Sparsity and Low Rank Dictionary Learning for Sparse Representation of Monogenic Signal

Ganggang Dong, Na Wang, Gangyao Kuang, and Hongbing Qiu

***Abstract*—This paper proposes a new framework of dictionary learning for a recently developed study, sparse representation of monogenic signal. The proposed framework is applied to target recognition in SAR image. Unlike the preceding works, where the sparse model is formed via an over-complete dictionary whose atoms are the training samples themselves, a new approach to learn a more discriminative dictionary is proposed. To achieve target classification, two specific implementation schemes, global learning and local learning, are developed. The global learning generates a single, universal dictionary for all target class. Since the class membership committed to each atom (element) is lost, the reconstruction error committed to each class is incapable to be computed. The conventional decision rule by the minimal residual could not be applied any more. Hence, this paper develops two decision rules for global learning. The first resorts to a third-party classifier, while the other recommends a non-parametric criterion. Different from the global learning, the local learning learns a sub-dictionary for each target class. These sub-dictionaries are concatenated to form a global one. Each sub-dictionary has been learned within certain target class, the generated atoms can be therefore labeled. The identification is predicted by evaluating which class of atom could produce the minimal reconstruction error. The effectiveness of proposed scheme is verified with multiple comparative studies.**

Keywords*—*Dictionary learning, sparse representation, low-rank, SAR, target recognition.

I. Introduction

AUTOMATIC target recognition is a challenging research topic of radar image interpretation [1]. It plays an increasingly important role in the military and civil fields. Though widely pursued over the past decades, it is still an open problem. The major obstacle of target recognition consists in these extended operating conditions, where a single operational parameter is significantly different between the images used for training and those for testing [2], [3]. The representatives include but are not limited to, variations in squint angle, target articulation and configurations, obscuration due to occlusion and/or layover, and intra-class target variability. The common optical imaging sensor typically projects target onto a plane perpendicular to the line of sight. Differently, the received backscattering returns of SAR sensor are actually cast into a "slant plane" defined by radar line-of-sight and the platform velocity vector. Precisely, SAR image projection plane contains the line-of-sight of sensor itself. Hence, multiple scattering mechanisms contribute to the backscattering signal, for example, the direct backscatter, the single-direction double bounce, the return-direct multi-bounce, the interactions with local ground plane, and the higher-order multiple bounce from cavity-like structures [2]. Due to the complex imaging mechanism, it is difficult to extract effective feature from SAR image itself. To circumvent this problem, the preceding works introduce a multi-dimensional analytic signal, the monogenic signal [4]. The target scattering phenomenology is characterized by the multi-resolution monogenic signal [5], [6], with which the broad spectral information and spatial localization of SAR image can be captured [7], [8]. The components of monogenic signal at different scale-space are used to generate the monogenic feature. The resulting feature is fed into the framework of sparse signal modeling. Specifically, the monogenic features of training sample are used to form an over-complete dictionary to encode the counterpart of query as a sparse linear combination of themselves. Sparsity is harnessed to generate the optimal representation, from which the final decision can be reached.

Though the preceding works achieve some improvement in terms of recognition performance, much more attention should be paid to the over-complete dictionary. These studies form the sparse model via an over-complete dictionary whose atoms are the training sample themselves. Since the generated dictionary is pre-specified, it often contains correlated versions of atom, each of which is usually densely sampled by the generic sensors. These limitations are adverse to target discrimination and computational efficiency [9]. Moreover, the computational cost and memory consumption could be unacceptable if a great many training samples are available.

Some recent studies prove that great advantage can be achieved by learning a discriminative and compact dictionary for sparse signal model, rather than the purely re-constructive and pre-specified ones [10]. The sparse models built by the learned dictionary demonstrate some advantages concerning the recognition accuracy and the computational efficiency. The first work on learned sparse model is the K. Kreutz-Delgado *et al.*'s [11]. They learn an environmentally adapted dictionary, and solve the presented problem by alternating the generation of a representative set of sparse representation and the update of dictionary using the resulting representation. Multiple studies have been then pursued following the road-map developed in Ref [11]. Michal Aharon *et al.* produce a dictionary that could generate the best representation for observed signals under specific sparsity constraint [12]. Gianluca Monaci *et al.* present a multimodal signals based on the sparse decomposition over a dictionary of learned multimodal structures [13], with which the shifted at all positions of the signal can be

Dr. Dong, Dr. N. Wang, and Prof. Kuang are with School of Electronics Science and Engineering, National University of Defense Technology, Changsha, China, 41073 (e-mail: dongganggang@nudt.edu.cn).

Prof. Qiu is with School of Information and Communication, Gulin University of Electronic Technology, Guilin, China, 541000.

solved. Ivana Tosic and Pascal Frossard propose a new method to learn an over-complete dictionary [14], in which the multi-view geometric structure of stereo images was exploited. Meng Yang *et al.* present a new dictionary learning method [15], [16], where a Fisher discrimination criterion for structured dictionary learning was developed. Julien Mairal *et al.* also presents a family of learning methods, online dictionary learning [17], [18], in which the stochastic approximations was used to formulate an convex optimization problem. Ignacio Ramirez *et al.* recommend a dictionary learning strategy for classification and clustering, where an incoherence promoting term that encourages dictionaries associated to different classes to be as independent as possible was developed [19]. This model is further utilized to action recognition [20], where the similarity constrained term was imposed on sparse representation.

Inspired by the works mentioned above, this paper proposes a dictionary learning framework. We aim to improving the preceding work, sparse representation of monogenic signal [7], [8], by a learned dictionary, whose discriminative power can be promoted. Unlike the previously published works, where the sparse model is formed by an over-complete dictionary, whose atoms are the training sample themselves, this paper proposes to build a learned sparse model, by which the recognition performance can be improved. Our goal is to generate a set of basis atoms, over which the training sample can be sparsely represented. To achieve target classification, two specific implementation schemes, global learning and local learning, are proposed. The global learning scheme generates a single and universal dictionary. Since the class membership associated with each atom is lost, the conventional decision with the minimal reconstruction error could not be applied any more. To predict the identity, two decision rules are developed. The first resorts to a third-party classifier, where the sparse representations over the learned dictionary are viewed as the new feature. The other draws the inference with a non-parameter criteria, the maximum correlation coefficient.

Unlike global learning scheme, the local learning produces a sub-dictionary for each target class [17]. The resulting sub-dictionaries are combined to form a global one. Since the sub-dictionary is learned within certain class, the class identity of generated atoms can be preserved. The decision is then made by seeking which class of atom could generate the minimal reconstruction error. To further enhance the discriminative power, we incorporate the low-rank representation into the generation of dictionary [21]. Although the presented objective is non-convex, it can be circumvented by alternating between the update of sparse representations with the old dictionary, and the renew of dictionary atom with the resulting representations. The flowchart of proposed framework is demonstrated in Fig 1.

Contributions. This paper develops a new framework of dictionary learning for the preceding works, sparse representation of monogenic signal. We learn an over-complete dictionary via sparsity and low-rank regularization. The learned dictionary is then used to build a sparse model, with which the classification can be implemented. Our contributions include:

- the development of dictionary learning via sparsity and low-rank regularization;
- the numeric realization of proposed method via Lasso and low-rank representation;
- the implementation of target classification of SAR image with the learned dictionary;
- the verification of proposed method with several groups of comparative studies.

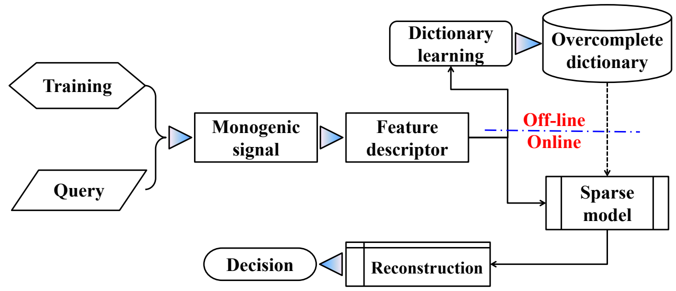


Fig. 1. The block diagram of proposed framework. The monogenic signal is used to characterize the target scattering phenomenology of SAR image. The monogenic features of training sample are pooled together to learn a dictionary. The ℓ_1-norm regularization and low-rank representation are simultaneously utilized to promote the discriminative power. The generated dictionary is then employed to encode the query as a sparse linear combination of their atoms, followed by a reconstruction procedure to draw an inference.

The reminder of this paper is organized as follows. Section II briefly reviewers the related works, including the prototype of sparse model, and sparse representation of monogenic signal. These works lay the solid foundation for the proposed strategy. Section III first summarizes two families of dictionary learning frameworks popularly used, ℓ_0-norm regularization and ℓ_1-norm regularization. The proposed method, sparsity and low-rank regularized dictionary learning strategy is then presented, according to which two specific implementation schemes, global learning and local learning are developed. Multiple comparative studies on MSTAR database are performed in Section IV, with which the effectiveness of the proposed strategy has been verified. Section V concludes this paper, and provides the future research outline.

