# SAR Target Recognition via Joint Sparse Representation of Monogenic Signal

Ganggang Dong, Gangyao Kuang, *Member, IEEE*, Na Wang, Linjun Zhao, and Jun Lu

***Abstract*—In this paper, the classification via sparse representation and multi-task learning is presented for target recognition in SAR image. To capture the characteristics of SAR image, a multi-dimensional generalization of the analytic signal, namely the monogenic signal is employed. The original signal can be then orthogonally decomposed into three components, local amplitude, local phase and local orientation. Since the components represent the different kinds of information, it is beneficial by jointly considering them in a unifying framework. However, these components are infeasible to be directly utilized due to the high-dimension and redundancy. To solve the problem, an intuitive idea is to define an augmented feature vector by concatenating the components. This strategy usually produces some information loss. To cover the shortage, this paper considers three components into different learning tasks, in which some common information can be shared. Specifically, the component-specific feature descriptor for each monogenic component is produced first. Inspired by the recent success of multi-task learning, the resulting features are then fed into a joint sparse representation model to exploit the inter-correlation among multiple tasks. The inference is reached in terms of the total reconstruction error accumulated from all tasks. The novelty of this paper includes (i) the development of three component-specific feature descriptors; (ii) the introduction of multi-task learning into sparse representation model; (iii) the numerical implementation of proposed method; (iv) extensive comparative experimental studies on MSTAR SAR dataset, including target recognition under standard operating conditions, as well as extended operating conditions, and the capability of outliers rejection.**

***Index Terms*—The monogenic signal, sparse representation, multi-task learning, convex optimization, joint covariate selection, SAR target recognition.**

## I. Introduction

RECENTLY, the monogenic signal has been introduced by Felsberg and Sommer [1]. It is defined as a linear combination of signal and its Riesz transformed one. As a multi-dimensional extension of the analytic signal, the monogenic signal could orthogonally decompose the original signal into three components, local amplitude, local phase, and local orientation. This decoupling strategy has made it possible to derive effective solutions to a number of image-processing problems, in particular when the traditional pixel intensity can not be regarded as a reliable feature. Moreover, it is able to capture the broad spectral information with maximal spatial localization. To consider scale-space theory and phase-based signal processing technique in a unifying framework, a linear scale-space named monogenic scale-space is proposed [2], [3], where the Poisson kernel is utilized instead of the Gaussian one. Since the Riesz transform itself yields the flux of the Poisson scale-space and the combination of flux and scale-space, the monogenic scale-space is capable to provide the local features phase-vector and attenuation in scale-space [2]. Thus, the monogenic signal produces deep insight to low-level image processing. In [2], edge detection is realized by considering the relation between phase congruency and local amplitude. In [4], a novel rotational invariant approach for analyzing i(ntrinsic)1D and i2D local features of 2D signal (image) without the use of heuristics is presented. In [5], the color monogenoic signal is introduced by the formalism of Clifford algebras. In [6], the analytic wavelet transform is extended to 2D monogenic wavelet transform. In [7], the monogenic signal is transposed to the wavelet domain by considering a complexified version of the Riesz transform which is capable to map a real-valued wavelet basis into a complex one. To form a general continuous-domain framework for designing steerable, reversible signal transformations in multi-dimensions, a self-reversible, $N$th-order extension of the Riesz transform is introduced by Unser and Ville [8]. Very recently, the monogenic signal has also been introduced to pattern recognition. In [9], two coding schemes, based on the images's responses to the 1st-order and 2nd-order Riesz transforms are presented for biometric recognition. In [10], three complementary components generated by the monogenic signal, local amplitude, local phase, and local orientation are binarized by three different schemes. In [11], local amplitude, local phase and local orientation generated from the monogenic signal are encoded to capture the texture and motion information.

Recently, the problem of recovering sparse linear representation of a query sample in terms of a redundant dictionary (a set of reference samples) has received great interests in signal processing literature. In [12], it claims that the spatial receptive fields of simple cell in mammalian striate cortex can be accounted for with respect to a strategy for producing a sparse distribution of input activity in response to natural images. Ref [13] considers decomposing an image in terms of a library of templates. It suggests to generate a sparse (or compact) representation for the image, and to account for occlusions and noise in the image by a weighted $\ell_p$ norm with $0 < p < 1$. Ref [14] proposes the application of sparse representation (factorization) of signals over a set of basis (overcomplete dictionary) for signal classification. Ref [15] presents an efficient sparse coding algorithm (the feature-sign search algorithm) by iteratively solving $\ell_1$-regularized least

This works was supported by the National Natural Science Foundation of China under Grant 61201338 and 61401477.

The authors are with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China, 410073 (e-mail: dongganggang@nudt.edu.cn)

squares problem and $\ell_2$-constrained least squares problem. The presented method is numerically achieved by optimizing an object function that includes the fidelity term and the sparsity level. Ref [16] presents a novel approach to speech recognition, where a word is sparsely represented in an over-complete basis of clear exemplar speech signals using only the uncorrupted time-frequency elements of the word. Ref [17] proposes a methodology for advanced image formation in wide-angle SAR, in which a heuristic, graph-structured, sparse signal representation algorithm for overcomplete dictionaries is presented. Ref [18] casts the recognition problem as one of classifying among multiple linear regression models and address it by optimizing $\ell_1$-norm minimization. Ref [19] focuses on the dictionary learning technique in sparse representation modeling. They summarize two approaches to the generation of overcomplete dictionary: i) forming a dictionary based on a mathematical model of the data, or ii) learning a dictionary to perform best on a training set. Ref [20] reviews the role of sparse representation model in image processing, its rationale and models related to it. Ref [21] reviews a few representative applications of how the interaction between sparse signal representation and computer vision can enrich both fields. Ref [22] utilizes the joint sparse representation model for multi-view target recognition, in which the multiple views of a single target share the similar sparsity pattern.

