Nonconvex Compressed Sensing by Nature-Inspired Optimization Algorithms

Fang Liu, *Senior Member, IEEE,* Leping Lin, Licheng Jiao, *Senior Member, IEEE,* Lingling Li, Shuyuan Yang, *Member, IEEE,* Biao Hou, *Member, IEEE,* Hongmei Ma, Li Yang, and Jinghuan Xu

***Abstract*—The l_0 regularized problem in compressed sensing reconstruction is nonconvex with NP-hard computational complexity. Methods available for such problems fall into one of two types: greedy pursuit methods and thresholding methods, which are characterized by suboptimal fast search strategies. Nature-inspired algorithms for combinatorial optimization are famous for their efficient global search strategies and superior performance for nonconvex and nonlinear problems. In this paper, we study and propose nonconvex compressed sensing for natural images by nature-inspired optimization algorithms. We get measurements by the block-based compressed sampling and introduce an overcomplete dictionary of Ridgelet for image blocks. An atom of this dictionary is identified by the parameters of direction, scale and shift. Of them, direction parameter is important for adapting to directional regularity. So we propose a two-stage reconstruction scheme (TS_RS) of nature-inspired optimization algorithms. In the first reconstruction stage, we design a genetic algorithm for a class of image blocks to acquire the estimation of atomic combinations in all directions; and in the second reconstruction stage, we adopt clonal selection algorithm to search better atomic combinations in the sub-dictionary resulted by the first stage for each image block further on scale and shift parameters. In TS_RS, to reduce the uncertainty and instability of the reconstruction problems, we adopt novel and flexible heuristic searching strategies, which include delicately designing the initialization, operators, evaluating methods, and so on. The experimental results show the efficiency and stability of the proposed TS_RS of nature-inspired algorithms, which outperforms classic greedy and thresholding methods.**

***Index Terms*—Clonal selection algorithm (CSA), genetic algorithm (GA), nature-inspired optimization algorithm, nonconvex compressed sensing, overcomplete dictionary, structured sparsity.**

Manuscript received February 28, 2014; revised June 6, 2014; accepted July 21, 2014. Date of publication August 19, 2014; date of current version April 13, 2015. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329402, in part by the National Natural Science Foundation of China under Grant 61173090, Grant 61173092, Grant 61303119, Grant 61271302, and Grant 61272282, in part by the National Research Foundation for the Doctoral Program of Higher Education of China under Grant 20110203110006, in part by the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) under Grant B07048, in part by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT1170, in part by the Fundamental Research Funds for the Central Universities under Grant JB140317 and in part by the Guangxi Natural Science Foundation under Grant 2013GXNSFDA019030. This paper was recommended by Associate Editor R. Lynch.

F. Liu, L. Lin, H. Ma, L. Yang, and J. Xu are with the School of Computer Science and Technology and also with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: f63liu@163.com; lin_leping@163.com).

L. Jiao, L. Li, S. Yang, and B. Hou are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: lchjiao@mail.xidian.edu.cn; linglingxidian@gmail.com; syyang@xidian.edu.cn; avcodec@163.com).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes a more detailed introduction and a sample of the running results of the programs presented in the paper. This material is 1.37 MB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2014.2343618

I. Introduction

COMPRESSED sensing (CS) is a new developed theoretic framework for information representation, acquisition, reconstruction and processing [1]–[4]. It not only inspires us to survey the linear problem again, but also enriches the optimization approaches for signal reconstruction to promote the combination of mathematics with engineering applications.

The CS problem of reconstructing a sparse signal $\mathbf{x} \in \mathbb{R}^n$ from its compressed measurement $\mathbf{y} \in \mathbb{R}^m$ ($m \ll n$) could be modeled by

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0, \ \ s.t. \ \mathbf{y} = \Phi\mathbf{x} \tag{1}$$

where the l_0 norm $\|\cdot\|_0$ counts the number of nonzero components of a vector. In real world, however, many signals are not sparse themselves but sparse in a transform domain

$$\mathbf{x} = \mathfrak{D}\mathfrak{s}$$

where $\mathfrak{D} \in \mathbb{R}^{n\times N}$ is the sparsifying dictionary and $\mathfrak{s} \in \mathbb{R}^N$ a sparse vector. The problem becomes to estimate the sparse coefficient vector $\mathfrak{s}$ by solving

$$\mathfrak{s}^* = \arg\min_{\mathfrak{s}} \|\mathfrak{s}\|_0, \ \ s.t. \ \mathbf{y} = \Phi\mathfrak{D}\mathfrak{s}. \tag{2}$$

Then $\mathbf{x}$ is estimated by $\mathfrak{D}\mathfrak{s}^*$. Taking into account the measurement noise and approximation error, the constrain in the model is relaxed to

$$\mathfrak{s}^* = \arg\min_{\mathfrak{s}} \|\mathfrak{s}\|_0, \ \ s.t. \ \|\mathbf{y} - \Phi\mathfrak{D}\mathfrak{s}\|^2 \le \varepsilon. \tag{3}$$

In the early theoretic research of CS [1]–[3], $\mathfrak{D}$ is an unit or orthonormal matrix, i.e., the signal $\mathbf{x}$ is sparse itself or sparse in an orthogonal basis. They are rather strong restrictions. Overcomplete dictionaries have found wide practical applications, and offered more adaptive and flexible representations. Mallat and Zhang [5] originally proposed a large dictionary composed of time-frequency atoms for adaptively decomposing signal structures. Olshausen and

Field [6], [7] argued that sparse coding with an overcomplete dictionary is the strategy employed by the visual receptive field of mammals. Nowadays, the classic theories of CS have been generalized to the scenario of employing overcomplete dictionaries [8]–[10].

In the model of (3), given the number of nonzero-valued components (also known as sparsity) in the sparse signal $\mathfrak{s}$ is K, the reconstruction is also given by

$$\mathfrak{s}^* = \arg\min_{\mathfrak{s}} \|\mathbf{y} - \Phi\mathfrak{D}\mathfrak{s}\|^2, \ s.t. \ \|\mathfrak{s}\|_0 \le K. \tag{4}$$

Rewrite the sparse representation formula as in [11]

$$\mathbf{x} = \mathfrak{D}\mathfrak{s} = \sum_{i\in\Lambda} \mathfrak{s}_i \mathbf{d}_i = \mathbf{D}\mathbf{s}. \tag{5}$$

By substituting (5) to (4), the model is cast as in [11]

$$(\mathbf{D}^*, \mathbf{s}^*) = \arg\min_{\mathbf{D},\mathbf{s}} \|\mathbf{y} - \Phi\mathbf{D}\mathbf{s}\|^2, \ s.t. \ \|\mathbf{D}^T\|_{p,0} \le K \tag{6}$$

where $\|\mathbf{D}^T\|_{p,0}$ denotes the number of nonzero distinct columns of $\mathbf{D}$. For the matrix norm $l_{p,q}$ and $l_{p,0}$, see [4], [11].

Though there are two variables in the problem, $\mathbf{s}$ could be calculated, when $\mathbf{D}$ is given, by the least squares formula

$$\mathbf{s} = (\Phi\mathbf{D})^+\mathbf{y} \tag{7}$$

where $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudo-inverse of a matrix. By (6) and (7), the reconstruction is modeled to be a combinatorial optimization problem. This is one of the synthesis CS models [12]. The task is nonconvex, multimodal, and NP-hard.

The methods available for (6) fall into one of the two types: greedy pursuit methods and thresholding methods. All of them adopt suboptimal fast search strategies. They iteratively estimate the support of a signal and the corresponding coefficients, or equivalently, iteratively evaluate $\mathbf{D}$ and $\mathbf{s}$. Orthogonal matching pursuit (OMP) [13], [14] is a classic greedy pursuit method. It identifies an atom of $\mathbf{D}$ at each iteration according to the current value of $\|\mathbf{y} - \Phi\mathbf{D}\mathbf{s}\|^2$, and then updates $\mathbf{s}$ followed by reevaluating the residual. Iterative hard thresholding (IHT) [15] is a thresholding method. It obtains an improved estimation of $\mathbf{D}$ at each iteration by a gradient descent step followed by hard thresholding. Both OMP and IHT estimate $\mathbf{D}$ at each iteration according to the current value of $\|\mathbf{y} - \Phi\mathbf{D}\mathbf{s}\|^2$ but not directly to $\mathbf{y}$. It is desired to establish a reconstruction method adopting global search strategies and simultaneously optimizing $\mathbf{D}$ and $\mathbf{s}$.

Nature-inspired algorithms [16] for combinatorial optimization are famous for their global search strategies and superior performance. It is quite natural to consider employing them for the nonconvex CS. Nature-inspired algorithms have been trying to model and borrow the mechanisms lying in intelligent systems to solve complex engineering problems [17]–[19]. Genetic algorithm (GA) [20] and clonal selection algorithm (CSA) [21], [22] are two of the most well-known nature-inspired algorithms. In our work [23], GA was introduced to solve a l_p $(0 < p < 1)$ regularized CS reconstruction subproblem. In [24], an evolutionary multiobjective approach was established for sparse reconstruction. In this paper, it is the first time that natural-inspired algorithms are employed for the l_0 regularized nonconvex CS reconstruction. It makes the simultaneous optimization of the values and nonzero indices of sparse coefficients possible.