II. Related Works

This section briefly reviews the related works, including the sparse model, sparse representation of monogenic signal. The details can be found in Ref [4], [7], [8], [22].

A. The Prototype of Sparse Model

Sparse model aims to representing a given signal $\mathbf{y} \in \mathbb{R}^m$ as a linear combination of small number of signals taken from a basis set, namely dictionary. The dictionary contains n prototype signal-atoms, each of which should be unit ℓ_2-norm. Denote by $\Phi = [\phi_1, \phi_2, \ldots, \phi_n] \in \mathbb{R}^{m \times n}$ the dictionary. It is over-complete ($n > m$) if it spans the signal space and its atoms are linearly dependent. In this way, the given signal can be represented as a linear combination of these atoms

$$\mathbf{y} = \phi_1\alpha_1 + \phi_2\alpha_2 + \cdots + \phi_n\alpha_n = \Phi\alpha \quad (1)$$

or be approximated as

$$\mathbf{y} \approx \Phi\alpha, \quad s.t. \quad \|\mathbf{y} - \Phi\alpha\|_p \leq \epsilon \quad (2)$$

where $\alpha \in \mathbb{R}^n$ is the representation over Φ, and ϵ is the allowed error tolerance. In the approximation scheme (2), the typical norms used for measuring the deviation are ℓ_p for $p = 1, 2$ and ∞. This paper mainly focuses on the case of $p = 2$. Since the dictionary Φ is a full-rank matrix ($m < n$), a set of infinite solutions are available for α. To produce the unique one, the feasible set of representation should be limited. This is where the sparsity constraint comes into play. Usually, the solution with the fewest number of nonzero coefficients is certainly an appealing representation. It refers to finding a sparse vector $\hat{\alpha}$ that contains a small number of significant coefficients, while the remaining ones are equal or close to zero. In other words, it aims to minimizing the number of atoms which participate in the representation of signal. The optimal representation is the solution of either the problem

$$(P_0) \qquad \min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \mathbf{y} = \Phi\alpha \tag{3}$$

or

$$(P_{0,\epsilon}) \qquad \min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \|\mathbf{y} - \Phi\alpha\|_2 \leq \epsilon \tag{4}$$

where the pseudo norm $\|\alpha\|_0 := \#\alpha_i, \neq 0$ counts the nonzero entries. Solving (3) or (4) is NP-hard. Fortunately, there are polynomial time approximation algorithms that find a suboptimal solution. These algorithms can be categorized into two classes, greedy algorithm and convex relaxation. The greedy algorithms iteratively search the locally optimal solution with a set of initial basis vectors. Differently, the convex relaxation convert the non-convex problem to the convex one with some approximation skill,

$$(P_1) \qquad \min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \mathbf{y} = \Phi\alpha \tag{5}$$

or

$$(P_{1,\epsilon}) \qquad \min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \|\mathbf{y} - \Phi\alpha\|_2 \leq \epsilon. \tag{6}$$

Given the optimal (or locally optimal) representation $\hat{\alpha}$, the class membership of query is predicted by searching which class of training sample could result in the minimal reconstruction error [22],

$$\min_{k=1,\ldots,K} \|\mathbf{y} - \Phi^{(k)}\hat{\alpha}^{(k)}\|_2^2 \tag{7}$$

where $\Phi^{(k)}$ and $\hat{\alpha}^{(k)}$ denote the atoms associated with the k-th class, and the corresponding weighted coefficients.

B. Sparse Representation of Monogenic Signal

To achieve target recognition with sparse model, G. Dong *et al.* develop a new framework, sparse representation of monogenic signal [7], [8]. Target scattering phenomenology is characterized by the monogenic signal, while the classification is implemented by sparse signal modeling. The monogenic signal is an extension of analytic signal to multi-dimensional space [4]. It is defined as the linear combination of signal $f(\mathbf{z})(\mathbf{z} = [x, y]^T)$ and the Riesz transformed one $f_{\mathcal{R}}(\mathbf{z})$

$$f_M(\mathbf{z}) = f(\mathbf{z}) - (i, j) f_{\mathcal{R}}(\mathbf{z}) \tag{8}$$

where $(i, j, 1)$ formulates a set of orthonormal basis in $\mathbb{R}^3$. The Riesz transform is an isotropic extension of Hilbert transform,

$$\mathcal{R}\{f\}(\mathbf{z}) = f(\mathbf{z}) * h(\mathbf{z}) = \begin{bmatrix} f(\mathbf{z}) * h_x(\mathbf{z}) \\ f(\mathbf{z}) * h_y(\mathbf{z}) \end{bmatrix} \tag{9}$$

where

$$h_x(\mathbf{z}) = -\frac{x}{2\pi\|\mathbf{z}\|^3}$$
$$h_y(\mathbf{z}) = -\frac{y}{2\pi\|\mathbf{z}\|^3}$$

are the first and second-order Riesz kernel. The Riesz transform satisfies the following properties, translation-invariance, scale-invariance, steerability, inner-product preservation [23], [24]. Since the monogenic signal is defined around the Riesz transform, these properties can be preserved. In addition, it inherits the property of analytic signal, such as symmetry, double energy, all-pass transfer function, invariance-equivariance, and hence provides great potential for signal analysis [4].

The practical signal is of finite length with periodic spectra, it is therefore needed to pursue infinite extension with a band-pass filter. The monogenic signal of a given signal is then converted to multiple components of a band-passed signal. The expression can be described as

$$f_M(\mathbf{z}) = (f * h_{bp})(\mathbf{z}) - (i, j) h_{bp} * f_{\mathcal{R}}(\mathbf{z}) \tag{10}$$

where h_{bp} is the band-pass filter. Following the preceding works, log-Gabor filter is employed to produce the multi-resolution monogenic signal. The expression of log-Gabor filter in the Fourier domain is

$$G(\omega) = \exp\Big\{ -\frac{[\log(\omega/\omega_0)]^2}{[\log(\sigma/\omega_0)]^2} \Big\}$$

where ω_0 is the center frequency, and σ is the scaling factor of the bandwidth. The multi-resolution monogenic signal is realized by tuning the scaling factor σ and center frequency ω_0,

$$\sigma_{ratio} = \frac{\sigma}{\omega_0}, \ \omega_0 = \frac{1}{\lambda_{\min}\mu^{s-1}}, \tag{11}$$

where $\lambda_{\min}$ is the minimal wavelength; μ is the multiplicative factor of wavelength; s is the scale index. The ratio σ_{ratio} is usually kept as a constant.

For signal $\mathbf{f}$, the S-scale monogenic signal can be formed by tuning the scale index $s = [1, 2, \ldots, S]$ in (11). The resulting multi-resolution monogenic signal is expressed as

$$\big\{\mathbf{f}_{M,1}, \mathbf{f}_{M,2}, \ldots, \mathbf{f}_{M,S}\big\},$$

with which an augmented monogenic feature can be generated,

$$\mathcal{M}(\mathbf{f}) = \big[vec(\mathbf{f}_{M,1}); vec(\mathbf{f}_{M,2}); \ldots; vec(\mathbf{f}_{M,S})\big]$$

where $vec(\cdot)$ concatenates the components of monogenic signal at a certain scale. The resulting feature is then fed into the framework of sparse signal modeling, where the dictionary are formed by the features of training sample. Denote the mapping of monogenic feature by $\mathcal{M}(\cdot)$. Given a set of training samples

$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, an over-complete dictionary $\mathcal{D}$ is formed by concatenating their augmented monogenic features,

$$\{\mathcal{M}(\mathbf{x}_1), \mathcal{M}(\mathbf{x}_2), \dots, \mathcal{M}(\mathbf{x}_n)\} \Rightarrow \mathcal{D}.$$

The resulting dictionary is then used to encode the counterpart of query (denote by $\mathbf{y}$) as a linear combination of their atoms,

$$\mathcal{M}(\mathbf{y}) = \mathcal{M}(\mathbf{x}_1)\alpha_1 + \mathcal{M}(\mathbf{x}_2)\alpha_2 + \cdots + \mathcal{M}(\mathbf{x}_n)\alpha_n.$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n] \in \mathbb{R}^n$ is the representation over $\mathcal{D}$. The optimal solution is obtained by limiting the feasible set of the representation with sparsity,

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \left\|\mathcal{M}(\mathbf{y}) - \sum_{i=1}^{n} \mathcal{M}(\mathbf{x}_i)\alpha_i\right\|_2 \leq \epsilon$$

Have obtained the optimal solution $\hat{\alpha}$, the decision is made by evaluating which class of atoms could produce the minimization reconstruction error,

$$\min_{k=1,\dots,K} \left\{\|\mathcal{M}(\mathbf{y}) - \Phi^{(k)}\alpha^{(k)}\|_2^2\right\}$$

where $\Phi^{(k)}$ is a sub-dictionary whose atoms are the training samples associated with the k-th class, and $\alpha^{(k)}$ is their corresponding coefficients.