To improve the performance for SAR target recognition, the preceding works [23]–[25] exploit the monogenic signal to capture the characteristics of SAR image, and draw the inference by sparse signal modeling. Ref [23] defines an augmented monogenic feature via uniformly downsampling, normalization, and concatenation of the monogenic components, and feeds the resulting feature into the framework of sparse representation model. Ref [24] generates a region covariance matrix descriptor via the monogenic components at different scales, and achieves target recognition by considering the feature descriptor on Riemannian manifolds rather than an Euclidean vector space. Actually, sparse signal model can be cast into a problem of sparse covariate selection via linear regression, where the training samples play the role of observation of covariate, and the query sample is viewed as the response. Motivated by the recent success of multi-task learning [26]–[28], this paper investigates the problem of multi-task joint sparse representation model under the monogenic scale-space. First, three component-specific feature descriptors are produced using local amplitude, local phase, and local orientation of the monogenic signal at different scale-spaces. Different from the conventional methods, where multiple components are considered into a single learning task, this paper produces different linear representation models for different monogenic components (*i.e.*, one task per feature). It searches a very few common classes of training samples that are most correlated to the query across related representation models. Since different tasks may favor different representation coefficients, the joint sparse representation model could provide additional information to target recognition, and hence the robustness towards coefficient estimation can be enhanced. The proposed method is realized by a constraint of joint sparsity term on the representations [29], [30]. The optimal representations can be estimated by optimizing $\ell_1/\ell_q$-norm regularized problem, and the decision is made in favor of the class with the lowest total reconstruction error accumulated from all tasks. The working mechanism of the proposed method is pictorially shown in Fig. 1

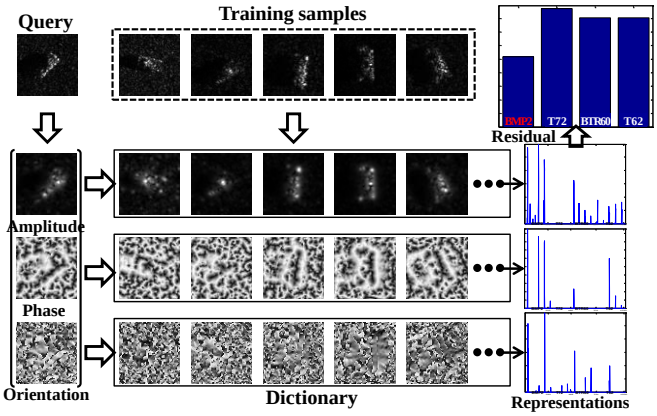



Fig. 1. Illustration of target recognition via joint sparse representation of monogenic signal. For a query sample, three features can be produced by the components of monogenic signal at various scale-spaces. For each feature, it is represented as a linear combination of the counterpart of training samples in a joint sparse constraint way across all features. The decision is reached by seeking the class with the lowest total reconstruction error.

**Contributions** As described above, this paper introduces a classification method via multi-task learning and sparse signal modeling. The components of monogenic signal at different scales are used to produce three component-specific features. The resulting features are then employed to form three linear regression models, where the features of the training samples are regarded as the regressors (or observations), and the one of the query are viewed as the corresponding response. Unlike the conventional works, this paper produces the representation coefficient vectors for these linear regression models with similar sparsity entries via multi-task learning. The inference is reached in terms of the total reconstruction error. Our contribution are therefore four-fold:

- We produce the discriminative features for SAR target recognition by the components of the monogenic signal.
- We propose a tri-task learning system for three monogenic features based linear regression models. Thus, it is able to produce three representation coefficient vectors with similar sparsity elements.
- We perform extensive comparative experiments under standard experimental setup, as well as non-literal experimental setup.
- We verify the performance under limited training resource, as well as the capability of outliers rejection.

The remainder of this paper is organized as follows. The sparse representation-based classification is reviewed in Section II. The proposed joint sparse representation of monogenic signal and classification is shown in Section III. The applications of the proposed method to SAR target recognition are given in Section IV. Section V concludes this paper finally.

## II. Sparse Representation

Over the years, a resurgent development of sparse signal representation has been witnessed. The representative applications include image restoration [31], pattern recognition [18], [32], image classification [33], [34]. The success of sparse signal modeling mainly consists in the fact that a certain signal has sparse representation with respect to a set of basis (dictionary). The optimal representation can be efficiently and provably retrieved by convex optimization strategy [21], [35].

Given sufficient training samples of the $k$-th ($k = 1, ..., K$) class, $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, ..., \mathbf{x}_{k,n_k}] \in \mathbb{R}^{m \times n_k}$, where the samples are arranged as the column vectors in $\mathbb{R}^m$, the new (query) sample from the same class $\mathbf{y} \in \mathbb{R}^m$ will approximately lie in the linear span of training samples associated with the $k$-th class,

$$\mathbf{y} = \mathbf{x}_{k,1}\alpha_{k,1} + \mathbf{x}_{k,2}\alpha_{k,2} + \cdots + \mathbf{x}_{k,n_k}\alpha_{k,n_k} \tag{1}$$

for scalar coefficients $\alpha_k = [\alpha_{k,1}, \alpha_{k,2}, ..., \alpha_{k,n_k}]^T \in \mathbb{R}^{n_k}$. Here, it is assumed that samples from a single class approximately span a linear subspace.