We proposed a novel TS_RS of nature-inspired optimization algorithms for the nonconvex CS of natural images. In the CS framework, the measurements are acquired by the block-wise compressed sampling. Then the sparsity prior is introduced by the overcomplete dictionary of Ridgelet [11]. Also based on the dictionary, the structured priors beyond classic sparsity are exploited to reduce the instability and flexibility of the reconstruction. Each atom of the dictionary is indexed by several parameters, of which direction parameter is important for capturing directional regularities. So the proposed scheme considers the atomic combination in direction at first, which is followed by further refinement on other parameters. In the first stage of TS_RS, local, and nonlocal similar blocks are supposed to be well represented by a common group of atoms. GA is employed to determine the common atomic combination in direction for each block class. In the second stage, the assumption that adjacent blocks share similar structures is cast. CSA is employed for each block to refine the results. In each stage, the reconstruction is implemented by delicately designing the components of the algorithms, e.g., initialization, operators, evaluating functions, and so on. They are quite different from that employed by l_1 regularized CS methods, which relax l_0 norm to l_1 norm, so as the obtained convex problems are solved by the available efficient l_1-based optimization methods (see [3]). Methods of this kind incorporate priors by regularizing the problem with convex items (see [25]). In this paradigm, some nonlinear and nonconvex priors are hard to be approximated or expressed by convex items, and the regularity coefficients are hard to be determined.

The rest of the paper is organized as follows. In Section II, the nonconvex CS framework and the proposed TS_RS are introduced. Section III presents the reconstruction models based on structured sparsity and the overcomplete dictionary of Ridgelet. In Section IV, the implemented details of TS_RS of nature-inspired algorithms is presented. In Section V, the experimental results and analysis are presented. The last section concludes the paper.

II. Nonconvex CS by Nature-Inspired Optimization Algorithms

A. Block Compressed Sampling

The block-based compressed sampling [26] is adopted. An image is divided into equal-sized and nonoverlapped blocks. Each block is sampled by the same measure operator.

This block strategy is advantageous for fast sampling and processing. Besides, according to the self-similarity property of images, an image block could easily find its structures in other blocks of the same image. For example, blocks of an image could be easily clustered into several classes. Blocks in the same class share similar structures, and could be represented by a common group of atoms. By incorporating such structured sparsity priors, the flexibility of the problems could be efficiently reduced as it is shown in later sections.

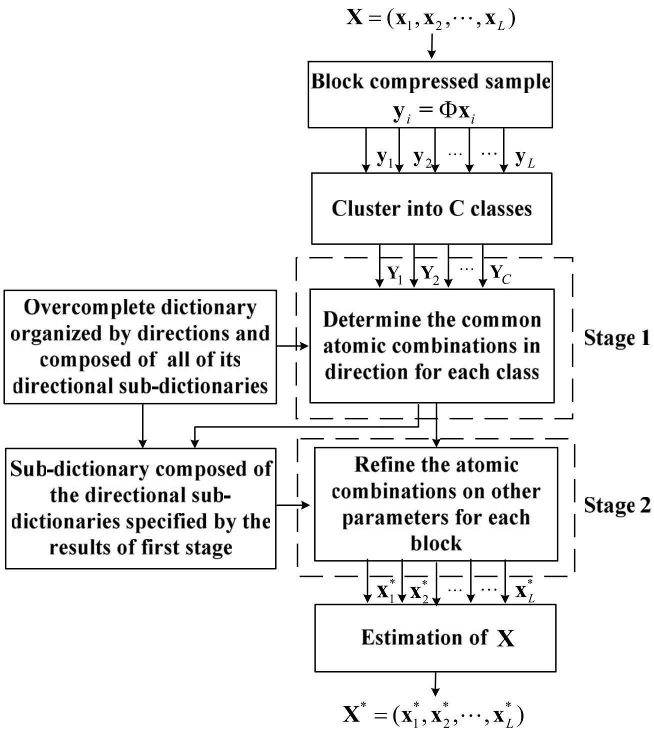


Fig. 1. Block diagram of the nonconvex CS framework of TS_CS.

B. Overcomplete Dictionary

The essential issues of CS reconstruction includes the following.

1) How to construct a sparsifying dictionary for the signals? The dictionary is required to adapt to arbitrary directions of localized edges and textures in natural images.
2) How to faithfully reconstruct the signal with the dictionary meeting above requirements? The algorithm must acquire a good solution within reasonable time, i.e., an atomic combination and the corresponding coefficients.

Sparsity is the primary prior of CS. In the block-based framework, an overcomplete dictionary for image blocks is constructed to introduce sparsity. The dictionary shares some translation invariance properties, which means the translated versions of an atom are still the members of the dictionary. The linear operators on an image block, such as scaling, shifting, and rotating, could be tightly correlated to the operators on the atoms used for representing the block.

All the atoms of the dictionary with a common direction are assembled into a directional sub-dictionary. It is desired to capture the local directional regularity of an image, since human eyes are sensitive to line-like regularities. The atomic combinations for image blocks in direction are considered at first, and followed by further refinement on other parameters.

C. Two-Stage Reconstruction Scheme (TS_RS)

Fig. 1 is the diagram of the CS framework of TS_RS. The reconstruction scheme is composed of two successive stages. In the first stage, after the L image blocks are clustered into C classes ($C \ll L$), the common atomic combination for each class is estimated. Meanwhile, a sub-dictionary is learned which is composed of several directional sub-dictionaries and adapts to the primary structure of each block class. In the second stage, better representations for each block are found by searching the sub-dictionaries obtained in the first stage.

Though the atoms in the dictionary do not change their shapes to adapt to a specific signal as other dictionaries learning methods have done (see [27]), each resulted sub-dictionary describes specific structure and reduces the scopes of further search. In this sense, TS_RS learns dictionaries to adapt to the structures of each block class or block.

III. Reconstruction Model Guided by Structured Sparsity on Overcomplete Dictionary

A. Block-Based CS Reconstruction

An image $I \in \mathbb{R}^{\sqrt{n}\times\sqrt{n}}$ is partitioned into L nonoverlapped image blocks of the size $\sqrt{B}\times\sqrt{B}$ ($\sqrt{B}=16$). Denote the measurements by $\mathbf{Y}=\Phi\mathbf{X}=(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_L)\in\mathbb{R}^{m_B\times L}$, where the image is $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_L)\in\mathbb{R}^{B\times L}$, and the measurement matrix is $\Phi\in\mathbb{R}^{m_B\times B}$ ($m_B<B$). Blocks are sparsely represented by the dictionary $\mathfrak{D}\in\mathbb{R}^{B\times N}$: $\mathbf{X}=\mathfrak{D}\mathfrak{S}$, where $\mathfrak{S}=(\mathfrak{s}_1,\mathfrak{s}_2,\ldots,\mathfrak{s}_L)\in\mathbb{R}^{N\times L}$, and $\mathbf{x}_i=\mathfrak{D}\mathfrak{s}_i=\mathbf{D}_i\mathbf{s}_i$, $i=1,2,\ldots,L$. The block-based reconstruction model [11] is given by

$$(\mathbf{D}_i^*,\mathbf{s}_i^*)=\arg\min_{\mathbf{D}_i,\mathbf{s}_i}\|\mathbf{y}_i-\Phi\mathbf{D}_i\mathbf{s}_i\|^2$$
$$s.t.\ \|\mathbf{D}_i^T\|_{p,0}\le K,\ i=1,2,\ldots,L. \quad (8)$$

B. Reconstruction Model Guided by Joint Sparsity

Compared with a full image, its local blocks have much more simple and consistent structures. Usually, an image block could easily find its structures in other blocks of the same image.

Fig. 2 shows the self-similarity property in nonoverlapped style. The image is divided into 16×16 blocks. Two groups of similar blocks are highlighted and shown below the image. The similar blocks are not necessarily spatially nearby. The self-similarity property have found successful applications in image inverse problems, such as de-noising [28]–[30], super-resolution [31], [32], CS [25], [33] and restoration [34], etc.

Guided by the assumption that a group of similar blocks could be represented by a common group of atoms, the L blocks of an image are divided into C ($C\ll L$) classes, and then a common atomic combination is found for each class. Equivalently, we impose joint sparsity [35], [36] constrains on the sparse matrix $\mathfrak{S}\in\mathbb{R}^{N\times L}$ as shown in Fig. 3. In this figure, the L sparse vectors have only C distinct sparse models. Each model could be determined by the collaboration of the blocks assigned to the class. By such constrains, the number of atomic combinations to be determined dramatically decreases, while information for determining each one of them increases. Consequently, more accurate and stable estimation is obtained.