III. The Proposed Dictionary Learning Strategy

The preceding works [7], [8] propose a framework, sparse representation of monogenic signal. The sparse model is built via a predefined dictionary, whose atoms are specified as the training samples directly. This kind of dictionary usually contains multiple correlated versions of atoms, each of which is densely sampled by generic sensors [10]. In addition, it is infeasible to be scaled up to large dataset. To solve this problem and promote the recognition performance, this paper proposes to learn a compact and discriminative dictionary, with which the sparse model can be built. The sparsity and low-rank regularization are employed to generate the basis vectors. We first summarize the methods popularly used for dictionary learning, followed by the proposed strategy, sparsity and low-rank regularization dictionary.

A. The Cost Function

The formulation of dictionary learning and sparse representation assumes that the input signal can be approximated with a linear combination of the dictionary atoms. The objective function typically includes a data fidelity term. It refers to minimizing the reconstruction error in the least square sense. Given a set of observed signals $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$, the objective is to search a dictionary $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_K]$, by which the given signal could be linearly and perfectly represented,

$$\min_{\Phi, \mathbf{X}} \mathcal{R}(\Phi, \mathbf{X}) \tag{12}$$

where

$$\mathcal{R}(\Phi, \mathbf{X}) = \|\mathbf{Y} - \Phi\mathbf{X}\|_F^2 = \min_{\Phi, \mathbf{X}} \sum_{i=1}^{N} \left\{\|\mathbf{y}_i - \Phi\mathbf{x}_i\|_2^2\right\}$$

is the overall reconstruction error, and $\|\cdot\|_F$ is the matrix Frobenius-norm. The resulting coefficient matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ are the representations of $\mathbf{Y}$ over the dictionary Φ. Here, we assume the reconstruction errors satisfy the Gaussian distribution. Other noise style, for example, Laplace distribution $\|\mathbf{y}_i - \Phi\mathbf{x}_i\|_1$ can be also utilized in our framework.

B. ℓ_0-norm regularization

Most of the present dictionary learning methods typically follow the thought of sparsity. It assumes that only a small portion of atoms contribute to the representation of observed signal. An intuitive idea to realize sparsity is to limit the number of nonzero entries with ℓ_0-norm regularization. The representative is the variants of K-Means clustering, namely K-SVD [12]. Different from the conventional clustering algorithms, this family of algorithms allow more than one nonzero entry of representation. Hence, much more flexible degree is introduced during the representation phase.

For given sample $\mathbf{y}_i$, its representation over the learned dictionary $\{\mathbf{x}_i\}_{i=1}^N$ is sparse, *i.e.*, only a small portion of nonzero entries are allowed. This problem is formulated as

$$\min_{\Phi, \mathbf{X}} \mathcal{R}(\Phi, \mathbf{X}) \quad s.t. \quad \forall i, \|\mathbf{x}_i\|_0 \leq T_0 \tag{13}$$

It can be solved by alternating the update of sparse coefficients ($\mathbf{X}$) with the previous dictionary and the renew of atom (Φ) with the representation coefficients.

To improve the performance, several variants of K-SVD have been presented. Q. Zhang and B. Li propose the discriminative K-SVD [25]. They incorporate the classification error into the objective function, and hence unify the dictionary and classifier learning processes. Z. Jiang *et al.* develop the label consistent K-SVD algorithm [26], where the class label information is associated with each dictionary item. Then a single over-complete dictionary and an optimal linear classifier can be jointly generated.

C. ℓ_1-norm regularization

Though ℓ_0-norm regularized algorithm and its variants possess some advantages, the objective function is non-convex and non-differential. Solving the ℓ_0-norm minimization is NP-hard, and hence the convergence is not guaranteed. Convex relaxation is a commonly used strategy. J. Mairal *et al.* propose an online optimization algorithm for dictionary learning via ℓ_1-norm regularization,

$$\min_{\Phi, \mathbf{X}} \mathcal{R}(\Phi, \mathbf{X}) \quad s.t. \quad \forall i, \|\mathbf{x}_i\|_1 \leq T_1 \tag{14}$$

where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ sums the absolute values of entries. An alternative strategy is to solve the unconstrained problem

$$\min_{\Phi, \mathbf{x}_i} \{\|\mathbf{y}_i - \Phi\mathbf{x}_i\|_2^2 + \lambda\|\mathbf{x}_i\|_1\} \tag{15}$$

H. Lee *et al.* empirically prove that the ℓ_1-norm minimization behaves better than the ℓ_0-norm minimization for dictionary learning [27]. The presented problem can be solved by alternating the procedure of sparse coding and dictionary update. The former procedure, sparse coding, can be solved by LAR algorithm [28]. The latter procedure, the renew of dictionary atom, can be achieved with block coordinate descent.

D. Low-rank and sparsity regularized dictionary learning

The recent development on machine learning has demonstrated the advantage of low-rank representation. It has been widely used in many fields. Examples include but are not limited to, background modeling, subspace clustering, biometric recognition, and multimedia analysis [21], [29], [30]. A peculiar property of low-rank representation is its ability to discover the underlying structure in noise signal [31]–[33], and hence inspire us to learn a discriminative dictionary [21], [34]. The objective function of low-rank representation is

$$\min_{\mathbf{X},\mathbf{E}} \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{Y} = \Phi\mathbf{X} + \mathbf{E} \tag{16}$$

where $\mathbf{E}$ is the error matrix, and $\|\cdot\|_*$ is the nuclear norm. S. Roweis and L. Saul claim that the training samples from the same class are linearly correlated and reside in a low dimensional subspace [35]. Hence, it is reasonable to assume that a sub-dictionary learned within a certain class is of low rank. In addition, this constrained term imposed on sub-dictionary could deal with the influence of noise.

To promote the discriminative power, we jointly consider sparsity and low-rank representation into dictionary learning. Sparsity is harnessed to restrict the representation, while low-rank representation is utilized to limit the feasible set of dictionary. The resulting optimization problem is expressed as

$$\min_{\Phi,\mathbf{X}} \mathcal{R}(\Phi,\mathbf{X}) + \lambda_1\mathcal{S}(\mathbf{X}) + \lambda_2\mathcal{D}(\Phi) \tag{17}$$

where $\mathcal{S}(\mathbf{X}) = \left\{\|\mathbf{X}_i\|_1\right\}_{i=1}^K$ measures the sparsity level of representation, and $\mathcal{D}(\Phi) = \sum_i^K \|\Phi_i\|_*$ indicates the matrix rank of dictionary. Although the presented problem (17) is infeasible to be solved directly due to the non-convex objective, it could be circumvented by two separate sub-convex-optimization procedures [27], [36].

1) Update $\{\mathbf{X}_i\}_{i=1}^K$ by fixing dictionary Φ and $\mathbf{X}_j$ $(j \neq i)$: Provided a dictionary[1], the representations $\mathbf{X}$ are updated class by class. The objective function is then reduced to

$$\min_{\mathbf{X}_i} \mathcal{R}(\mathbf{X}_i) + \lambda'\mathcal{S}(\mathbf{X}_i) \tag{18}$$

where $\mathcal{R}(\mathbf{X}_i) = \|\mathbf{Y}_i - \Phi_i\mathbf{X}_i^i\|_F^2 + \|\mathbf{Y}_i - \Phi\mathbf{X}_i\|_F^2$ is the reconstruction error[2], and $\mathcal{S}(\mathbf{X}_i) = \|\mathbf{X}_i\|_1$ measures the sparsity level. The introduction of parameter λ' is to make a tradeoff between fidelity and sparsity. It is actually a typical Lasso problem, and can be solved by LAR [28]. The representations $\mathbf{X}_i$ can be then updated by fixing the old dictionary Φ^{old} and these unrelated representations $\mathbf{X}_j (j \neq i)$.