Since the class membership of the query sample is unknown initially, it is represented with respect to the whole training set, $\mathbf{X} = \left[\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\right] \in \mathbb{R}^{m \times n}$, where $n = \sum_{k=1}^{K} n_k$ is the total number of training samples. Then, the linear representation of $\mathbf{y}$ can be rewritten as as

$$\mathbf{y} = \mathbf{X}_1\alpha_1 + \mathbf{X}_2\alpha_2 + \cdots + \mathbf{X}_K\alpha_K = \mathbf{X}\alpha, \tag{2}$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_K]^T \in \mathbb{R}^n$ is the representation vector. Theoretically, the optimal representation $\alpha$ is a coefficient vector whose entries are zero except those associated with the $k$-th class.

To retrieve the representation $\alpha$, it is tantamount to solve the linear system $\mathbf{y} = \mathbf{X}\alpha$. Thus, the recognition problem can be cast as one of classifying among multiple linear regression models, where the training samples play the role of regressors, and the query sample is the corresponding response.

Considering the under-determined system ($m < n$), the solution to problem (2) is not unique. The popularly used solver is to seek the most parsimonious representation with a constraint term on the sparsity level [20],

$$\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon \tag{3}$$

where $\|\cdot\|_0 : \mathbb{R}^n \mapsto \mathbb{R}$ counts the total number of nonzero entries, and $\varepsilon$ is the allowed error tolerance. Since $\ell_0$-minimization optimization is NP-hard, problem (3) can be only approximated with some algorithms, *e.g.*, matching pursuit, orthogonal matching pursuit [29], [36], [37]. Due to the development in the theory of compressed sensing [38], if the solution $\hat{\alpha}$ is sparse enough, (3) can be relaxed as $\ell_1$-norm minimization

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon \tag{4}$$

where $\|\cdot\|_1$ sums up the absolute weights of all entries. Since (4) is convex and differential, it can be easily solved by the common optimization methods, as reviewed in Ref [39].

Given the optimal representation $\hat{\alpha}$, the class label of query sample $\mathbf{y}$ is decided by searching which class of samples could result in the minimum reconstruction error:

$$identity(\mathbf{y}) = \min_{k=1,...,K} \|\mathbf{y} - \mathbf{X}_k\hat{\alpha}_k\|_2. \tag{5}$$

This model is known as sparse representation-based classification (SRC) [18], [21].

## III. Classification via Tri-Task Joint Sparse Representation of Monogenic Signal

In the conventional approach to SAR target recognition, explicit pose estimation is usually required [40]. Moreover, multiple preprocessing procedures, *e.g.*, noise filter, image segmentation, are needed to extract the discriminative feature [41]. To circumvent these problems, a sparse representation approach for classifying different targets in SAR images is proposed in [32], where random projections is used to produce a dimensionally acceptable feature descriptor. However, some issues should be further considered. First, different from the conventional optical imaging sensors, where the images are mainly formed via diffuse reflection from a non-coherent source of light, SAR images are mainly generated from a specular reflection of a coherent source. Thus, the characteristics of SAR images vary drastically even with small changes in pose and depression, and hence exhibits broad spectral information yet spatial localization. Due to the unique properties of SAR image formation, *e.g.*, specular reflection, multiple bounces, low-resolution and non-literal natures, it is not effective to represent SAR image via the conventional linear feature, as used in the common recognition task. Specifically, the standard features as recommended previously may not be effective to handle target recognition under non-literal experimental setup, *e.g.*, different poses, depression angles and configurations, because the characteristics of SAR image vary abruptly with the slight change of operating conditions. Furthermore, it is costly to collect sufficient SAR dataset, and hence the valid training resources available for target recognition are scarce to access.

To effectively represent SAR image and improve target recognition performance, the preceding works jointly consider the monogenic signal and sparse representation modeling into a unifying framework [23], [24]. They generate a single feature descriptor from the components of monogenic signal at different scale-spaces, and pool the resulting feature into the framework of SRC. However, these methods could not transfer some knowledge from one task to another. To cover the shortage, this paper introduces a novel classification scheme via multi-task joint sparse representation learning. We construct three component-specific classification systems using the three features derived from the components of monogenic signal at different scales. By limiting the representations of these classifications with a joint sparse constraint term, it is able to produce the representations that share the similar sparse pattern with respect to their own dictionary. The resulting representations are much more discriminative than the ones produced by the individual classifications. The final decision is made in terms of the total reconstruction error.

### A. The Monogenic Signal

The monogenic signal is an extension of the analytic signal to multi-dimensional space [1]. Though it is valid for a general number $N$-dimension ($N > 1$), this paper only considers the case at hand of 2-D signal with finite length (image). An intriguing topic is to extend it into high-dimensional signal processing (*e.g.*, video-based pattern recognition).

The monogenic signal has been built around the Riesz transform, a natural multi-dimensional generalization of Hilbert transform. Given signal $f(\mathbf{z}), \mathbf{z} = (x, y)^T$ and its Fourier spectra $F(\mathbf{u}), \mathbf{u} = (u, v)^T$, the expression of Riesz-transformed signal in the frequency domain is

$$\begin{aligned} F_R(\mathbf{u}) &= H(\mathbf{u})F(\mathbf{u}) = \frac{i\mathbf{u}}{|\mathbf{u}|}F(\mathbf{u}) \\ &= [H_u(\mathbf{u})F(\mathbf{u}), H_v(\mathbf{u})F(\mathbf{u})]^T, \end{aligned} \tag{6}$$

while the corresponding one in the spatial domain is

$$\begin{aligned} f_R(\mathbf{z}) &= h(\mathbf{z}) * f(\mathbf{z}) = -\frac{\mathbf{z}}{2\pi|\mathbf{z}|^3} * f(\mathbf{z}) \\ &= [h_x(\mathbf{z}) * f(\mathbf{z}), h_y(\mathbf{z}) * f(\mathbf{z})]^T \end{aligned} \tag{7}$$

with the first- and second-order filters $h_x, h_y$ characterized by the frequency response $H_u = -\frac{u}{|\mathbf{u}|}i$ and $H_v = -\frac{v}{|\mathbf{u}|}i$.