Denote a partition of the image blocks by $Q:\mathbb{R}^B\to\{1,2,\ldots,C\}$, and the class label of $\mathbf{x}_i$, $i=1,2,\ldots,L$, by $\mathbf{q}_i=Q(\mathbf{x}_i)\in\{1,2,\ldots,C\}$. The reconstruction model

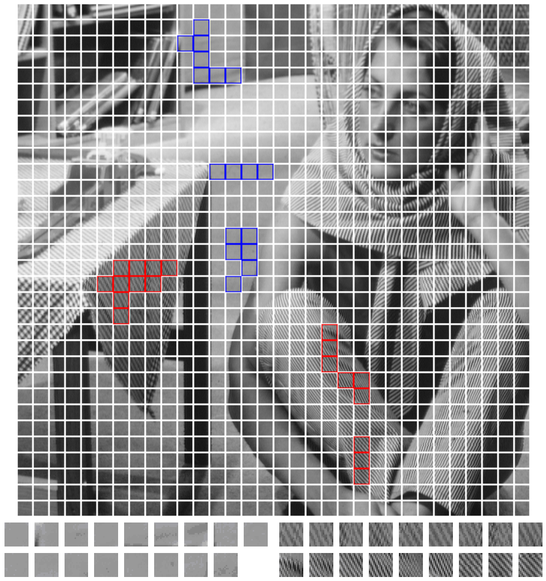


Fig. 2. Blocks of Barbara and two groups of similar blocks. The groups of similar blocks are highlighted and shown below the image. Each group shares similar structures and could be represented by a common atomic combination but different combinational coefficients.

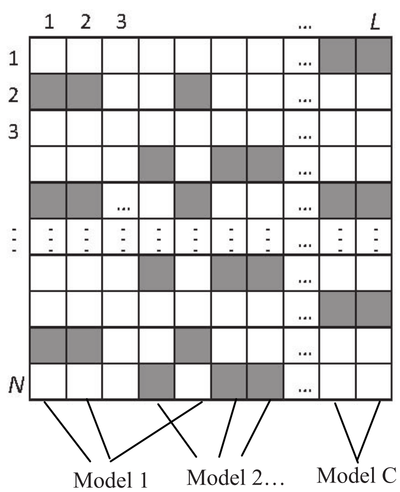


Fig. 3. Diagram of sparse matrix $\mathfrak{S} \in \mathbb{R}^{N\times L}$ with joint sparsity structure. Each of the columns is the sparse coefficient vector of an image block vector. A gray box is a nonzero component, and a white box a zero component. The L columns are supposed to have only C ($C \ll L$) distinct sparse models.

constrained by the joint sparsity structure is given by

$$(\hat{\mathbf{D}}_i^*, \mathbf{S}_i^*) = \arg\min_{\hat{\mathbf{D}}_i, \mathbf{S}_i} \|\mathbf{Y}_i - \Phi\hat{\mathbf{D}}_i\mathbf{S}_i\|_F^2$$
$$s.t.\ \|\hat{\mathbf{D}}_i^T\|_{p,0} \le K,\ i = 1, 2, \ldots, C \tag{9}$$

where $\mathbf{Y}_i = \{\mathbf{y}_j|\mathbf{q}_j = i\}$, $i = 1, 2, \ldots, C$, and the coefficient matrix $\mathbf{S}_i$ is calculated by $(\Phi\hat{\mathbf{D}}_i)^+\mathbf{Y}_i$. Each column of $\mathbf{S}_i$ is the coefficients of $\hat{\mathbf{D}}_i^T$ for $\mathbf{y}_j \in \mathbf{Y}_i$.

The obtained $\hat{\mathbf{D}}_i^*$ is the common atomic combination for all the blocks in the ith class

$$\bar{\mathbf{D}}_j = \hat{\mathbf{D}}_i^*,\ \mathbf{q}_j = i,\ j = 1, 2, \ldots, L,\ i = 1, 2, \ldots, C. \tag{10}$$

Then the image $\mathbf{X}$ is evaluated by $\bar{\mathbf{X}} = (\bar{\mathbf{D}}_1\bar{\mathbf{s}}_1, \bar{\mathbf{D}}_2\bar{\mathbf{s}}_2, \ldots, \bar{\mathbf{D}}_L\bar{\mathbf{s}}_L)$, where $\bar{\mathbf{s}}_j$, $j = 1, 2, \ldots, L$, is computed by (7) according to $\bar{\mathbf{D}}_j$ and $\mathbf{y}_j$.

Here is another question: How to get the partition Q, since only the compressed measurements of blocks are available? Distances between measurement vectors could be used as the substitutes of distances between image blocks, since Johnson-Lindenstrauss Lemma [37] shows that Gaussian projection is approximately distance preserving. In this paper, the partition Q is acquired by applying Affinity Propagation (AP) [38] clustering method to the measurement vectors $\mathbf{Y}$. AP is simple, fast and efficient. It does not require the number of clusters in advance.

C. Overcomplete Dictionary of Ridgelet

The overcomplete dictionary of Ridgelet proposed in [11] is employed to introducing sparsity for image blocks. The dictionary is denoted by $\mathfrak{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N)$, where $\mathbf{d}_i$, $i = 1, 2, \ldots, N$, is the ith column of $\mathfrak{D}$. The dictionary is constructed by scaling, shifting and rotating a prototype Ridgelet [39] function. The parameters of direction, scale and shift of $\mathbf{d}_i$ are denoted by $\gamma_i = (\theta_i, a_i, b_i)$, which is a sample of the parameter space $\Gamma_\gamma = \{(\theta, a, b)\}$. The sampling intervals are $\pi/36$, 0.2, and 1 for θ, a and b respectively. The dictionary is composed of 11281 atoms, which could be divided into 36 directional sub-dictionaries.

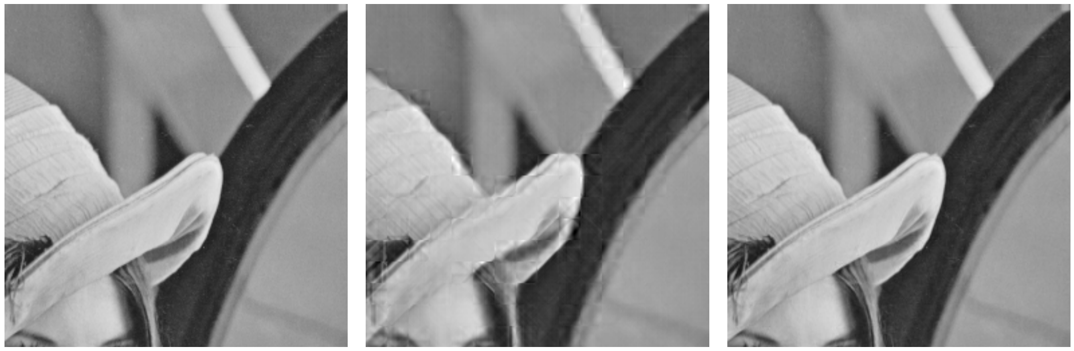
Of the three parameters, direction is important for adaptively representing orientated structure. When a block is sparsely coded by a given dictionary, the atoms are activated whose geometric structures are coherent with that of the block. To ensure the dictionary that is enriched by atoms with different orientations, the sampling intervals must be small enough. Fig. 4 exhibits the sparse representation results of a 256×256 image which is a quarter of *Lena*. OMP [13] is used as the sparse coding algorithm. Two dictionaries are employed respectively, named by DIC1 and DIC2. Their parameters are the same except the sampling intervals of direction: $\pi/4$ for DIC1 and $\pi/45$ for DIC2. As shown in the figure, the resulted image approximated by DIC1 is blurred and has obvious block artifacts; while the image approximated by DIC2 is visually faithful to the original image.

IV. TS_RS OF NATURE-INSPIRED OPTIMIZATION ALGORITHMS

This section represented the implementing details of TS_RS of nature-inspired optimization algorithms. Specifically, the designs of GA and CSA for each stage of TS_RS are represented. It is also displayed how the structured priors of images are incorporated by novel and flexible ways into the nature-inspired optimization algorithms.

A. Genetic Algorithm for Determining Common Atomic Combinations in Direction for Block Class

In the first stage, blocks are clustering into C ($C \approx 0.1L$) classes by applying AP method to the compressed measurements. Then GA [20] is employed to determine the atomic



(a) (b) (c)

Fig. 4. Comparison of the images approximated by two dictionaries of different direction sampling intervals. Each 16×16 block is represented by 32 atoms. (a) Original image. (b) Image using DIC1 composed of 1201 atoms, 32.85dB (0.9824). The directional sampling interval is $\pi/4$. (c) Image using DIC2 composed of 14116 atoms, 41.42dB (0.9975). The directional sampling interval is $\pi/45$.