2) Renew dictionary $\{\Phi_i\}_{i=1}^K$ with the resulting representation $\mathbf{X}$ and $\Phi_j (j \neq i)$: The next task is to renew each sub-dictionary Φ_i with the representation $\mathbf{X}_j$ and the sub-dictionary unrelated $\Phi_j (j \neq i)$. Both the sub-dictionary Φ_i and the sparse representations of $\mathbf{Y}_i$ over Φ_i are renewed simultaneously. It is therefore much more complicated than the preceding step. The objective function can be simplified as

$$\min_{\Phi_i,\mathbf{X}_i^i} \mathcal{R}(\Phi_i,\mathbf{X}_i^i) + \lambda''\mathcal{D}(\Phi_i) \tag{19}$$

where

$$\begin{aligned}\mathcal{R}(\Phi_i,\mathbf{X}_i^i) =& \|\mathbf{Y}_i - \Phi_i\mathbf{X}_i^i - \sum_{j=1,j\neq i}^K \Phi_j\mathbf{X}_i^j\|_F^2 \\ &+ \sum_{j=1,j\neq i}^K \|\Phi_j\mathbf{X}_i^j\|_F^2 + \|\mathbf{Y}_i - \Phi_i\mathbf{X}_i^i\|_F^2\end{aligned}$$

and $\mathcal{D}(\Phi_i) = \|\Phi_i\|_*$. Following the standard form of low rank representation (16), the objective function (19) can be also re-written as

$$\begin{aligned}\min_{\Phi_i,\mathbf{E}_i,\mathbf{X}_i^i} & \|\mathbf{X}_i^i\|_1 + \lambda'\|\Phi_i\|_* + \lambda''\|\mathbf{E}\|_{2,1} + \lambda'''\mathcal{R}(\Phi_i) \\ s.t. \quad & \mathbf{Y}_i = \Phi_i\mathbf{X}_i^i + \mathbf{E}_i\end{aligned} \tag{20}$$

The problem (20) can be then solved with the inexact Augmented Lagrange Multiplier method [31], [37]. When the representations are renewed, the next round of update is pursued. The iteration is terminated if the convergence condition is met.

The procedure of low rank and sparsity regularized dictionary learning:

1) **Initialization:**
Pick a set of element from the training samples, each of which should be unit ℓ_2 energy.
2) **Update the coefficient:**
Generate the representation of training sample over the previous dictionary, *i.e.*, solve the Lasso problem (18).
3) **Renew the dictionary:**
Renew the dictionary atoms based on the resulting coefficients, *i.e.*, search the low-rank representation (19).
4) **Convergence judgment:**
Repeat the preceding two steps until the convergence condition is activated.

E. The Implementation of Target Classification

The previous subsection develops a new framework of dictionary learning. The subsequent task is to pursue target classification with the learned dictionary.

Given a set of training samples from K distinct target classes, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]$, their monogenic features can be generated, $\{\mathcal{M}(\mathbf{y}_1), \mathcal{M}(\mathbf{y}_2), \ldots, \mathcal{M}(\mathbf{y}_n)\}$. Those feature are then used to produce a learned dictionary. To achieve target classification, two specific schemes are proposed.

1) Global Learning: Our first thought is to pool the training samples of all K target classes together, and then learn a single, universal dictionary, over which both the training and the query can be sparsely represented. Given a set of training samples, we compute their augmented monogenic features,

$$\mathcal{M}(\mathbf{Y}) = [\mathcal{M}(\mathbf{y}_1), \mathcal{M}(\mathbf{y}_2), \ldots, \mathcal{M}(\mathbf{y}_n)].$$

[1] A initialized dictionary Φ_0 is needed before the iteration.

[2] $\mathbf{X}_i^j \in \mathbb{R}^{n_j \times n_i}$ is the coefficients of $\mathbf{Y}_i$ over the sub-dictionary Φ_j.

The batch of features $\mathcal{M}(\mathbf{Y})$ is then used to learn a dictionary, where ℓ_1-norm regularization is employed to realize sparsity[3]. The procedure is formulated as the optimization problem

$$\min_{\Phi,\mathbf{X}} \{\|\mathcal{M}(\mathbf{Y}) - \Phi\mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_1\} \tag{21}$$

The learned dictionary Φ is then used to predict the identity of query. Since all training samples are bundled together to learn a single dictionary, the class membership associated with each atom has been inevitably lost. The conventional rule that make the decision according to the minimal reconstruction error could not be applied any more. To draw an inference with the learned dictionary, this paper introduces two rules. One rule resorts to a third-party classifier, the other creates a non-parameteric criteria.

Third-party classifier. The first rule draw the inference with a third-party classifier. Given the learned dictionary, both the training samples and the query can be sparsely represented. The representations of training samples over the learned dictionary are then used to train a third-party classifier, with which the identity of query can be estimated in terms of its representation.

For training sample $\mathbf{y}$, its sparse representation over the learned dictionary is obtained by

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$

Similarly, the representations of all training samples over the learned dictionary can be obtained, $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$, where $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,n_k}]$ is the representations of samples from k-th class, and n_k is the number of training sample of the k-th class. The resulting representations are then used to train a third-party classifier. For a query sample, its representation over the learned dictionary can be generated. We then predict its identity by feeding the generated representation into the trained third-party classifier.

The procedure of decision with a third-party classifier:

- Learn a single dictionary Φ with (21).
- Compute the representations of training samples over the learned dictionary, as shown in (6).
- Train a third-party classifier with this set of representations.
- Produce the representation of query over the dictionary.
- Feed the representation into the trained classifier, and reach an inference.

Maximum correlation coefficient criterion. The alternative rule is non-parameteric. For the samples of the k-th class, their sparse representations over the learned dictionary are $\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,n_k}$. We then produce a specific statistic by these representations, for example, mean

$$\mathbf{x}^{(k)} = mean(\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,n_k}) = \frac{1}{n_k}\sum_{j=1}^{n_k} \mathbf{x}_{k,j} \tag{22}$$

For a query sample, its sparse representation over the learned dictionary is $\hat{\alpha}$. Its correlation coefficient with mean representation of each class can be computed,

$$corr(k) = \hat{\alpha}^T \mathbf{x}^{(k)}, k = 1, 2, \dots, K \tag{23}$$

The decision is reached by searching the maximum correlation coefficient

$$identity(\mathbf{y}) = \arg\max_k corr(k). \tag{24}$$

The procedure of decision with the maximum correlation

- Learned a single dictionary Φ with (21),
- Generate sparse representations of training samples over the learned dictionary,
- Produce the mean representation of each target class with (22), $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)}$,
- Solve the sparse representation of query $\hat{\alpha}$ with (6),
- Compute the correlation coefficient of query with each mean representation $\mathbf{x}^{(k)}$,
- Reach an inference by seeking the maximum correlation coefficient, as defined in (24).

The implementation flow of global learning is demonstrated in Fig. 2, where the maximum correlation coefficient criteria is displayed.

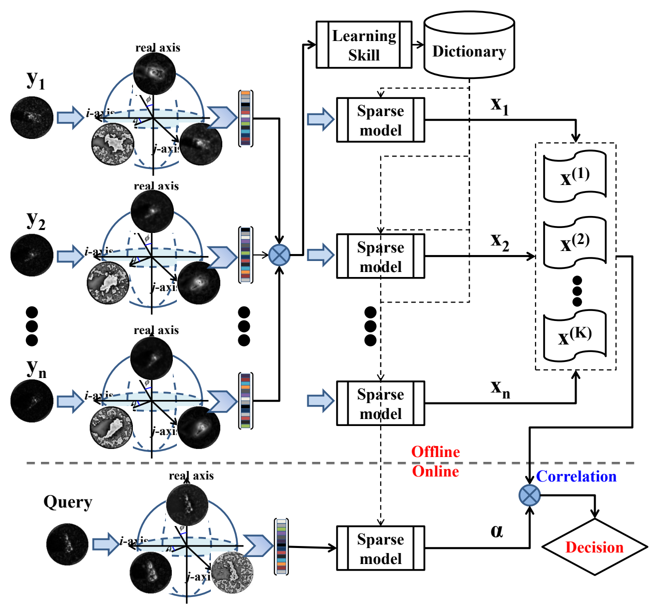


Fig. 2. The implementation flow of global learning. The monogenic features of training sample are bundled together to learn an universal dictionary, over which the training samples and the query can be sparsely represented.

2) Local Learning: Unlike global learning, the local learning scheme generates several class-specific sub-dictionaries. These sub-dictionaries are combined to form a global one. The class label associated with each atom can be then preserved. This strategy has been also studied in the preceding works, namely supervised dictionary learning [19], [20], [38]. Since each sub-dictionary is learned independently, there is no need to pursue repeated training when new target class is available. The decision is reached by evaluating which class of sample could result in the minimal reconstruction error.

[3]In global learning scheme, the low-rank representation is not employed. Only the sparsity constraint is enforced on the representation.

For the k-th target class, suppose there are n_k training samples, $\mathbf{Y}_k = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n_k}]$. We compute their monogenic features, $\mathcal{M}(\mathbf{Y}_k) = [\mathcal{M}(\mathbf{y}_1), \mathcal{M}(\mathbf{y}_2), \ldots, \mathcal{M}(\mathbf{y}_{n_k})]$. It is used to learn a class-specific sub-dictionary,

$$\min_{\Phi_k, \mathbf{X}_k} \|\mathcal{M}(\mathbf{Y}_k) - \Phi_k \mathbf{X}_k\|_F^2 + \lambda_1 \|\mathbf{X}_k\|_1 + \lambda_2 \|\Phi_k\|_* \tag{25}$$

where Φ_k is the sub-dictionary to be learned, and $\mathbf{X}_k$ are the representations. The sub-dictionaries of all K target class are then concatenated to form a global one $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_K]$.