The linear combination of signal $f(\mathbf{z})$ and the Riesz transformed one $f_R(\mathbf{z})$ yields a sophisticated 2-D analytic signal namely monogenic signal,

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

where $i$ and $j$ are the imaginary units and $\{i, j, 1\}$ forms an orthonormal basis in $\mathbb{R}^3$. Obviously, the original signal and Riesz transformed one comprise the real and imaginary parts of the monogenic signal. Then, it is capable to orthogonally decompose the original signal into three components, local amplitude, local phase, and local orientation, which are formulated by

$$\begin{aligned} amplitude &: \mathbf{A} = \sqrt{f(\mathbf{z})^2 + |f_R(\mathbf{z})|^2} \\ phase &: \phi = atan2(|f_R(\mathbf{z})|, f(\mathbf{z})) \in (-\pi, \pi] \\ orientation &: \theta = atan(f_y(\mathbf{z})/f_x(\mathbf{z})) \in (-\frac{\pi}{2}, \frac{\pi}{2}] \end{aligned} \tag{9}$$

where $f_x(\mathbf{z}) = f(\mathbf{z}) * h_x(\mathbf{z})$ and $f_y = f(\mathbf{z}) * h_y(\mathbf{z})$ are the $i$-imaginary component and $j$-imaginary component of the monogenic signal. In (9), local amplitude represents the local energetic information. It is similar to the one in the analytic signal. Difference from 1-D analytic signal, the monogenic phase composes of two different components, local phase and local orientation, where the former describes the local structural information, and the latter represents the local geometric information.

As a multi-dimensional extension of the analytic signal, the monogenic signal could preserve the properties of the standard analytic signal, *e.g.*, symmetry, doubling energy of the raw signal, nonnegative part of spectra, allpass transfer function $H(\mathbf{u})$, as well as the invariance-equivariance property of signal decomposition.

The practical signal is usually of finite length, and hence result in infinite spectra in the frequency domain. Thus, it is necessary to extend the signal to be infinite by a bandpass filter. Give the bandpass filters $h_b(\mathbf{z})$, the monogenic signal can be rewritten as

$$f_M(\mathbf{z}) = (h_b(\mathbf{z}) * f(\mathbf{z})) - (i, j)(h_b(\mathbf{z}) * f_R(\mathbf{z})). \tag{10}$$

Then, the next problem is how to determine the bandpass filter.

The property of Riesz transform requires that $h_b$ should be symmetric to maintain the odd character of Riesz kernel $h$ and with null direct current component [42]. Considering the mathematical, computational, and empirical reasons, log-Gabor filter bank is utilized in this paper. Moreover, it has several remarkable properties. First, it is capable to capture the broad spectral information of the signal with compact spatial filter, as recommended in the preceding works [10], [11]. In addition, it has a Gaussian Shaped response along the logarithmic frequency scale rather than a linear one, thus it allows more information to be captured in the high frequency band and endows desirable high pass characteristics accordingly. The frequency response of log-Gabor filter can be described as

$$G(\omega) = \exp\Big\{ -\frac{[\log(\omega/\omega_0)]^2}{[\log(\sigma/\omega_0)]^2} \Big\}, \tag{11}$$

where $\omega_0$ is the center frequency, and $\sigma$ is the scaling factor of the bandwidth. Usually, the ratio $\sigma_r = \frac{\sigma}{\omega_0}$ is kept as a constant. By tuning the scale index $s$, it is able to produce multiresolution monogenic signal representation. Then, the monogenic signal is actually embedded in the log-Gabor scale-space by convolving the components with the log-Gabor kernel, and hence the monogenic scale-space can be formed accordingly.

Provided $S$-scale log-Gabor filters, the monogenic signal at different scales can be shown as $\{f_M^1, f_M^2, \cdots, f_M^S\}$, and the corresponding monogenic components can be generated as

$$\Big\{ \underbrace{\mathbf{A}_1, \phi_1, \theta_1}_{f_M^1}, \underbrace{\mathbf{A}_2, \phi_2, \theta_2}_{f_M^2}, \cdots, \underbrace{\mathbf{A}_S, \phi_S, \theta_S}_{f_M^S} \Big\}, \tag{12}$$

where $A_k, \phi_k, \theta_k$ denote local amplitude, local phase, and local orientation of the monogenic signal at the $k$-th scale.

The feature set $\mathbf{f}$ is incapable to be directly applied to learning system due to the high dimension and redundancy. To solve the problem, the preceding work [23] defines an augmented monogenic feature vector via uniform down-sampling, normalization, and concatenation of the multiple components at various scales. The work [24] combines the monogenic components by constructing a region covariance matrix whose entries are the correlation of the components themselves.