Algorithm 1 Reconstruction Algorithm in the First Stage

Input: $\mathbf{Y}$, $\mathfrak{D}$, Φ, K;
Output: Image estimation $\bar{\mathbf{X}}$, solution population $\bar{\mathbf{P}}_i$, and its best individual $\bar{\mathbf{b}}_i$, $i = 1, 2, \ldots, L$.
- **Cluster** $\mathbf{Y}$ into C classes: $\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_C}$;
for $j = 1, 2, \ldots, C$, **do**
1: Initialization $\mathbf{P}^{(0)}$, $t = 0$;
2: Perform crossover to $\mathbf{P}^{(t)}$ and get $\mathbf{P}_{Cr}^{(t)}$;
3: Perform mutation to $\mathbf{P}_{Cr}^{(t)}$ and get $\mathbf{P}_{M}^{(t)}$;
4: Selects individuals among $\mathbf{P}_{M}^{(t)} \bigcup \mathbf{P}^{(t)}$ to get $\mathbf{P}^{(t+1)}$;
5: $t \leftarrow t + 1$; if $t > \bar{t}_{max}$, stop; else turn to 2;
6: $\hat{\mathbf{b}}_j^* = \arg\max f_s(\mathbf{b}^{(t)})$, $\mathbf{b}^{(t)} \in \mathbf{P}^{(t)}$;
7: $\forall i \in \{k|\mathbf{q}_k = j\}$, $\bar{\mathbf{P}}_i = \mathbf{P}^{(t)}$, $\bar{\mathbf{b}}_i = \hat{\mathbf{b}}_j^*$.
end for
- **Estimate** $\bar{\mathbf{x}}_i$ by $\bar{\mathbf{b}}_i$: $\bar{\mathbf{D}}_i = dec(\bar{\mathbf{b}}_i)$, $\bar{\mathbf{x}}_i = \bar{\mathbf{D}}_i[(\Phi\bar{\mathbf{D}}_i)^{+}\mathbf{y}_i]$; and $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_L)$.

combination for each block class. The engineering implementation of GA includes designing evolution encoding and decoding, fitness function and several operators, etc. The overall reconstruction algorithm in this stage is exhibited in Algorithm 1.

1) Evolution Encoding and Decoding: Throughout the work, the reconstruction is modeled as a combinatorial problem: choosing no more than K out of L atoms for sparse representation. Since the subscript i uniquely identifies the atom $\mathbf{d}_i$, $i \in \{1, 2, \ldots, L\}$, an atomic combination $\mathbf{D} = (\mathbf{d}_{n_1}, \mathbf{d}_{n_2}, \ldots, \mathbf{d}_{n_K}) \subset \mathfrak{D}$ is uniquely corresponding to its subscript sequence $n_1 n_2 \cdots n_K$. So an atomic combination is encoded into an integer sequence that is composed of the indices of the atoms. Denote the encoding as $\mathbf{b} = enc(\mathbf{D})$, while the decoding as $\mathbf{D} = dec(\mathbf{b})$. They are shown below

$$\mathbf{D} = (\mathbf{d}_{n_1}, \mathbf{d}_{n_2}, \ldots, \mathbf{d}_{n_K}) \underset{\text{dec}}{\overset{\text{enc}}{\rightleftharpoons}} \mathbf{b} = n_1 n_2 \cdots n_K$$

where $n_j \in \{1, 2, \ldots, L\}$, $j = 1, 2, \ldots, K$, and n_j is the basic operating unit. Before any operators on $\mathbf{D}$ or $\mathbf{b}$, a repetition check will be carried out; the repetitive ones will be deleted to ensure that the chosen atoms are different. A group of encoded solutions (also called individuals) is assembled into a solution population. The tth generation of population is denoted by $\mathbf{P}^{(t)} = (\mathbf{b}_1^{(t)}, \mathbf{b}_2^{(t)}, \ldots, \mathbf{b}_{Ps}^{(t)})$, where Ps is the population size.

2) Initialization: Two initialization ways are made a combined use in GA. In the first way, a block is considered as smooth, a totally random way is adopted to initialize the population as it is usually done in GA. In the second way, each individual in the population is initialized by a directional sub-dictionary. In other words, each directional sub-dictionary generates one individual for the population. The population initialized by this way could represent adequate single-directional structures.

Determining whether a block is smooth is according to the variance of it. Calculate the variance σ_i of $\mathbf{y}_i$, $i = 1, 2, \ldots, L$. The average variance $\bar{\sigma}$ is used as the threshold. If $\sigma_i < \bar{\sigma}$, $\mathbf{x}_i$ is considered as smooth.

The initializing way for a block class is voted by the blocks assigned to the class. The way taken by the majority is the way used for initialization.

3) Fitness Function: A fitness function is used for evaluating individuals. The greater the function value, the better the individual. The fitness function used for evaluating how $\mathbf{b}_i^{(t)} \in \mathbf{P}^{(t)}$ fits the target $\mathbf{Y}_j$ is defined by

$$f_s(\mathbf{b}_i^{(t)}) = 1 \big/ \|\mathbf{Y}_j - \Phi dec(\mathbf{b}_i^{(t)})\mathbf{S}_j\|_F^2. \tag{11}$$

4) Operators: GA simulates Darwin's evolutionary theory. The operators include: genetic crossover, genetic mutation and genetic selection.

Genetic crossover is the major operator. Here a one-point crossover strategy is employed. Firstly, choose a crossover mate for each individual in the population at random. For each individual couple, the crossover happens with probability p_c. Then the crossover pair $(\mathbf{b}_i^{(t)}, \mathbf{b}_j^{(t)})$ exchange over a random position l, $1 < l < K$, as shown below

$$\begin{array}{l} \mathbf{b}_i^{(t)} = n_1 \cdots n_l n_{l+1} \cdots n_K \\ \mathbf{b}_j^{(t)} = n'_1 \cdots n'_l n'_{l+1} \cdots n'_K \end{array} \rightarrow \begin{array}{l} \bar{\mathbf{b}}_i^{(t)} = n_1 \cdots n_l n'_{l+1} \cdots n'_K \\ \bar{\mathbf{b}}_j^{(t)} = n'_1 \cdots n'_l n_{l+1} \cdots n_K. \end{array}$$

The resulted individuals, $\bar{\mathbf{b}}_i^{(t)}$ and $\bar{\mathbf{b}}_j^{(t)}$, join the crossover population $\mathbf{P}_{Cr}^{(t)}$, which is composed of $(2 \times Ps \times p_c)$ individuals

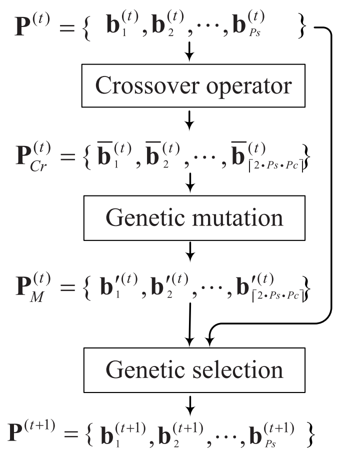


Fig. 5. Diagram of the operators of GA in an iteration.

in total. By the crossover operator, the atoms with multiorientations are assembled into an individual to adapt to the multioriented blocks.

The genetic mutation is operated on $\mathbf{P}_{Cr}^{(t)}$. Each position of the individual $\forall n_l \in \mathbf{b}_i^{(t)} \in \mathbf{P}_{Cr}^{(t)}$ is replaced by a random selected integer $n_l' \in \{1, 2, \ldots, N\}$ with probability $\bar{p}_m$. Supposed that the mutation takes place on positions: $l_1, l_2, \ldots$, the operator is shown below

$$\mathbf{b}_i^{(t)} = n_1 \cdots n_{l_j} \cdots n_K \rightarrow \mathbf{b'}_i^{(t)} = n_1 \cdots n_{l_j}' \cdots n_K$$

where $j = 1, 2, \ldots$. The resulted population is denoted by $\mathbf{P}_M^{(t)}$.

The genetic selection operator chooses Ps individuals with the largest fitness values out of $\mathbf{P}_M^{(t)} \bigcup \mathbf{P}^{(t)}$, and gets the next generation of population $\mathbf{P}^{(t+1)}$. The diagram of an iteration of GA is shown in Fig. 5.

5) Others: Taking time cost into account, we stop the search up to $\bar{t}_{\text{mat}}$ iterations.

There are quit a few parameters, and it is hard to determine them by theoretical derivation. In experiments, a population contains $Ps = 36$ individuals. The crossover probability $p_c = 0.6$ and the mutation probability $\bar{p}_m = 0.02$. The maximum number of iterations $\bar{t}_{\text{mat}} = 200$.

The computational time is correlated to the number of calculating the fitness function. Take the action of implementing the pseudo inverse of a matrix as the basic operator, the total complexity in this stage is $O(C \times Ps \times p_c \times \bar{t}_{max})$. The time needed for the basic operator is related to the size of the matrix $(\Phi\mathbf{D}) \in \mathbb{R}^{m_B \times K}$. As we get more samples or set a bigger value for sparsity, more running time is needed.

B. Clonal Selection Algorithm for Determining Atomic Combinations in Scale and Shift for Each Block

Natural images are characterized by piecewise smooth and changing locally slowly. Hence the blocks spatially nearby usually share similar structures. The second stage of TS_RS makes use of the property to refine the results of the first stage.