Given a query sample $\mathbf{q}$, its monogenic feature $\mathcal{M}(\mathbf{q})$ can be represented as a linear combination of atoms of Φ,

$$\mathcal{M}(\mathbf{q}) = \Phi\alpha = \sum_{j=1}^{K} \Phi_j \alpha_j \tag{26}$$

where $\alpha_j \in \mathbb{R}^{n_j}$ is the representation with regarding to the j-th sub-dictionary. Following the preceding studies [22], [39], the optimal representation vector $\hat{\alpha}$ is obtained by solving

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \|\mathcal{M}(\mathbf{q}) - \Phi\alpha\|_2 \le \epsilon. \tag{27}$$

The inference is reached by evaluating which class of samples could result in the minimal reconstruction error,

$$\min_{k} \|\mathcal{M}(\mathbf{q}) - \Phi_k \alpha_k\|_2^2. \tag{28}$$

The procedure of classification via local dictionary learning

- Learn a sub-dictionary for each target class with (21),
- Construct a global dictionary by concatenating all class-specific sub-dictionaries,
- Solve the sparse representation of query over the global dictionary,
- Predict the identity by evaluating which class of atom could produce the minimal reconstruction error.

The implementation flow of target classification for local dictionary learning is displayed in Fig. 3.

IV. Experiments and Discussions

The effectiveness of proposed framework is evaluated with multiple sets of experiments on MSTAR database, a standard testbed for SAR image interpretation. The gallery is collected by a 10 GHz SAR sensor with 0.3-meter resolution in range and azimuth. Images are captured at a range of depression angle $\{15°, 17°, 30°, 45°\}$ over $[0, 359°]$ range aspect view. They are of around 128×128 pixels in size, and cropped to 80×80 pixels. The resulting image is used to pursue multi-resolution monogenic signal. The parameters for monogenic signal representation is same to the preceding works [7], [8]. The third-party classifier for global dictionary learning is specified as support vector machine learning (SVM), where the linear kernel function is selected, and the cost function is experimentally fixed as $c = 8$.

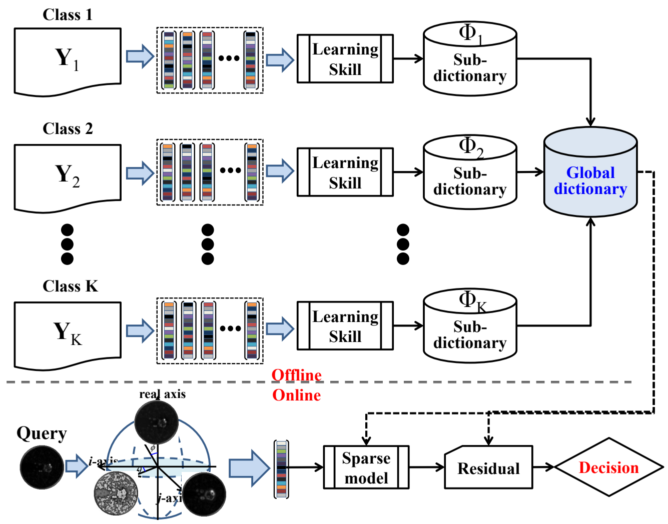


Fig. 3. The implementation flow of local learning. A sub-dictionary is generated for each class. These learned sub-dictionaries are concatenated to form a global one, over which sparse representation of query can be produced.

A. Fundamental Verification

This paper develops a dictionary learning framework for the recent work, sparse representation of monogenic signal. The previously published studies build the sparse model with an over-complete dictionary whose atoms are the training samples directly. Differently, this paper forms sparse model via a learned discriminative dictionary. The ℓ_1-norm minimization is used to limit the feasible set of representation, with which the sparsity can be harnessed. Meanwhile, low-rank representation is employed to confine the dictionary, by which the discriminative power can be promoted. To validate the proposed strategy, a series of fundamental experiments are performed.

Four vehicle target classes, BMP2, T72, BTR60, and T62 are employed to conduct recognition experiments. The number of images used for training and those available for testing are tabulated in TABLE I. The numeric entries are the number of aspect view for different target. The items in bracket are series number, denoting different configuration with structural modifications[4]. The details on dataset will be explained in the subsequent subsection.

TABLE I. The number of aspect views for parameter setting.

Depr.	BMP2	T72	BTR60	T62	***Total***
17°[Training]	**233**(*SN_9563*)	**232**(*SN_132*)	256	299	1020
15°[Testing]	195(*SN_9563*) 196(*SN_9566*) 196 (*SN_c21*)	196(*SN_132*) 195(*SN_812*) 191 (*SN_s7*)	195	273	1637

1) The Representation Model: This paper develops a dictionary learning scheme for the preceding work, sparse representation of monogenic signal. A multi-dimensional generalization

[4] The series number denotes target with a certain configuration. SN_9563, SN_9566, SN_c21 are three configuration variants of T72. Similarly, SN_132, SN_812, SN_s7 are three structural modifications of BMP2.

of analytic signal, the monogenic signal, has been employed to characterize SAR target scattering phenomenology. To test the effectiveness of representation model, a set of comparative studies are pursued, where Haar wavelet transform and Gabor filter bank are compared with the monogenic signal. The experimental results are shown in TABLE II.

TABLE II. THE EFFECTIVENESS OF REPRESENTATION MODEL.

Classifier	**Haar**	**Gabor**	**Monogenic**
SVM	0.7842	0.8925	0.9041
SRC	0.8002	0.9053	0.9230

Two classifier, support vector machine learning (SVM) and sparse representation based classification (SRC), are used to implement target classification. As can be seem from TABLE II, whichever classifier is implemented, the recognition performance obtained using the monogenic feature is consistently better than the reference features, Gabor filter and Haar wavelet. The recognition accuracy obtained using the monogenic feature is 12.28%, 1.77% better than Haar wavelet transform and Gabor filter bank. The result demonstrate the monogenic feature is much more effective than the remaining for characterizing target scattering phenomenology. The results also corroborate the studies in the preceding works [7], [8].

2) The Regularization Term: The present approaches to harness sparsity can be roughly categorized into two classes, ℓ_0-norm regularized learning and ℓ_1-norm one. The representative of ℓ_0-norm regularized learning is the famous K-SVD algorithm [12]. The typical application of ℓ_1-norm regularized learning is the online dictionary learning [17]. To evaluate the effectiveness of regularization term on dictionary learning, a group of experiments are performed. The ℓ_0-norm regularization, and the ℓ_1-norm regularization are compared with the proposed strategy. Local learning scheme is employed to implement target classification. The experimental results are given in TABLE III.

TABLE III. THE COMPARISON OF REGULARIZATION TERM.

Regularization	ℓ_0	ℓ_1	$\ell_1 + LR$
Accuracy	0.9278	0.9304	0.9392

The recognition accuracy listed in TABLE III demonstrate that the performance obtained using ℓ_1-norm regularization dictionary learning is better than ℓ_0-norm regularization dictionary learning. This conclusion has also been verified in Lee's work [27]. The proposed sparsity and low-rank regularization dictionary learning achieves the best performance. Its recognition accuracy is 0.88% better than the ℓ_1-norm regularization, and 1.14% better than the ℓ_0-norm regularization.

3) The Learning Manner: To achieve target classification, this paper introduces two specific schemes, global learning (unsupervised) and local learning (supervised). The subsequent experiments devote to the comparison of two presented schemes. We verify the performance from two perspective, recognition accuracy and computational efficiency. The size of learned dictionary is set as $k=\{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ times of the original set. Specifically, the number of atom in the learned dictionary is $round(n \cdot k)$, where n is the total number of training sample. The experimental result is shown in TABLE IV.

TABLE IV. THE STUDY ON LEARNING MANNER.

(a) Global learning with a third-party classifier (SVM).

Size	408	510	612	714	816	918
Accuracy	0.9194	0.9206	0.9249	0.9261	0.9261	0.9304
CPU-Time	118s	157s	195s	240s	292s	336s

(b) Global learning with the maximum correlation decision.

Size	408	510	612	714	816	918
Accuracy	0.9096	0.9151	0.9200	0.9236	0.9297	0.9304
CPU-Time	114s	150s	197s	236s	281s	335s

(c) Local learning with minimal residual decision.