### B. Classification via Tri-Task Joint Sparse Representation of Monogenic Components (TJSR)

Unlike the preceding works [23], [24], where a single feature descriptor are produced by concatenating the monogenic components, this paper generates three component-specific features by three components of the monogenic signal at different scales. Specifically, the components of the monogenic signal (local amplitude, local phase, and local orientation) at $S$ scales are first downsampled by a factor of $\rho$ to reduce the

dimension and redundancy. For each component, the down-sampled coefficient matrices at different scales are reshaped to be vectors by raster-scanning. The resulting vectors are then normalized and concatenated to generate a component-specific feature vector. The overall process can be formulated by

$$\begin{aligned}\chi_{\mathbf{A}} &= [vec_\rho(\mathbf{A}_1); vec_\rho(\mathbf{A}_2); ...; vec_\rho(\mathbf{A}_S)]\\ \chi_\phi &= [vec_\rho(\phi_1); vec_\rho(\phi_2); ...; vec_\rho(\phi_S)],\\ \chi_\theta &= [vec_\rho(\theta_1); vec_\rho(\theta_2); ...; vec_\rho(\theta_S)]\end{aligned} \tag{13}$$

where $vec(\cdot)$ reshapes a matrix to be a single vector by raster-scanning. The generation of the component-specific features are illustrated in Fig. 2.

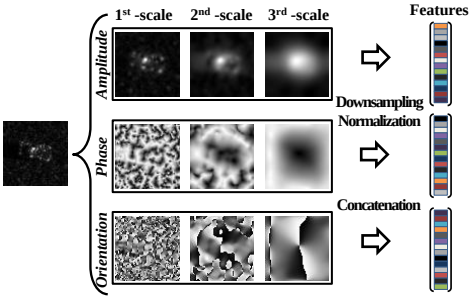



Fig. 2. The generation of monogenic component-specific features.

In the monogenic signal representation, local amplitude represents the local energetic information, while the monogenic phase (local phase and local orientation) provide the local structural information and the local geometric information. Thus, the combination technique via simple concatenation usually produces some information loss. To cover the shortage, this paper generates a component-specific feature for each component, *i.e.*, we derives three different feature descriptors by local amplitude, local phase, and local orientation, respectively. The resulting features are then used to construct three component-specific classification systems. Then, the problem of classifying the query samples converts to three component-specific learning tasks. For each task, the features resulting from the training samples play the role of regressors (or observations, predictors), while the one of query sample is viewed as the corresponding response. Thus, it is actually a tri-task learning problem. In machine learning and computer vision fields, when the tasks to be learned share some latent factors, it may be advantageous to take into account these cross-task relations. For instance, when the number of samples for learning is small, transferring some knowledge from one task to another can be advantageous in term of generalization performance. The detail studies on multi-task learning can be found in [27], [30].

Given a training set with $n$ training samples from $K$ different classes, $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, ..., \mathbf{x}_{k,n_k}], k = 1, ..., K$, it is able to generate the redundant dictionary for sparse representation modeling. For each sample $\mathbf{x}_{k,j}$, we first produce its monogenic features, $\chi_{\mathbf{A}}(\mathbf{x}_{k,j}), \chi_\phi(\mathbf{x}_{k,j}), \chi_\theta(\mathbf{x}_{k,j})$, by (13). By stacking the monogenic features of all training samples, it is able to build three component-specific overcomplete dictionaries $(card(\chi_{\mathbf{A}}), card(\chi_\phi), card(\chi_\theta) < n)$, which can be formulated as

$$\begin{aligned}\mathbf{X}^{(1)} &= [\chi_{\mathbf{A}}(\mathbf{x}_{1,1}), \chi_{\mathbf{A}}(\mathbf{x}_{1,2}), ..., \chi_{\mathbf{A}}(\mathbf{x}_{K,n_K})]\\ \mathbf{X}^{(2)} &= [\chi_\phi(\mathbf{x}_{1,1}), \chi_\phi(\mathbf{x}_{1,2}), ..., \chi_\phi(\mathbf{x}_{K,n_K})].\\ \mathbf{X}^{(3)} &= [\chi_\theta(\mathbf{x}_{1,1}), \chi_\theta(\mathbf{x}_{1,2}), ..., \chi_\theta(\mathbf{x}_{K,n_K})]\end{aligned} \tag{14}$$

For any query sample $\mathbf{y}$, the corresponding features can be obtained by (13) similarly,

$$\mathbf{y}^{(1)} = \chi_{\mathbf{A}}(\mathbf{y}), \mathbf{y}^{(2)} = \chi_\phi(\mathbf{y}), \mathbf{y}^{(3)} = \chi_\theta(\mathbf{y}). \tag{15}$$

Following the thought of the preceding works [18], [23], [32], this paper casts the recognition problem as one of classifying among multiple linear regression models, while the features of training samples play the role of regressors, and the counterpart of query is the corresponding response. Then, the formulation of linear regression models can be expressed as,

$$\mathbf{y}^{(k)} = \mathbf{X}^{(k)}\alpha^{(k)} + \varepsilon^{(k)} \tag{16}$$

where $k = 1, 2, 3$ index three different learning tasks formulated by local amplitude, local phase, and local orientation; $\alpha^{(k)}$ and $\varepsilon^{(k)}$ denote the representation and residual for the $k$-th learning tasks. To solve the problem (16), an intuitive idea is to conduct each learning task individually, and combine their results. Suppose the residuals are independently identical distribution Gaussian noise, a commonly used approach to estimate the representation is the least square fitting technique,

$$\min_\alpha \{g(\mathcal{A}) = \sum_{l=1}^{3} \|\mathbf{y}^{(l)} - \sum_{k=1}^{K} \mathbf{X}_k^{(k)}\alpha_k^{(k)}\|_2\}, \tag{17}$$

where $\mathcal{A} = [\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}]$ is the representation matrix. Each column of $\mathcal{A}$ are resulting from a classification task. However, this strategy is incapable to exploit the inter-correlation among each learning task, and hence its solver is not discriminative.

