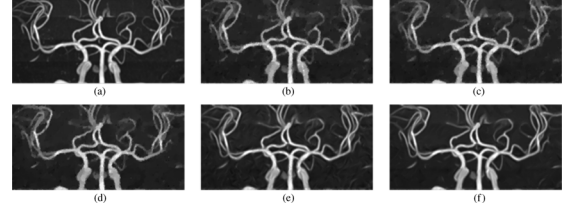


Fig. 2. CS recovered MRA-Vessels images (part) with $M = 0.2 N$ measurements. (a) Original; (b) TV recovery (26.06 dB); (c) [10] recovery (25.84 dB); (d) [11] recovery (26.08 dB); (e) MARX recovery (30.84 dB); (f) MARX-PC recovery (32.98 dB).

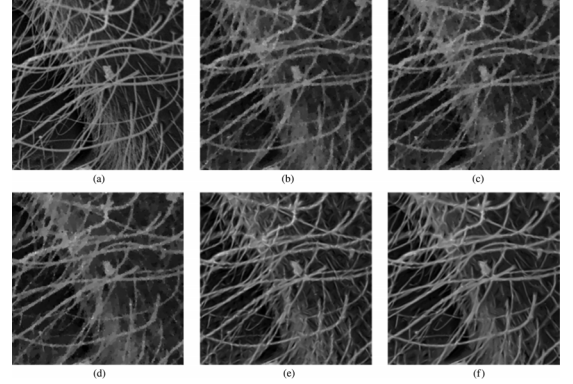


Fig. 3. CS recovered micrograph 1 images (part) with $M = 0.2 N$ measurements. (a) Original; (b) TV recovery (22.72 dB); (c) [10] recovery (22.63 dB); (d) [11] recovery (22.43 dB); (e) MARX recovery (24.72 dB); (f) MARX-PC recovery (25.86 dB).

competing methods of MARX. For each test image, conventional CS recovery is performed in wavelet, DCT, and gradient space separately, and the space of best reconstruction quality is selected. It turns out that for all test images but one, the TV method is better than the CS recovery in either wavelet or DCT space. Only for test image Barb, the CS reconstruction in DCT space is better than the TV reconstruction. To demonstrate the advantage of the proposed method, some other recently proposed adaptive CS recovery methods [10], [11] are also included in the comparison group.

As shown in Table I, the MARX algorithm outperforms other competing methods on most of test images and the numbers of CS measurements. With pattern classification, the proposed MARX algorithm consistently outperforms other methods on all test images and over different numbers of CS measurements and further increases the performance advantage of the proposed MARX algorithm over other methods. The average PSNR gain over [11] that is ranked number 3 in the comparison group can be up to 3.35 dB (at $M = 0.3 N$ measurements). In most cases, the MARX-PC algorithm can obtain comparable and even higher PSNR than the other methods using 0.1 N fewer measurements. By relating the PSNR gains of the MARX algorithms to the types of images shown in Fig. 1, we see that the MARX algorithm enjoys greater advantage over competing methods for images of rich anisotropic spatial structures (mostly in the form of edges). This observation manifests the capability of MARX to adapt to spatially varying second-order statistics of the image signal and the great benefit of utilizing structured sparsity in CS recovery.

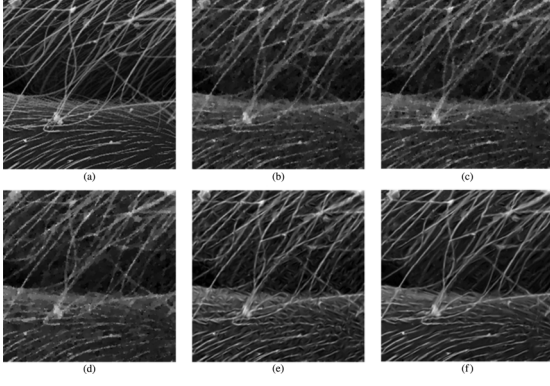


Fig. 4. CS recovered micrograph 2 images (part) with $M = 0.2 N$ measurements. (a) Original; (b) TV recovery (21.25 dB); (c) [10] recovery (21.09 dB); (d) [11] recovery (21.02 dB); (e) MARX recovery (22.84 dB); (f) MARX-PC recovery (24.33 dB).