Define the set $\Sigma_{\mathbf{d}_i}$ for the atom $\mathbf{d}_i$. It is composed of the atoms with the same direction and scale as $\mathbf{d}_i$ but different

Algorithm 2 Reconstruction Algorithm in the Second Stage

Input: $\mathbf{Y}$, $\mathfrak{D}$, Φ, $\bar{\mathbf{P}}_i$, $\bar{\mathbf{b}}_i$, $i = 1, 2, \ldots, L$.
Output: Atomic combinations $\mathbf{D}_i^*$, $i = 1, 2, \ldots, L$, and image estimation $\mathbf{X}^*$.
for $i = 1, 2, \ldots, L$, **do**
1: Initialization: $\mathbf{P}^{(0)} = \bar{\mathbf{P}}_i \bigcup \{\bar{\mathbf{b}}_j | j \in \mathcal{N}_i \bigcup \bar{\mathcal{N}}_i\}$, $t = 0$, supposed $\mathbf{P}^{(t)} = \{\mathbf{b}_1^{(t)}, \mathbf{b}_2^{(t)}, \cdots\}$;
2: Clone $\mathbf{P}^{(t)}$ and get $\mathbf{P}_C^{(t)} = \{\mathbf{P}_C^{(t)}(1), \mathbf{P}_C^{(t)}(2), \cdots\}$;
3: Mutate $\mathbf{P}_C^{(t)}$ and get $\mathbf{P}_M^{(t)} = \{\mathbf{P}_M^{(t)}(1), \mathbf{P}_M^{(t)}(2), \cdots\}$;
4: Generate $\mathbf{P}^{(t+1)} = \{\mathbf{b}_1^{(t+1)}, \mathbf{b}_2^{(t+1)}, \ldots\}$, where $\mathbf{b}_j^{(t+1)}$is the best individual of $\mathbf{b}_j^{(t)} \bigcup \mathbf{P}_M^{(t)}(j)$, $j = 1, 2, \ldots$, $t \leftarrow t + 1$;
5: If $t = 1$, keep the best Ps individuals in $P^{(t)}$;
6: If $t > t_{max}$, stop; else turn to 2;
7: $\mathbf{b}_i^* = \arg\max f_s(\mathbf{b}^{(t)})$, $\mathbf{b}^{(t)} \in \mathbf{P}^{(t)}$; $\mathbf{D}_i^* = dec(\mathbf{b}_i^*)$;
end for
- $\mathbf{X}^* = (\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_L^*)$, where $\mathbf{x}_i^* = \mathbf{D}_i^*[(\Phi\mathbf{D}_i^*)^+\mathbf{y}_i]$, $i = 1, 2, \ldots, L$.

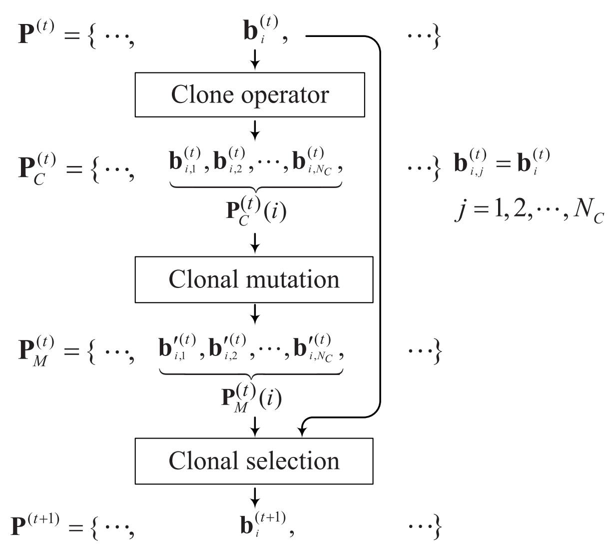


Fig. 6. Diagram of the operators of CSA in an iteration.

shift parameters

$$\Sigma_{\mathbf{d}_i} \triangleq \{\mathbf{d}_j | \theta_p = \theta_i,\ a_j = a_i,\ \mathbf{d}_j \in \mathfrak{D}\}.$$

Define the set $\Sigma_\mathbf{D}$ for the atomic combination $\mathbf{D} = (\mathbf{d}_{n_1}, \mathbf{d}_{n_2}, \ldots, \mathbf{d}_{n_K})$

$$\Sigma_\mathbf{D} \triangleq \left\{(\mathbf{d}_{n_1}', \mathbf{d}_{n_2}', \ldots, \mathbf{d}_{n_K}') \middle| \mathbf{d}_{n_j}' \in \Sigma_{\mathbf{d}_{n_j}}, j = 1, 2, \ldots, K\right\}.$$

Given an image block $\mathbf{x}_i$ is well approximated by $\mathbf{D}_i$, and one of its adjacent block $\mathbf{x}_j$ $(i \neq j)$ shares similar structure, the assumption is cast that a member of $\Sigma_{\mathbf{D}_i}$ could sparsely represent $\mathbf{x}_j$. In this stage, the blocks potentially share similar structure to $\mathbf{x}_i$, $i = 1, 2, \ldots, L$, are found. Their atomic combinations obtained in the first stage and their shift versions are evaluated for $\mathbf{x}_i$, and the best one of them survives. This process is implemented by CSA algorithm.

CSA simulates Burnet's clonal selection principles [21], [22]. The algorithm of this stage is shown in Algorithm 2.

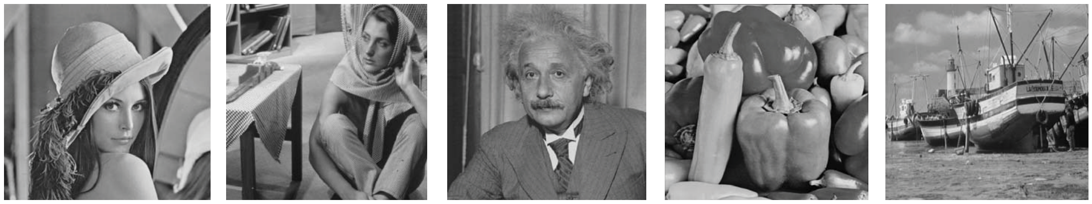


Fig. 7. Natural images used in the experiments.

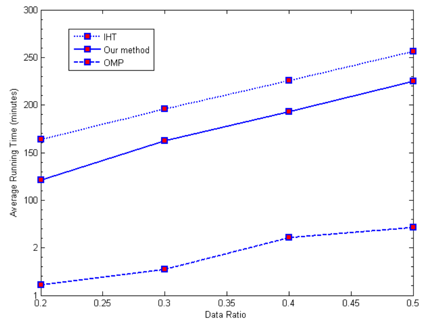


Fig. 8. Average time needed by different methods for reconstructing Barbara.

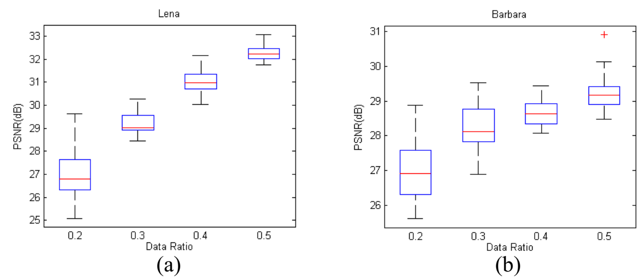


Fig. 9. Box plots for depicting the stability of the proposed TS_RS. Each box is the statistical analysis of 30 reconstruction results. (a) Box plot for Lena. (b) Box plot for Barbara.

1) Initialization: The initial population for $\mathbf{x}_i$ is composed of three parts: the output population $\bar{P}_i$ of the first stage, the best individuals of its nonlocal and local neighbors. Denote the indices of n_1 ($n_1 = 4$) blocks most similar (measured by the Euclidean distance between measurement vectors) to $\mathbf{x}_i$ by $\mathcal{N}_i$, and the indices of n_2 ($n_2 = 8$) neighbors of $\mathbf{x}_i$ by $\bar{\mathcal{N}}_i$. The population is initialized by $\bar{\mathbf{P}}_i \bigcup \{\bar{\mathbf{b}}_j | j \in \mathcal{N}_i \bigcup \bar{\mathcal{N}}_i\}$.

2) Affinity function: The evaluating function in CSA is named affinity function. The affinity of $\mathbf{b}_j^{(t)} \in \mathbf{P}^{(t)}$ for $\mathbf{y}_i$ is defined by

$$f_a(\mathbf{b}_j^{(t)}) = 1 / \|\mathbf{y}_i - \Phi dec(\mathbf{b}_j^{(t)})\mathbf{s}_i\|_2^2. \quad (12)$$

3) Operators: CSA operators include: clone, clonal mutation and clonal selection.

Clone operator replicates each individual of $\mathbf{P}^{(t)}$ for N_c times to get the clonal population $\mathbf{P}_C^{(t)}$. Denote the sub-population generated by $\mathbf{b}_i^{(t)} \in \mathbf{P}^{(t)}$ by $\mathbf{P}_C^{(t)}(i)$. Then $\mathbf{P}_C^{(t)} = \{\mathbf{P}_C^{(t)}(1), \mathbf{P}_C^{(t)}(2), \cdots\}$.