To take the relationship among the three learning tasks into account, this paper jointly considers the three monogenic component-specific learning tasks into a unifying framework for target recognition. Specifically, we assume that three classification systems share a common sparse pattern, *i.e.*, the representations derived from each task have the similar nonzero entries with respect to their own dictionary, yet the weight values may be different. Thus, three sparse representation-based classification tasks are simultaneously optimized, rather than individually solved.

The proposed tri-task joint sparse representation model is realized by imposing a special sparsity-inducing term on the representation matrix. Then, the solution to problem (16) can be obtained by optimizing a $\ell_1/\ell_q$-norm regularization,

$$\min_{\mathcal{A}} g(\mathcal{A}) + \lambda\|\mathcal{A}\|_{q,1}, \tag{18}$$

where $\lambda$ is a parameter to make a tradoff between the fidelity (the accuracy of reconstruction error) and the sparsity level. The mixed-norm $\|\alpha\|_{q,1}$ means applying $\ell_1$-norm on each row, followed by $\ell_q$-norm on the resulting vector. Fig. 3 draws the diagram of triple-task joint sparse representation model.

Given the optimal coefficient matrix $\hat{\mathcal{A}}$, the query sample can be approximately reconstructed, $\mathbf{y}^{(k)} = \mathbf{X}^{(k)}\hat{\alpha}^{(k)}$. Then, the decision can be made by evaluating which class of training

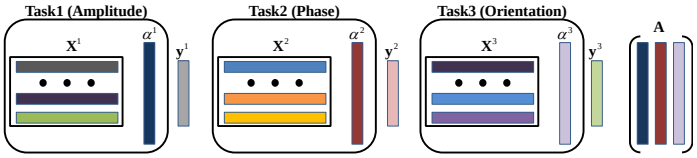



Fig. 3. The block diagram of tri-task joint sparse representation. Each block denotes a linear regression model formed by a monogenic feature. The interested reader can be addressed to [43] for a more intuitive description.

samples (features) could produce the lowest total reconstruction error,

$$identity(\mathbf{y}) = \min_{k=1,...,K} \sum_{l=1}^{3} \|\mathbf{y}^{(k)} - \mathbf{X}_k^{(l)} \hat{\alpha}_k^{(l)}\|_2. \tag{19}$$

where $\hat{\alpha}_k^{(l)}$ denotes the coefficients associated with the $k$-th class in the $l$-th task.

### C. Numerical Implementation

Suppose $q = 2$, the problem (17) can be rewritten as

$$\min_{\mathcal{A}} g(\mathcal{A}) + \lambda \|\mathcal{A}\|_{1,2}. \tag{20}$$

The method (20) evolves to a typical $\ell_1$-regularized optimization if the number of the task is 1, $l = 1$. Denote by $\alpha^i \in \mathbb{R}^3, i = 1, ..., n$ the $i$-th row of the representation matrix $\mathcal{A}$, and $\alpha_j \in \mathbb{R}^n, j = \{1, 2, 3\}$ the $j$-th column.

Since the objective of (17) is differential everywhere (for $0 < q < \infty$), the optimal representations $\hat{\mathcal{A}}$ satisfies the following conditions,

$$\begin{aligned} \hat{\alpha}_j = 0,\ \|\nabla_{\alpha_j} g(\hat{\mathcal{A}})\|_2 \leqslant \lambda,\ or \\ \hat{\alpha}_j \propto -\nabla_{\alpha_j} g(\hat{\mathcal{A}}),\ \|\nabla_{\alpha_j} g(\hat{\mathcal{A}})\|_2 = \lambda \end{aligned} \tag{21}$$

where $\nabla_{\alpha_j} g(\hat{\mathcal{A}})$ are partial gradients in each of the subspace corresponding to the representation vectors resulting from the component-specific classifications.

Similar to the feature-sign search algorithm [15], only the active elements, whose norm of the gradient vector is not strictly less than $\lambda$ ($\lambda \frac{\hat{\alpha}_j}{\|\hat{\alpha}_j\|} = \nabla_{\alpha_j} g(\hat{\mathcal{A}})$), participate in the solution.

Since the condition (21) is too strict to be realized, we resort to approximate subgradient one,

$$\begin{aligned} \hat{\alpha}_j = 0,\ \|\nabla_{\alpha_j} g(\hat{\mathcal{A}})\|_2 < \lambda + \xi_0,\ or \\ \|\nabla_{\alpha_j} g(\hat{\mathcal{A}}) + (\lambda - \xi) \frac{\alpha_j}{\|\alpha_j\|}\| \leqslant \xi \end{aligned} \tag{22}$$

where $\xi$ and $\xi_0$ are slack parameters. Then, the solution to (17) can be obtained by alternatively iterating two following steps,

$$\begin{aligned} \alpha^{(t+1)} = \alpha^{(t)} - \epsilon \frac{\nabla_\alpha g}{\|\nabla_\alpha g\|}\ and \\ \lambda^{t+1} = \min(\lambda^t, \frac{g(\mathcal{A}^t) - g(\mathcal{A}^{t+1})}{\epsilon}) \end{aligned} \tag{23}$$

where $t$ is the iteration index, and $\epsilon$ is the step index. We address the interested readers to Ref [30], [43]–[45] for more detail descriptions. The proposed method can be pictorially shown in Fig. 4, the corresponding recognition procedure is given in Algorithm 1.