Size	408	511	612	713	816	918
Accuracy	0.9230	0.9304	0.9340	0.9389	0.9426	0.9407
CPU-Time	115s	134s	145s	161s	193s	201s

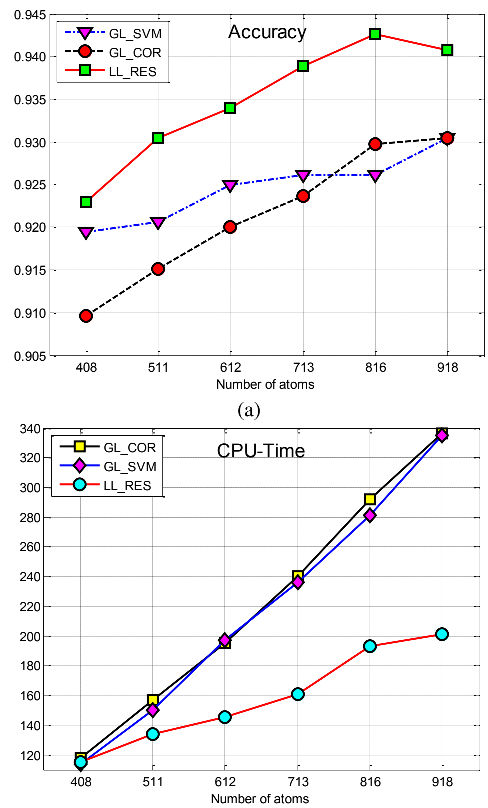


Fig. 4. The comparison of global learning and local learning in terms of (a) recognition accuracy; (b) CPU time. GL_COR and CL_SVM denote two global learning schemes, and LL_RES is the local dictionary learning method.

The variations of recognition accuracy across the size of learned dictionary are drawn in Fig. 4 (a), while the changes of computational times are plotted in Fig. 4 (b). The recognition rates listed in TABLE IV and the curves drawn in Fig. 4 illustrate that the performance obtained using local learning is slightly better than global learning. First, the recognition accuracy of local learning is always higher than global learning if the size of learned dictionary is tuned from 0.4 to 0.9 time of original set. This result can be attributed to the introduction of low-rank regularization, with which the discriminative power of dictionary is promoted. Second, the computational consumption of local learning is much less than global learning. However, it should be noted that most of computational cost is consumed in the off-line dictionary learning phase. The consumption of online prediction phase is similar for these two schemes. Although the performance of global learning is not as good as the local learning, it is needed to demonstrate the advantage of low-rank representation.

The curves drawn in Fig. 4 demonstrate that the decision with a third-party classifier slightly outperforms the maximum correlation coefficient inference. This phenomenon can be explained by the advantage of classifier. The former rule trains a linear model with all sparse representations, and draws the inference with the maximum margin bound criterion. Contrarily, the latter represent each class with the mean representation, and reach the decision by seeking the maximum correlation.

4) Illustrative Example: To delineate the sparse representation over the training samples and the learned dictionary, a set of examples are provided in Fig. 5. The query sample is selected from T72. The top row displays the result of sparse representation over the raw training samples, while the bottom row gives the sparse model formed by the learned dictionary. The total number of images available for training is 1020, compared to 765-atom learned dictionary. Though the representation over the training samples is much more sparse than the learned dictionary, it predicts the error identity. The sparse model formed by the training samples mis-classifies the query as T62. Contrarily, the learned dictionary reaches the correct decision, T72. This set of experiment prove that the discriminative power is enhanced with the proposed method.

B. Performance Comparison

The previous experiments devote to the fundamental verification of proposed strategy. Next, multiple comparative studies are pursued. The proposed strategies are compared with the baseline algorithms in terms of recognition accuracy. The size of learned dictionary is set as the 0.75 times of original set. The methods to be studied are summarized in TABLE V. KSVD and ODL are implemented in the framework of local dictionary learning. M. Yang *et al.* develop a Fisher discrimination criteria for supervised dictionary learning [15]. I. Ramirez *et al.* recommend a structured incoherence constraint between class-specific dictionaries [19]. H. Wang *et al.* introduce a improved version of [19], where similarity limitation is incorporated into [20]. Hence, these works can be viewed as the representative local dictionary learning. The methods to be compared with are tabulated in TABLE V, where the 'Learned' item refers to sparse model formed over training samples directly or the learned dictionary.

1) Target Recognition under Configuration Variation: We first evaluate target recognition under different configurations. The configuration refers to physical difference and small structural modifications. The representative of T72 main-battle tank are shown in Fig. 6. Some exemplars are listed in the following items.

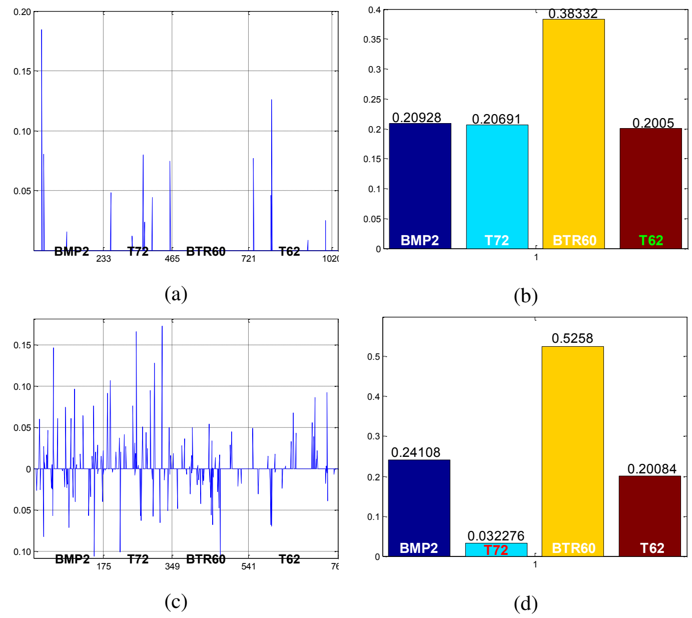


Fig. 5. The comparison of sparse model formed by pre-defined samples and learned dictionary. (a) the representation over the pre-defined dictionary, (b) the reconstruction error resulting from (a), (c) the representation over the learned dictionary, (d) the reconstruction error resulting from (c).

TABLE V. THE METHODS TO BE COMPARED WITH.

Abbre.	Full name (description)	*Ref.*	Learned (Y/N)
SRC	Sparse representation-based classifier	[22]	N
MSRC	Sparse representation of monogenic signal	[8]	N
MSRF	Score-level fusion of sparse representation	[7]	N
KSVD	K-SVD dictionary learning	[12]	Y
LCKSVD	Label consistent K-SVD learning	[26]	Y
ODL	Online dictionary learning	[17]	Y
FDDL	Fisher Discrimination Dictionary Learning	[15]	Y
DLSI	Dictionary Learning with Structured Incoherence	[19]	Y
LCSI	Local-Constrained Structured Incoherence	[20]	Y
LRSR	Low-Rank and Sparsity Regularized DL	III-E2	Y

- **Version Variant:** *Smoke Grenade Launchers, Side Skirts*
- **Configuration Variant:** *Two Cables, Fuel Barrels*
- **Structural Modifications:** *Dented Fenders, Broken Antenna Mount*

Fig. 6. Exemplar on configuration variation (T72 tank).

Four vehicle targets, BMP2, T72, BTR60, and T62 have been employed, among which BMP2 and T72 have several configuration variants. The details can be found in TABLE I. Different to the previous experiments, the configurations used for training, SN_9563 (BMP2) and SN_132 (T72) are no longer available for testing. The algorithms are trained using a certain configuration, and tested with another different ones.

The experimental results are shown in Fig. 7.

	MSRC [8]⇒0.8747				**KSVD** [12]⇒0.9007			
	TG1	TG2	TG3	TG4	TG1	TG2	TG3	TG4
TG1	0.9286	0.0230	0.0255	0.0230	0.9337	0.0434	0.0128	0.0102
TG2	0.0415	0.6736	0.0259	0.2591	0.0389	0.7565	0.0078	0.1969
TG3	0.0	0.0	0.9897	0.0103	0.0051	0.0051	0.9846	0.0051
TG4	0.0	0.0	0.0	1.0	0.0	0.0037	0.0	0.9963
	LCKSVD [26]⇒0.8536				**ODL** [17]⇒0.9133			
	TG1	TG2	TG3	TG4	TG1	TG2	TG3	TG4
TG1	0.8801	0.0587	0.0408	0.0204	0.9209	0.0510	0.0153	0.0128
TG2	0.0907	0.6917	0.0052	0.2124	0.0155	0.8161	0.0052	0.1632
TG3	0.0308	0.0051	0.9538	0.0103	0.0103	0.0154	0.9692	0.0051
TG4	0.0	0.0147	0.0	0.9853	0.0	0.0	0.0	1.0
	DLSI [19]⇒0.8868				**LCSI** [20]⇒0.8900			
	TG1	TG2	TG3	TG4	TG1	TG2	TG3	TG4
TG1	0.9388	0.0102	0.0255	0.0255	0.9388	0.0128	0.0306	0.0179
TG2	0.0492	0.7073	0.0026	0.2409	0.0570	0.7124	0.0026	0.2280
TG3	0.0103	0.0	0.9897	0.0	0.0103	0.0	0.9897	0.0
TG4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
	FDDL [15]⇒0.8740				**LRSR**⇒0.9278			
	TG1	TG2	TG3	TG4	TG1	TG2	TG3	TG4
TG1	0.8827	0.0842	0.0179	0.0153	0.9413	0.0357	0.0128	0.0102
TG2	0.0181	0.7358	0.0	0.2461	0.0155	0.8368	0.0052	0.1425
TG3	0.0154	0.0051	0.9538	0.0256	0.0051	0.0103	0.9795	0.0051
TG4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0

Fig. 7. The confusion matrices of target recognition on configuration variations. TG1, TG2, TG3, and TG4 denote BMP2, T72, BTR60, and T62.