Clonal mutation is the major operator. Denote the clonal mutation population by $\mathbf{P}_M^{(t)} = \{\mathbf{P}_M^{(t)}(1), \mathbf{P}_M^{(t)}(2), \ldots\}$, where the sub-population $\mathbf{P}_M^{(t)}(j)$ is obtained by mutating $\mathbf{P}_C^{(t)}(j)$. supposed $\mathbf{b}_i^{(t)} = n_1 n_2 \cdots n_K \in \mathbf{P}_C^{(t)}$, and the mutation takes place on positions: $l_1, l_2, \ldots$, the operator is shown below

$$\mathbf{b}_i^{(t)} = n_1 \cdots n_{l_j} \cdots n_K \rightarrow \mathbf{b}_i'^{(t)} = n_1 \cdots n'_{l_j} \cdots n_K$$

where $n'_{l_j} \in \{k | \mathbf{d}_k \in \Sigma_{\mathbf{d}_{n_{l_j}}}\}$, $j = 1, 2, \cdots$. The clonal mutation and genetic mutation are similar except two differences. One is the mutation probability p_m is much higher than $\bar{p}_m$. The other one is that the mutation scope is the dictionary $\mathfrak{D}$ in GA, while is limited in a much smaller sub-dictionary in CSA.

Clonal selection operator selects the best individuals of $\mathbf{P}_M^{(t)}(i) \bigcup \mathbf{b}_i^{(t)}$, $i = 1, 2, \cdots$ to get $\mathbf{P}^{(t+1)}$. To keep the population size to be Ps, an extra selection is needed for $\mathbf{P}^{(1)}$.

The diagram of an iteration of CSA is shown in Fig. 6. It can be seen that each individual initializes a search. Each resulted individual is uniquely correlated to its parental individual. Compared Figs. 5 and 6, both GA and CSA take the elitism strategy, which always keeps the best solutions obtained so far. The new generation of population will be guaranteed to be better than its previous ones. The decrease of the objective functions and the convergence of the algorithms are also guaranteed. The combined use of the two algorithms is according to the characteristics of them and different demands of the models of TS_RS.

4) Others: We stop the search up to t_{mat} iterations. A population contains $Ps = 36$ individuals; except the initial population contains $(Ps + n_1 + n_2)$ individuals. The clone size of each individual is $N_c = 10$. There are $Ps \times N_c$ individuals in the clonal population. The mutation probability $p_m = 0.3$. The maximum number of iterations $t_{\text{mat}} = 20$. The computational complexity is $O(L \times Ps \times N_c \times p_c \times t_{\max})$.

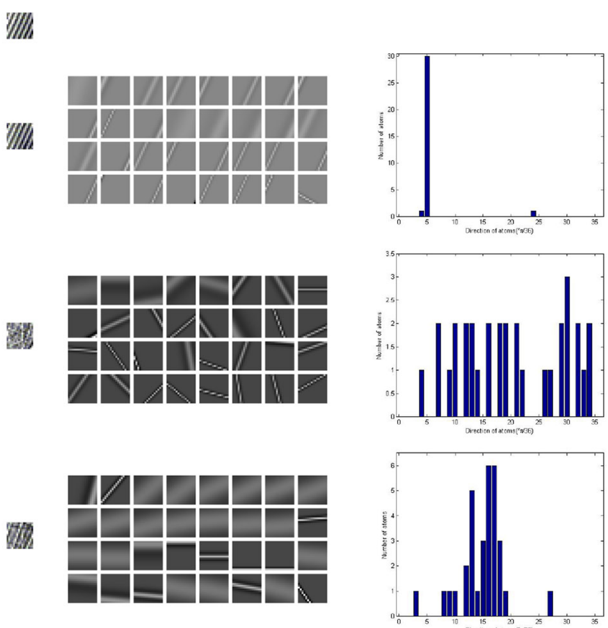
V. Experiments

The experiments are committed on five natural images of size 512×512 depicted in Fig. 7. An image will be divided into $L = 1024$ blocks. Each block is assumed to be sparsely represented by $K = 32$ atoms out of the dictionary composed of $N = 11281$ atoms. All the results are based on the average of 5 trials due to the randomness of the compressed sampling operator. For each trial, a random Gaussian measurement matrix is used for sampling, and the TS_RS will be run for 6 times due to the randomness of its nature-inspired optimization algorithms. In a word, for each data ratio, the result is the average of 30 times tests for TS_RS and 5 times for other methods, i.e., OMP [14] and IHT [40]. Data ratio $= m_B/B$, which is the ratio of the length of the measurement to the length of a block vector. All the experiments are run under windows XP and MATLAB 7.11 on PCs with Intel dual core CPU at

(a) (b) (c) (d)

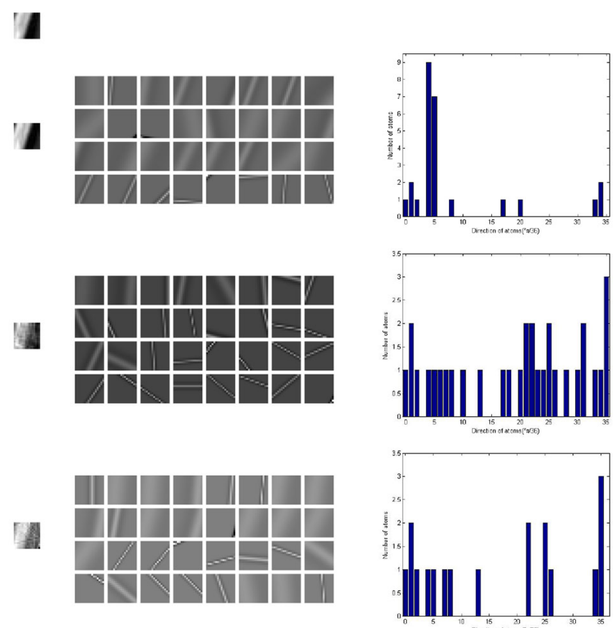
Fig. 10. Comparison of the images resulted by the two stages of TS_RS, and the images resulted by initializing the population totally by the first way, data ratio = 0.5. (a) and (b) show the images resulted by each stage of TS_RS; the PSNR and SSIM values are 26.32 and 0.8629 for (a), 30.91 and 0.9288 for (b). (c) and (d) show the resulted images of each stage by initializing the population totally by the first way; the PSNR and SSIM values are 25.25 and 0.8474 for (c), 27.84 and 0.9182 for (d).

TABLE I
TEXTURE BLOCKS RESULTED BY DIFFERENT METHODS*



*Data ratio=0.3. The first column is the blocks resulted by (top to bottom): original block, TS_RS(PSNR:23.66dB, SSIM:0.9444), OMP(PSNR:14.45dB, SSIM:0.3522) and IHT(PSNR:16.50dB, SSIM:0.7131). The second column shows the atoms for the blocks on their left. The last column are the histogram of the directions of the atoms.

TABLE II
EDGE BLOCKS RESULTED BY DIFFERENT METHODS*



*Data ratio=0.3. The first column is the blocks resulted by (top to bottom): original block, TS_RS(PSNR:32.62dB, SSIM:0.9894), OMP(PSNR:20.03dB, SSIM:0.8761) and IHT(PSNR:18.89dB, SSIM:0.8635). The second column shows the atoms for the blocks on their left. The last column are the histogram of the directions of the atoms.

3GHz and 4GB of memory. The reconstruction performance is evaluated in terms of PSNR and SSIM [41].

Fig. 8 compares the average reconstructing time needed for Barbara. OMP takes the least time. Its computational complexity is $O(L \times K)$, where the basic operator is computing the pseudo inverse of a matrix. The time computational complexity of IHT is $O(L \times t_{IHT})$, where t_{IHT} is the iteration number and the basic operator is the multiplication of matrix. To compare the methods in comparable time, set $t_{IHT} = 1500$. TS_RS could be accelerated by parallel computation since the individuals of a population could be processed simultaneously.

Fig. 9 depicts the box plots for illustrating the stability of TS_RS. Each box is the statistical analysis of 30 times tests for a given data ratio. TS_RS is shown stable. There are only few outliers. As the data ratio increases, more information about a signal is known, and the boxes become smaller.

Tables I and II depict the single-oriented blocks of Barbara reconstructed by different methods. The reconstructed blocks, the atoms chosen for the blocks and the histograms of the directions of the atoms are shown. By comparing the reconstructed blocks, it is shown that TS_RS results faithful reconstructions, while OMP and IHT result the blocks with blurred edges and orientations. By comparing the histograms, it can be seen that the directions of the atoms employed by TS_RS are highly concentrated. Most atoms fall into one or two directional sub-dictionaries. While the directions of the atoms employed by OMP and IHT are scattered. The atoms scattered in more than half of the directional sub-dictionaries.