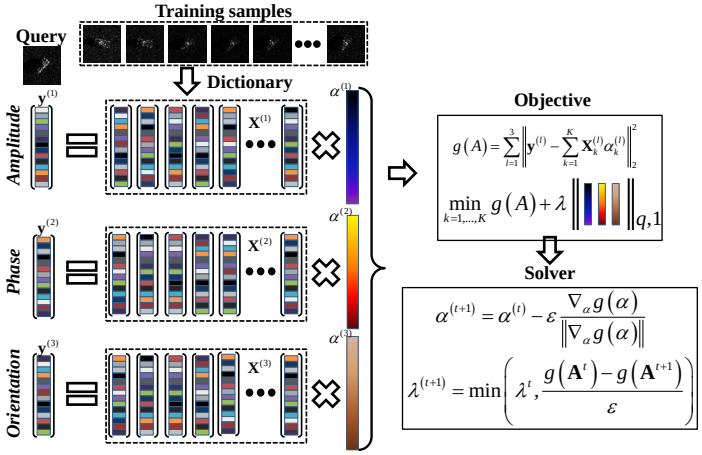



Fig. 4. The illustration of proposed method. Three features are generated by the components of the monogenic signal at different scale-spaces. Then, multi-task learning is used to produce the representations with a similar sparse pattern with respect to their own dictionary. The decision is reached in favor of the class with the lowest total reconstruction error accumulated from all classification tasks.

### D. Illustrative Examples

To demonstrate the proposed method, a set of four-target-recognition experiments are performed, where BMP2, BTR60, T72, and T62 are used. The details on dataset can be found in the next section. Some examples are shown in Fig. 5.

From Fig. 5, it can be seen: local amplitude based classification correctly predict the class labels of first three query samples; local phase based classification correctly infer the class memberships of last two query samples; local orientation based classification correctly estimate the label of the fourth query sample. The result demonstrate that each learning task favors different representation coefficients with respect to their own dictionary. By jointly considering three classifications into a unifying framework for target recognition via multi-task learning, the proposed method make the correct decisions for all query samples.

To demonstrate the advantage for exploiting the inter-correlation among multiple learning tasks, an exemplar classification is given in Fig. 6. The query sample is chosen from

**Algorithm 1** Classification via Tri-Task Joint Sparse Representation of Monogenic Components (**TJSR**)

**Input:** $\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, ..., \mathbf{x}_{k,n_k}$: a set of training samples from $K$ different target classes; $\mathbf{y}$: a query sample; $\lambda$: a regularization parameter; $\rho$: the downsampling ratio;

**Output:** the identity of $\mathbf{y}$.

1: Orthogonally decompose each training sample into three components by (13), from which three monogenic feature vectors can be produced by (14);
2: Build three overcomplete dictionaries by stacking the features of all the training samples, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \mathbf{X}^{(3)}$;
3: Produce the optimal representations by optimizing (18);
4: Compute the total reconstruction error by (19);
5: Reach the inference by seeking the minimum total residual.

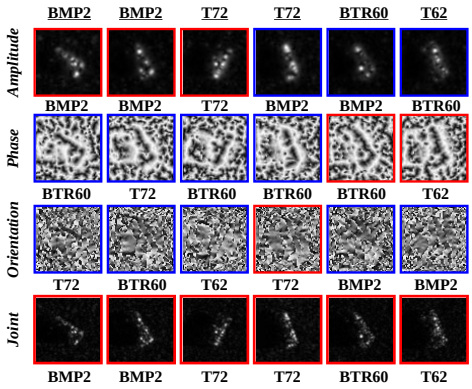



Fig. 5. Example classifications. Each column record a group of experiments. The ground-truth of query sample is in the top (underlined). Images with red border indicate correct prediction, while the ones with blue border mark incorrect inference. The first three rows give the results of individual sparse representation model, while the last row provides the results of joint sparse representation model.

T72 set. The representations and the corresponding residuals of the individual classification are in blue border, while the ones of the joint sparse representation are in red border. As can be seen, though amplitude-specific classification and phase-specific one predict the query sample as T62 and BMP2, the proposed method still reaches the correct decision, T72. This is because the proposed method could transfer some knowledge from one task to another. Thus, the resulting representations are much more discriminative than the one produced by the individual classification.

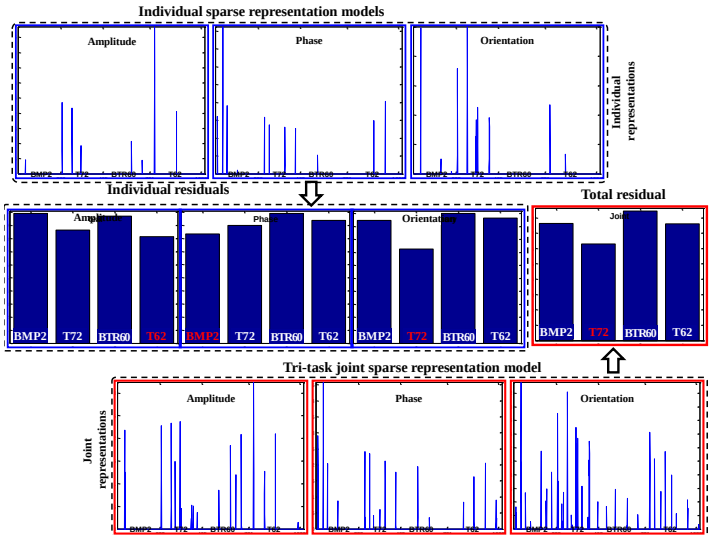



Fig. 6. The comparison of the individual sparse representation model and joint sparse representation one. The coefficients obtained by thee component-specific classifications are drawn in the first row, while the ones obtained using joint sparse representation are plotted in the third row. The corresponding residuals are shown in the middle.