Fig. 5. CS recovered Barbara images (part) with $M = 0.2 N$ measurements. (a) TV recovery (24.78 dB); (b) [10] recovery (24.51 dB); (c) [11] recovery (24.70 dB); (d) MARX-PC recovery (30.24 dB).

Perhaps more importantly than higher PSNR, the MARX algorithm also offers far superior visual quality. To verify this, we present and invite the reader to compare, in Figs. 2–8, the recovered CS-sampled images by competing algorithms for various numbers of CS measurements. In many cases, the MARX algorithm can reproduce sharper and cleaner images than the other methods while using only two thirds as many CS measurements. This should be evident by comparing the MARX-PC result of $M = 0.2 N$ with the results of other competing methods of $M = 0.3 N$ [i.e., Fig. 5(d) versus Fig. 6(a)–(c) and Fig. 7(d) versus Fig. 8(a)–(c)]. Such large differences in visual quality are due to the fact that the human visual system is highly sensitive to distortions of spatial coherence of edges, and this semantically vital structural information is more faithfully reproduced by MARX.

Finally, we examine the computational complexity of the MARX method. In MARX, computations are performed in three stages: 1) the initial TV-based estimation; 2) the computation of the PAR model; and 3) the image recovery using



Fig. 6. CS recovered Barbara images (part) with $M = 0.3 N$ measurements. (a) TV recovery (26.72 dB); (b) [10] recovery (26.48 dB); (c) [11] recovery (26.81 dB); (d) MARX-PC recovery (33.76 dB).

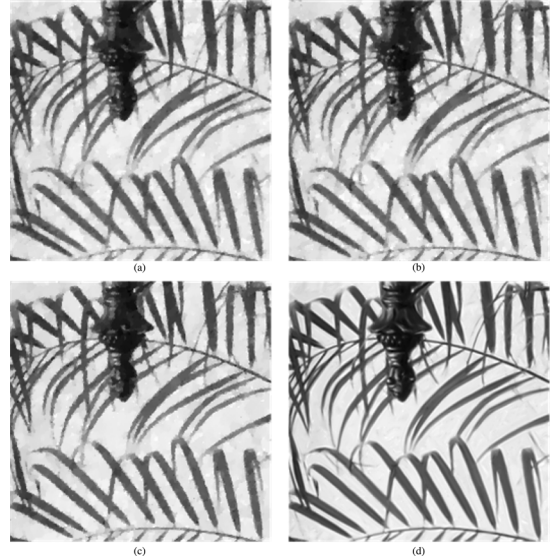


Fig. 7. CS recovered leaf images (part) with $M = 0.2 N$ measurements. (a) TV recovery (23.51 dB); (b) [10] recovery (23.54 dB); (c) [11] recovery (23.78 dB); (d) MARX-PC recovery (30.60 dB).

the learned PAR model. We apply the basis pursuit algorithm [12] to obtain an initial estimate of the image in $O(N \log N)$ operations. Since the determination of the PAR parameters is posed as a set of $N \ell_1$ minimization problems (14), the total computational cost for computing the PAR model is $O(NS \log S)$, where S is the size of the sample set used to estimate the PAR parameters. With the estimated PAR model \mathbf{A} , the reconstruction of the image requires $O(N \log N)$ operations. Therefore, the total complexity of the MARX method is $O(N(\log N + S \log S))$.

V. CONCLUSION

We have argued for the importance of adapting to local 2-D waveforms in the recovery of CS-acquired images and departed

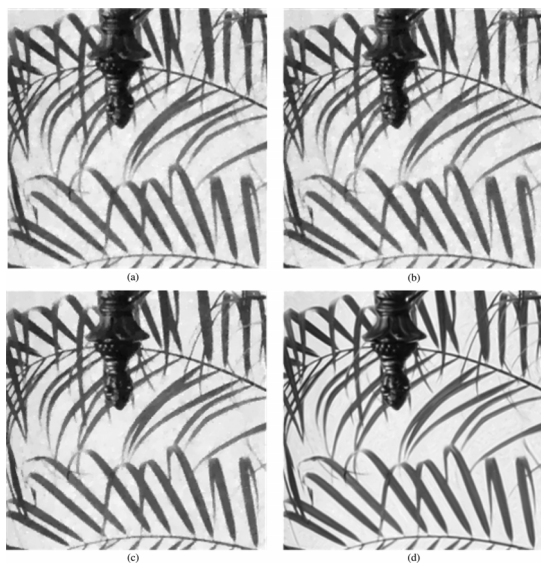


Fig. 8. CS recovered leaf images (part) with $M = 0.3 N$ measurements. (a) TV recovery (27.22 dB); (b) [10] recovery (27.37 dB); (c) [11] recovery (27.82 dB); (d) MARX-PC recovery (34.24 dB).