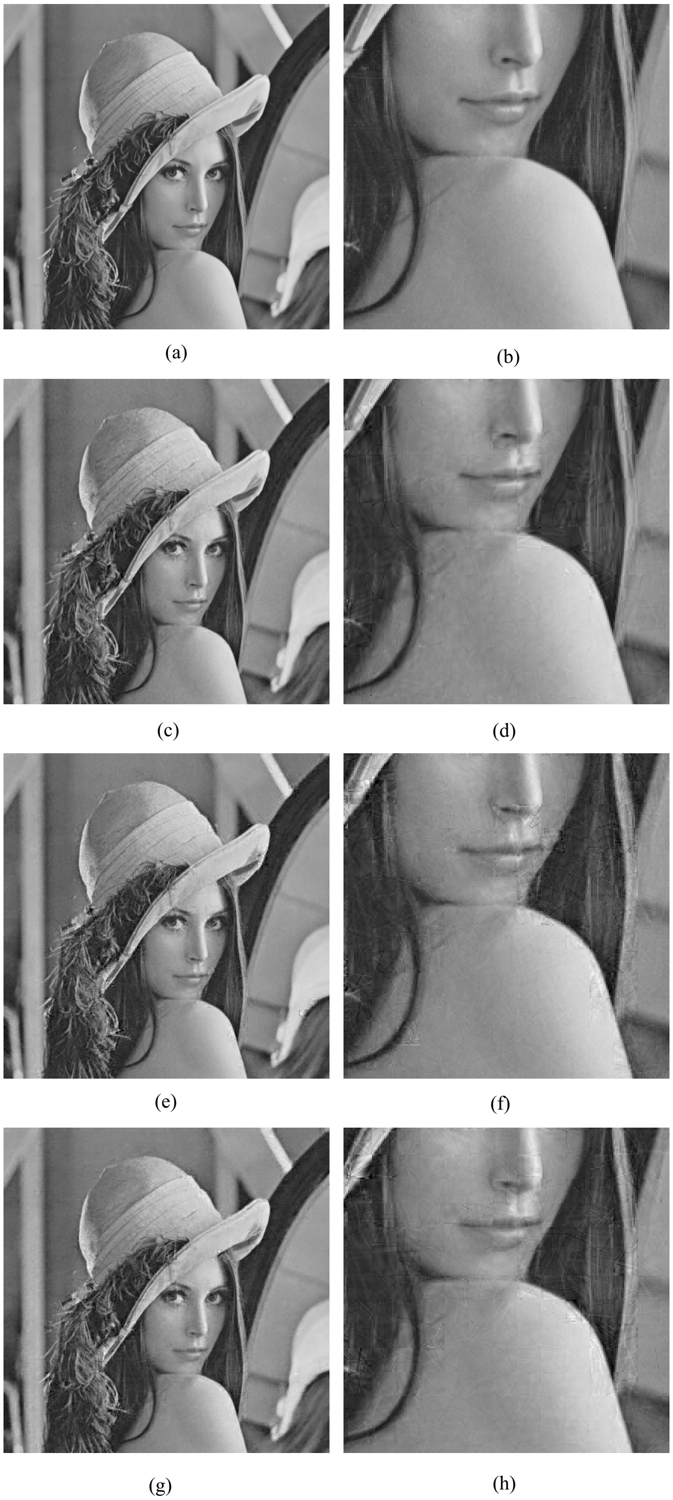


Fig. 11. Best reconstruction results of *Lena* by different methods (data ratio = 0.3). (a) Original. (b) Part of (a). (c) TS RS, 32.16dB (0.9331). (d) Part of (c). (e) OMP, 28.87dB (0.9165). (f) Part of (e). (g) IHT, 30.83dB (0.9163). (h) Part of (g).

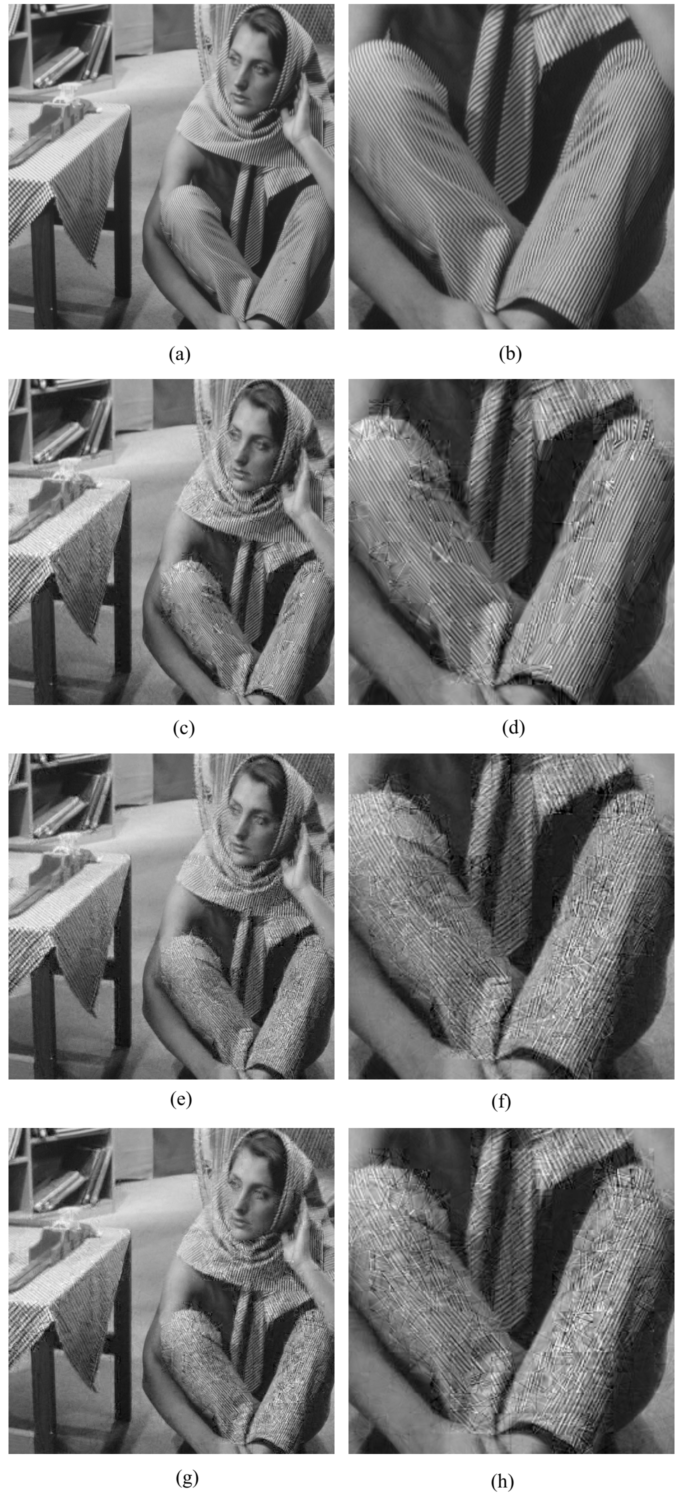


Fig. 12. Best reconstruction results of Barbara by different methods (data ratio = 0.3). (a) Original. (b) Part of (a). (c) TS RS, 29.52dB (0.8891). (d) Part of (c). (e) OMP, 23.86dB (0.8570). (f) Part of (e). (g) IHT, 27.46dB (0.8509). (h) Part of (g).

Fig. 10 compares the images resulted by each stage of TS_RS and that resulted by initialing the population only by the fist way. The reconstructed images of the second stage have a lot of improvements on that of the first stage. The figures of (b) and (d) have much more accurate and consistent details than that of (a) and (c). Compared (a) and (b) with (c) and (d), the combined initializing method adopted by TS_RS is advantageous in handling single-oriented blocks, which results sharper edges and clearer regular textures.

Table III shows the average reconstructed results in PSNR and SSIM. TS_RS overpasses OMP and IHT in almost all items. Figs. 11 and 12 compare the resulted images.

TABLE III
AVERAGE PSNR(dB) AND SSIM RESULTS BY DIFFERENT METHODS

Image	Method	Data ratio			
		0.2	0.3	0.4	0.5
Lena	TS_RS	29.22(0.8711)	31.05(0.9306)	32.23(0.9530)	32.85(0.9612)
	OMP	25.50(0.8472)	28.08(0.9080)	30.51(0.9442)	32.17(0.9599)
	IHT	27.88(0.8496)	30.11(0.9188)	31.56(0.9468)	32.24(0.9590)
Barbara	TS_RS	26.91(0.7869)	28.23(0.8820)	28.67(0.9181)	29.25(0.9368)
	OMP	21.69(0.7701)	22.87(0.8241)	25.43(0.8991)	27.08(0.9288)
	IHT	24.79(0.7517)	26.40(0.8477)	27.94(0.9027)	28.91(0.9288)
Einstein	TS_RS	31.37(0.8440)	33.15(0.9165)	34.11(0.9414)	34.86(0.9520)
	OMP	28.32(0.8444)	29.86(0.8902)	32.67(0.9341)	34.24(0.9528)
	IHT	30.38(0.8317)	32.15(0.9013)	33.32(0.9329)	34.34(0.9486)
Peppers	TS_RS	29.29(0.8421)	30.58(0.9109)	31.56(0.9396)	32.13(0.9521)
	OMP	24.72(0.8227)	27.97(0.8920)	30.05(0.9273)	31.51(0.9470)
	IHT	28.09(0.8392)	29.40(0.8995)	30.86(0.9335)	31.74(0.9481)
Boats	TS_RS	26.93(0.7803)	28.60(0.8841)	29.74(0.9213)	30.26(0.9368)
	OMP	21.68(0.7416)	25.48(0.8628)	27.56(0.9120)	29.14(0.9372)
	IHT	25.97(0.7733)	27.69(0.8663)	29.12(0.9141)	29.87(0.9348)

Obviously, the images resulted by TS_RS are more faithful to the original images, and have clearer and better visual effect than that resulted by other methods. Let us compare the local enlarged images in the figures. In Fig. 11, TS_RS produces more consistent edges and less artifacts. In Fig. 12, TS_RS produces clearer and sharper effect in the regular texture part of the image. It is reasonable to conclude that TS_RS results the atomic combinations more adaptive to the structures of image blocks.