Obviously, BMP2 and T72 are much more difficult to be recognized than T62 and BTR60. Two versions of main-battle tanks, T72 and T62 are hard to be distinguished, and two kinds of armored personnel carriers, BMP2 and BTR60, are easy to be confused. Even 25.91% samples of T72 are misclassified as T62 by MSRC. Meanwhile, 4.08% samples of BMP2 are mis-recognized as BTR60 by LCKSVD. This is because both the configuration and depression angle are drastically different between the images used for training and those available for testing. The recognition rates for DLSI, LCSI, and FDDL are bellow 0.9, more inferior than the remaining dictionary schemes. The results can be attributed to their learning criterion (Fisher discrimination and structured incoherence). They may be inefficient to deal with the scattering phenomenology resulting from configuration variation.

The recognition accuracies are 0.9278 for our proposed method, compared to 0.9007 for K-SVD, and 0.8536 for LCKSVD, 1.45%, 7.42%, 2.71% better than these competitors.

2) Target Recognition under Articulation and Occlusion: Target articulation and occlusion, for example, hatch open and close, turret straight and rotation, are of universal phenomenon in the real battlefields. A set of examples are shown in Fig. 8. Hence, target recognition on articulation and occlusion is important to real-world applications.

To evaluate the performance of target recognition on articulation and occlusion, a set of experiments are conducted. Three military targets, 2S1, BRDM2, and ZSU23/4 are employed,

Fig. 8. Exemplar on articulation and occlusion (ZSU23/4 antiaircraft gun).

among which BRDM2 and ZSU23/4 have some articulated variants. The standards are used to train the algorithms, while the variants are specified as the query. The details are listed in TABLE VI. The number of aspect view for articulated and occluded variants are in bracket. The experimental result is given in Figure 9.

TABLE VI. ASPECT VIEWS FOR ARTICULATION AND OCCLUSION.

	2S1	BRDM2	ZSU23	*Total*
Training (17°)	**299**	**298**	**299**	**896**
Testing (45°)	303	303(120)	303(119)	1148

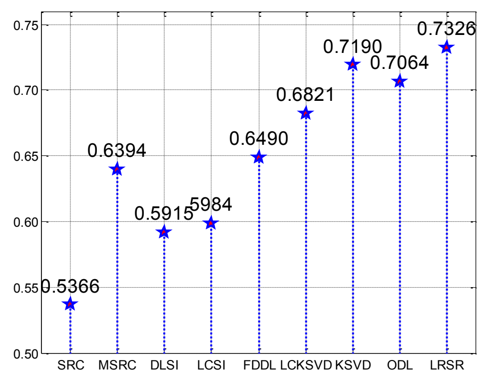


Fig. 9. The performance of target recognition on articulation and occlusion.

The performance of this set of experiments are much more inferior than the previous one. The best recognition rate is 0.7326, while the wost performance is only 0.5366. The result can be attributed to the rigorous experimental settings. The query samples are collected at an operating condition of 45° depression angle, while the training samples are taken at an operating condition of 17° depression angle. A drastic change of 28° from 17° to 45° exists between the images used for training and those available for testing. In addition, two military vehicles, BRDM2 and ZSU23/4 still have the articulated and occluded variants. Images of articulated variants are 120 and 119 for BRDM_2 and ZSU23/4. The settings are much more difficult than the previous. The proposed method achieves the best performance. The recognition accuracy is 0.7326 for our proposed method, 2.62%, 5.05%, 1.46% better than the main competitors, ODL, LCKSVD, and K-SVD. The performance obtained using FDDL, DLSI, and LCSI (0.6490, 0.5915, and 0.5984) is much more inferior than the remaining learning algorithms.

C. Target Recognition of Ten-Classes

To further evaluate the performance of target recognition, a set of experiments have been performed. All of 10-class vehicle targets in MSTAR database have been verified. Some exemplars are shown in Fig. 10. The details on aspect view images are listed in TABLE VII. Images collected at a depression angle of 17-degree are used for training, while the ones captured at a depression angle of 15-degree are available for testing. For two multi-configuration targets, BMP2 and T72, the division is same to Section IV-B1, *i.e.*, SN_9563 and SN_132 taken at a depression angle of 17-degree are used to train the algorithm, while the remaining collected at a depression angle of 15-degree are available for testing.

Fig. 10. Illustration of 10-class target.

TABLE VII. ASPECT VIEWS FOR 10-CLASS TARGET RECOGNITION.

	BMP2	BTR70	T72	BTR60	2S1	BRDM	D7	T62	ZIL	ZSU	*Total*
17°	**233**	**233**	**232**	**256**	**299**	**298**	**299**	**299**	**299**	**299**	**2747**
15°	392	196	386	195	274	274	274	273	274	274	2812

The confusion matrices as well as the overall recognition accuracy are displayed in Fig. 11. In each sub-figure, the diagonal entry lists the recognition accuracy of each target class, while the non-diagonal one provides the mis-recognition rate. The overall recognition rates are arranged at the top in bracket.

The experimental results are similar to the one shown in Section IV-B1. The recognition accuracies are 0.6862, 0.7577, 0.7423, 0.8393, 0.8265, and 0.8878 for BMP2. Similarly, the recognition rates are 0.4974, 0.5259, 0.5440, 0.7332, 0.7383, and 0.7824 for T72. The accuracies are much lower than the remaining. This is because both the configuration and depression angle are significantly different between the image used for training and those available for testing. The classification rates generated by the proposed method, 0.8878 and 0.7824, consistently better than the competitors, FDDL, DLSI, LCSI, KSVD, and ODL. The overall recognition accuracy is 0.9495 for our proposed method, 6.27%, 5.90%, 6.05%, 1.71%, and 2.21% better than the baseline methods, FDDL, DLSI, LCSI, KSVD, and ODL. The results fully prove the advantage of low-rank and sparsity regularized dictionary.

V. Conclusion

This paper introduces a dictionary learning strategy for the preceding study, sparse representation of monogenic signal. The proposed strategy is applied to target recognition in SAR image. Unlike the preceding works, where the sparse model is formed by a dictionary whose atoms are the training samples directly, this paper proposes to learn a discriminative dictionary. The main contribution consists in the development of dictionary learning via low-rank and sparsity regularization. Two specific implementation schemes have been developed to achieve target classification. They demonstrate the advantage of sparsity and low-rank representation respectively. To verify the proposed strategy, we pursue multiple comparative studies, from which the conclusions can be come.

- The discriminative power of the learned dictionary is promoted by limiting the feasible set of dictionary with low-rank representation.
- The proposed strategy could deal with the configuration variation, as well as target articulation and occlusion.
- Though the sparsity can be realized via ℓ_0-norm minimization, the convergence is not guaranteed. Its performance is therefore inferior to ℓ_1-norm minimization.
- The proposed strategy achieve state-of-the-art performance, consistently exceeding the competitors.

The good performance shown by proposed method is the result of coupling monogenic signal representation with the dictionary learning. It is inspired by a recently developed algorithm, online dictionary learning, with an additional regularization term, low-rank representation incorporated into. An intriguing question for future work is to learn a task-specific dictionary. Another research focus is to adopt dictionary learning in Reproducing Kernel Hilbert Space. Though some preliminary trials have been pursued in our preceding work [7], the full potential in target recognition has been yet uncovered.

Acknowledgment

The authors would like to thank the handling Associate Editor and the anonymous reviewers for their constructive comments and suggestions for this paper. Ganggang Dong would like to thank Julien Mairal for sharing the source code of sparse modeling. The similar thank is paid to Tiep Vu for his help on low rank representation.

References

[1] L. Novak, G. Owirka, and C. Netishen, "Performance of a high-resolution polarimetric SAR automatic target recognition system," *Lincoln Lab J.*, vol. 6, no. 1, pp. 11–23, 1993.

[2] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions-a tutorial," in *Algorithms for Synthetic Aperture Radar Imagery III*. SPIE, Jun. 1996, pp. 228–242.