from the current practice of recovering a CS-acquired signal in a fixed “sparse” space. Instead, we have proposed an adaptive CS recovery strategy MARX that fits the recovered signal to a PAR model. By adjusting the model parameters, the model-based CS recovery can track spatially varying second-order statistics and thus outperform existing CS recovery methods by 2–7 dB. In addition to obtaining higher PSNR, the spatially adaptive recovery of MARX conveys signal semantics much better than other CS recovery methods.

REFERENCES

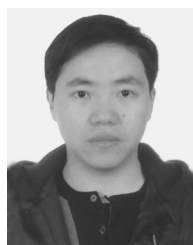
- [1] E. Candès, “Compressive sampling,” in *Proc. Int. Congr. Math.*, Madrid, Spain, 2006, pp. 1433–1452.
- [2] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [4] M. Figueiredo, R. Nowak, and S. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 568–597, Dec. 2007.
- [5] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [6] I. Daubechies, M. D. Friese, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [9] A. Buades, B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one,” *Multisc. Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [10] R. Chartrand and W. Yin, “Iteratively reweighted algorithms for compressive sensing,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2008, pp. 3869–3872.
- [11] E. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [12] S. Chen, D. Donoho, and M. Saunders, “Atomic decompositions by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.



Xiaolin Wu (SM’96–F’10) received the B.Sc. degree from Wuhan University, Wuhan, China, in 1982 and the Ph.D. degree from the University of Calgary, Calgary, AB, Canada, in 1988, both in computer science.

He started his academic career in 1988 and has since been with the faculty at the University of Western Ontario, London, ON, Canada; New York Polytechnic University, Brooklyn; and, currently, McMaster University, Hamilton, ON, Canada, where he is a Professor with the Department of Electrical and Computer Engineering, and where he holds the NSERC-DALSA Industrial Research Chair in Digital Cinema. He has published over 180 research papers and is the holder of two patents. His research interests include image processing, multimedia compression, joint source-channel coding, multiple description coding, and network-aware visual communication.

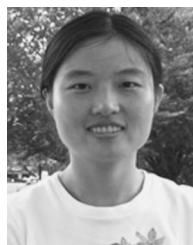
Dr. Wu currently serves as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING.



Weisheng Dong received the B.S. degree in electronic engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004 and the Ph.D. degree in circuits and system from Xidian University, Xi’an, China, in 2010.

From September to December 2006, he was a Visiting Student at Microsoft Research Asia, Beijing, China. From January 2009 to June 2010, he was a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong. Since September 2010, he has been a Lecturer with Xidian University. His research interests include inverse problems in image processing, sparse signal representation, and image compression.

Dr. Dong was the recipient of the Best Paper Award at The International Society for Optical Engineers Visual Communication and Image Processing 2010.



Xiangjun Zhang received the B.Sc. degree from Xi’an Jiaotong University, Xi’an, China, in 1997, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree from McMaster University, Hamilton, ON, Canada in 2009.

She is currently with the Department of Electrical and Computer Engineering, McMaster University. Her research interests include image/video processing, multimedia communication, and signal compression.



Guangming Shi (SM’10) received the B.S. degree in automatic control in 1985, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi’an, China, in 1985, 1988, and 2002, respectively.

He joined the School of Electronic Engineering, Xidian University, in 1988. From 1994 to 1996, as a Research Assistant, he cooperated with the Department of Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, and in 2004, he was the Head of the National Instruction Base of Electrician and Electronic. From June to December in 2004, he studied in the Department of Electronic Engineering, University of Illinois at Urbana-Champaign, Urbana. He is currently the Deputy Director of the School of Electronic Engineering, Xidian University, and the Academic Leader in the subject of circuits and systems. He has authored or coauthored over 60 research papers. His research interests include compressed sensing, theory and design of multirate filter banks, image denoising, low-bit-rate image/video coding and implementation of algorithms for intelligent signal processing (using digital signal processor and field-programmable gate array).