VI. CONCLUSION

In this paper, the TS_RS of nature-inspired optimization algorithms is proposed. It is the first time that natural-inspired optimization algorithms are employed for the l_0 regularized nonconvex CS reconstruction. The work presents the great potential in applying nature-inspired algorithms for nonconvex CS. Besides, this paper presents novel and flexible ways to incorporate structured priors of images. The structured sparsity and other priors based on the overcomplete dictionary are implemented by delicately designing the components of the nature-inspired methods, such as coding and decoding, evolutionary strategies, initialization, operators and evaluation functions, etc. It shows the great capability of nature-inspired optimization algorithms in flexibly handling nonconvex, nonlinear, and even nonformalized priors.

The shortcomings inherent in GA and CSA include: longer convergence times and fixed setting of parameters. In fact, the running time could be further reduced by employing parallel computation. We have been working toward faster algorithms by exploiting more structured constrains and delicately designing nature-inspired algorithms. As regards the parameters, they could be adaptively tuned according to the structure of the signal or the status of the solution population [42], [43]. Since lots of advanced nature-inspired methods have been available, say, [43], [44], better reconstruction results are expected to appear in the short term.

Throughout the paper, the CS reconstruction problem is modeled as finding an atomic combination for well approximating the signal. The maximum sparsity of the signals are prefixed, and the combinational coefficients of the atoms are calculated by the least squared formula according to the atomic combination. In fact, the sparsity estimation for individual blocks is one of the important issues in CS. The combinational coefficients could be better estimated by incorporating statistic priors, say, [33], [45]. Our research lists include: exploiting more structured priors of images, establishing advanced reconstruction models and their realization based on nature-inspired optimization algorithms, etc.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. Candès, "Compressed sampling," in *Proc. Int. Congr. Math.*, 2006, pp. 1433–1452.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] E. Yonina and K. Gitta, Eds., *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[6] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 23, pp. 3311–3325, Dec. 1997.

[7] B. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, Jun. 1996.

[8] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[9] H. Rauhut, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[10] E. Candès and Y. Eldar, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmonic Anal.*, vol. 31, pp. 59–73, Jul. 2011.

[11] L. Lin, F. Liu, and L. Jiao, "Compressed sensing by collaborative reconstruction on overcomplete dictionary," *Signal Process.*, vol. 103, pp. 92–102, Oct. 2014.

[12] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[13] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals Syst. Comput.*, vol. 1. Pacific Grove, CA, USA, 1993, pp. 40–44.

[14] 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.

[15] T. Blumensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmonic Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[16] X. Yang, *Nature-Inspired Metaheuristic Algorithms*. 2nd ed. Frome, U.K.: Luniver Press, 2010.

[17] L. Jiao and L. Wang, "A novel genetic algorithm based on immunity," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 5, pp. 552–561, Sep. 2000.

[18] L. Jiao, H. Du, F. Liu, and M. Gong, *Immunological Computation for Optimization, Learning, and Recognition*. Beijing, China: Science, 2006.

[19] M. Gong, L. Jiao, F. Liu, and W. Ma, "Immune algorithm with orthogonal design based initialization, cloning, and selection for global optimization," *Knowl. Inf. Syst.*, vol. 25, no. 3, pp. 523–549, 2010.

[20] H. Holland, *Adaptation in Natural and Artificial System: An Introductory Analysis With Application to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[21] L. Castro and F. Zuben, "The clonal selection algorithm with engineering applications," in *Proc. Genet. Evol. Comput. Conf. (GECCO)*, 2000, pp. 36–37.

[22] L. Castro and F. Zuben, "Learning and optimization using the clonal selection principle," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 239–251, Jun. 2002.

[23] J. Wu, F. Liu, L. Jiao, and X. Wang, "Compressive sensing SAR image reconstruction based on Bayesian framework and evolutionary computation," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1904–1911, Jul. 2011.

[24] L. Li, X. Yao, R. Stolkin, M. Gong, and S. He, "An evolutionary multiobjective approach to sparse reconstruction," *IEEE Trans. Evol. Comput.*, to be published.

[25] X. Wu, W. Dong, X. Zhang, and G. Shi, "Model-assisted adaptive recovery of compressive sensing with imaging applications," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 451–458, Feb. 2012.

[26] L. Gan, "Block compressed sensing of natural images," in *Proc. Int. Conf. Digital Signal Process.*, 2007, pp. 403–406.

[27] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[28] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[29] A. Buades, B. Coll, and M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 60–65.

[30] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, Jul. 2014.

[31] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[32] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4016–4028, Sep. 2012.

[33] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, May 2012.

[34] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, 2009, pp. 2272–2279.

[35] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solution to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[36] J. Tropp, A. Gilbert, and M. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. Int. Conf. Acoustics Speech Signal Process.*, 2005, pp. 1520–6149.

[37] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Struct. Algorith.*, vol. 22, no. 1, pp. 60–65, 2003.

[38] B. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[39] E. Candès, "Ridgelets: Theory and application," Ph.D. thesis, Dept. Statist., Stanford Univ., Stanford, CA, USA, 1998.

[40] T. Blumensath, "Accelerated iterative hard thresholding," *Signal Process.*, vol. 92, no. 3, pp. 752–756, 2012.

[41] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] M. Srinivas and L. Patnaik, "Adaptive probabilities of crossover and mutation in genetic algorithms," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 4, pp. 656–667, Apr. 1994.

[43] Z. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1362–1381, Dec. 2009.

[44] H. Wang and X. Yao, "Corner sort for Pareto-based many-objective optimization," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 92–102, Jan. 2014.

[45] J. Wu, F. Liu, L. Jiao, X. Wang, and B. Hou, "Multivariate compressive sensing for image reconstruction in the wavelet domain: Using scale mixture models," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 489–509, Dec. 2011.

Fang Liu (SM'07) received the B.S. degree in computer science and technology from the Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. degree in computer science and technology from the Xidian University, Xi'an, in 1995.

She is currently a Professor at Xidian University, Xi'an, China. She has authored or co-authored of five books and over 80 papers in journals and conferences. Her current research interests include image perception and pattern recognition, machine learning, and data mining.

Prof. Liu won the second prize of National Natural Science Award in 2013.

Leping Lin received the B.S. degree in computer science and application and the M.S. degree in computer application technology from Xidian University, Xi'an, China, in 2002 and 2005, respectively, where she is currently pursuing the Ph.D. degree in computer application technology.

Since 2005, she has been a Lecture of the Guilin University of Electronic Technology, Guilin, China. Her current research interests include the applications of nature-inspired computation, pattern recognition, sparse representation, and compressed sensing.

Licheng Jiao (SM'89) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He is currently a Distinguished Professor with the School of Electronic Engineering, Xidian University, Xi'an. He has led approximately 40 important scientific research projects and has published over 10 monographs and 100 papers in international journals and conferences. He is the Author of three books *Theory of Neural Network Systems* (Xidian University Press, 1990), *Theory and Application on Nonlinear Transformation Functions* (Xidian University Press, 1992), and *Applications and Implementations of Neural Networks* (Xidian University Press, 1996). He has authored or coauthored of over 150 scientific papers. His current research interests include signal and image processing, natural computation, and intelligent information processing.

Prof. Jiao is a member of the IEEE Xi'an Section Executive Committee and the Chairman of the Awards and Recognition Committee, and an Executive Committee Member of the Chinese Association of Artificial Intelligence. He won the second prize of National Natural Science Award in 2013.

Lingling Li received the B.S. degree from the School of Electronic Engineering, Xidian University, Xi'an, China, where she is currently pursuing the Ph.D. degree from the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China.

Her current research interests include data mining in networks and computational intelligence.

Shuyuan Yang (M'07) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

She is currently a Post-Doctoral Fellowship at the Institute of Intelligent Information Processing, Xidian University. Her current research interests include image processing and machine learning.

Biao Hou (M'07) received the B.S. and M.S. degrees in mathematics from Northwest University, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, in 2003.

Since 2003, he has been with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, where he is currently a Professor. His current research interests include dictionary learning, deep learning, compressive sensing, and synthetic aperture radar image processing.

Hongmei Ma received the M.S. degree in computer application technology from Xidian University, Xi'an, China, in 2012.

Her current research interests include compressed sensing by clonal selection algorithms.

Li Yang received the M.S. degree in computer application technology from Xidian University, Xi'an, China, in 2012.

Her current research interests include structured compressed sensing by genetic algorithms.

Jinghuan Xu received the M.S. degree in computer application technology from Xidian University, Xi'an, China, in 2011.

Her current research interests include sparse representation by overcompete dictionaries.