[3] T. D. Ross, J. J. Bradley, L. J. Hudson, and M. P. O'Connor, "SAR ATR: so what's the problem? an MSTAR perspective," in *Algorithms for Synthetic Aperture Radar Imagery VI*. SPIE, Aug. 1999.

[4] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, 2001.

[5] ——, "The monogenic scale-space: A unifying approach to phase-based image processing in scale space," *J. Math. Imag. Vis.*, vol. 21, no. 1, pp. 5–26, 2004.

[6] M. Felsberg, R. Duits, and L. Florack, "The monogenic scale space on a rectangular domain and its features," *Int'l J. Comput. Vis.*, vol. 64, no. 2, pp. 187–201, 2005.

[7] G. Dong and G. Kuang, "Classification on the monogenic scale space: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2527–2539, Aug. 2015.

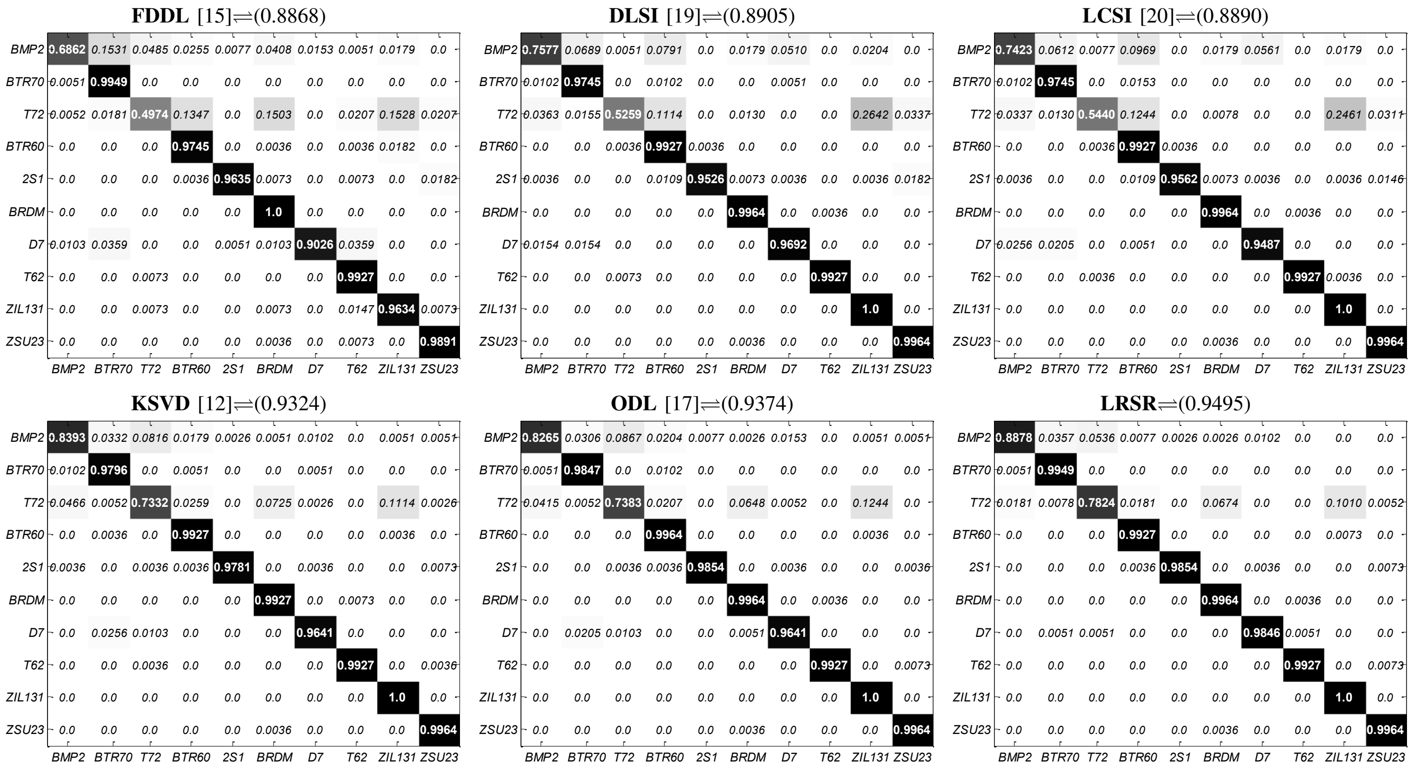


Fig. 11. Confusion matrices of 10-class target recognition. The vertical axis provides the ground-truth, while the horizontal axis shows the predicted class.

[8] G. Dong, N. Wang, and G. Kuang, "Sparse representation of monogenic signal: with application to target recognition in SAR images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 952–956, Aug. 2014.

[9] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[10] I. T. and P. Frossard, "Dictionary learning, what is the right representation for my signal?" *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[11] K. Kreutz-Delgado, J. Murray, B. Rao, K. Engan, T. Lee, and T. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, no. 2, pp. 349–396, Feb. 2003.

[12] 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. 4311–4321, Nov. 2006.

[13] G. Monaci, P. Jost, P. Vandergheynst, BorisMailh, S. Lesage, and R. Gribonval, "Learning multimodal dictionaries," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2272–2283, Sep. 2007.

[14] I. T. and P. Frossard, "Dictionary learning for stereo image representation," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 921–934, Apr. 2011.

[15] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int'l Conf. Comput. Vis. (ICCV)*, 2011, pp. 543–550.

[16] ——, "Sparse representation based fisher discrimination dictionary learning for image classification," *Int'l J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, 2014.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26^{th} International Conference on Machine Learning*, Jun. 2009, pp. 689–696.

[18] ——, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, Jan. 2010.

[19] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3501–3508.

[20] H. Wang, C. Yuan, W. Hu, and C. Sun, "Supervised class-specific dictionary learning for sparse modeling in action recognition," *Pattern Recognition*, vol. 45, no. 11, pp. 3902–3911, Nov. 2012.

[21] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2586–2593.

[22] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[23] M. Unser, D. Sage, and D. Ville, "Multiresolution monogenic signal analysis using the Riesz-Laplace wavelet transform," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2042–2058, Nov. 2009.

[24] G. Dong, G. Kuang, N. Wang, and W. Wang, "Classification via sparse representation of steerable wavelet frames on Grassmann manifold: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2892–2904, Jun. 2017.

[25] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 2691–2698.

[26] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1697–1704.

[27] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 801–808.

[28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[29] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2439–2446.

[30] C. Zhang, J. Liu, QiTian, C. Xu, H. Lu, and S. Ma, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1673–1681.

[31] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2080–2088.

[32] R. S. Cabral, F. D. la Torre, J. P. Costeira, and A. Bernardino, "Matrix completion for multi-label image classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 190–198.

[33] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image and Vision Computing*, vol. 32, pp. 814–823, 2014.

[34] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2618–2625.

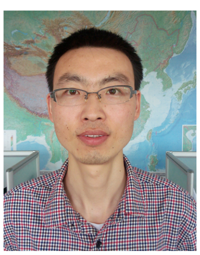
[35] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality analysis by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[36] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[37] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv:1009.5055*, 2010.

[38] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Supervised dictionary learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1033–1040.

[39] A. Tropp and J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.



Ganggang Dong Ganggang Dong received the B.S. degree in UAV application engineering from the Academy of Artillery, Hefei, China, in 2004, and M.S. and Ph.D. degrees in information and communication engineering from National University of Defense Technology, Changsha, China, in 2012 and 2016, respectively. He has authored 14 scientific papers in peer-reviewed journals, including IEEE TIP, IEEE TGRS, IEEE JSTARS, IEEE GRSL, IEEE SPL, and SPIE JARS. His research interests include the applications of compressed sensing and sparse representations, and manifolds learning.

Na Wang received the B.S. degree and M.A. Eng. degree in automated command and control, information and communciation engineering from the Air Force Engineering University, Xian, China, in 2005, 2008, and the Ph. D. degree in information and communciation engineering from the National University of Defense Technology, Changsha, China, in 2012.

Her research interests include the polarmetric SAR image interpretation, automatic target detection.

Gangyao Kuang(M11) received the B.S. and M.S. degrees from the Central South University of Technology, Changsha, China, in 1998 and 1991, respectively, and the Ph.D. degree from the National University of Defense Technology, Changsha, in 1995. He is currently a Professor and Director of the Remote Sensing Information Processing Laboratory in the School of Electronic Science and Engineering, National University of Defense Technology. His current interests mainly include remote sensing, SAR image processing, change detection, SAR ground moving target indication, and classification with polarimetric SAR images.

Hongbing Qiu received the B.S., M.S. and Ph.D. degree in radio communication, communication and electronics system, information and communication engineer from Xidian University, Xi'an, China, in 1984, 1991, and 2004. He is Professor and Dean of School of Information and Communication at Gulin University of Electronic Technology currently.

His research interests include mobile communication, Ultra WideBand wireless communication, wide band communication network and communication signal processing.