red

## IV. Experiments and Discussions

This section verifies the proposed method on MSTAR database, a gallery collected using a 10 GHz SAR sensor with $1ft \times 1ft$ resolution in range and azimuth. Images are captured at various depressions over $0 \sim 359^\circ$ range of aspect view. They are of around $128 \times 128$ pixels in size, and cropped to $64 \times 64$ pixels region of interest. The details on the number of aspect view for each target are tabulated in TABLE I. The downsampling factor in (13) is experimentally set as $\rho = \frac{1}{64}$, which corresponds the downsampled image with $8 \times 8$ pixels in size. For the multiresolution monogenic signal representation, 3-scale log-Gable filter bank ($S = 3$) is built. To quantitatively assess the performance, several state-of-the-art algorithms, including support vector machine learning, the extended maximum average correlation height, discriminative graphical models, SRC, and our preceding works [23], [24] are used as the reference. The methods to be studied are summarized in TABLE II.

In the remainder of this section, several kinds of experiments are performed to demonstrate the merits of the proposed method. The experimental scenario includes target recognition under standard operating conditions (**SOC**), extended operating conditions (**EOC**) and limited training resource. Moreover, the capability to outlier rejection is verified.

### A. Target Recognition under Standard Operating Conditions

We first consider target recognition under standard operating condition. Images acquired at an operating condition of a $17^\circ$ depression angle are used to train the algorithm, while the ones captured at an operating condition of a $15^\circ$ depression angle are used for testing, as shown in TABLE I. All of ten targets are employed, among which BMP2 and T72 have several variants with small structural modifications (denote by series number). Only the standards (SN_9563 for BMP2 and SN_132 for T72) captured at $17^\circ$ depression (in bold in TABLE I) are available for training.

We provide the quantitative comparison of tri-task joint sparse representation with the three component-specific sparse representations first. The experimental result is tabulated in TABLE III. From TABLE III, we are able to see that the recognition rate of joint sparse representation is 4.22%, 8.40%, and 10.52% better than three component-specific sparse representations. The result corroborates that the three monogenic components are correlated for target recognition. It is beneficial by jointly considering them in a unifying classification framework. This conclusion is an important motivation for a joint sparsity model.

Then, the comparison of the proposed method with the reference methods are given in Fig. 7, where the confusion matrices, as well as the overall recognition rates are tabulated. The classification accuracies of SVM, EMACH, AdaBoost, and IGT are consistent with the results reported in the preceding works [40], [41], [48]. For SVM, AdaBoost, and IGT, two specific experimental cases, classification with pose estimation and the one without pose estimation, are available. It can be seen the algorithms with pose estimation perform much better than the ones without pose estimation. The overall recognition rate for IGT, AdaBoost and SVM under pose estimation are 0.9475, 0.9296, and 0.9042, compared to 0.8884, 0.8742, 0.8428 for IGT, AdaBoost, and SVM without pose estimation.

TABLE I
THE NUMBER OF ASPECT VIEWS AVAILABLE FOR DIFFERENT TARGETS. THE SERIES NUMBERS FOR BMP2 AND T72 ARE IN PARENTHESES.

| *Depr.* | BMP2 | BTR70 | T72 | BTR60 | 2S1 | BRDM2 | D7 | T62 | ZIL131 | ZSU23/4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 17° *Train* | 233 **(SN_9563)** 232 (SN_9566) 233 (SN_c21) | 233 | 232 **(SN_132)** 231 (SN_812) 228 (SN_s7) | 256 | 299 | 298 | 299 | 299 | 299 | 299 |
| 15° *Test* | 195 (SN_9563) 196 (SN_9566) 196 (SN_c21) | 196 | 196 (SN_132) 195 (SN_812) 191 (SN_s7) | 195 | 274 | 274 | 274 | 273 | 274 | 274 |

TABLE II
THE METHODS TO BE STUDIED IN THIS PAPER.

| *Abbre.* | Full name (or description) | *Ref.* | Input |
|---|---|---|---|
| k-NN | k-Nearest-Neighbor classifier | [46] | Intensity values |
| SVM | Linear support vector machine learning | [47] | Intensity values |
| EMACH | The extended maximum average correlation height filter | [48] | DFT coefficients |
| AdaBoost | Feature fusion via boosting on individual neural net classifiers | [41] | Intensity values and DFT coefficients |
| IGT | Target recognition using discriminative graphical models | [40] | Wavelet features (LL, LH, and HL bands) |
| SRC | Sparse representation-based classification | [18], [32] | Intensity values |
| MSRC | Sparse representation of monogenic signal | [23] | Augmented monogenic feature vector |
| SRCR | Sparse representation on Riemannian manifolds | [24] | Covariance feature of monogenic components |
| TJSR | Joint sparse representation of monogenic signal | — | Monogenic component-specific feature vectors |

TABLE III
THE COMPARISON OF JOINT SPARSE REPRESENTATION WITH THREE COMPONENT-SPECIFIC SPARSE REPRESENTATION.

| *Amplitude* | *Phase* | *Orientation* | **Joint** |
|---|---|---|---|
| 0.8919 | 0.8501 | 0.8289 | 0.9341 |

Thus, we could come the conclusion that the performances of IGT, AdaBoost, and SVM are highly reliant on the accuracy of pose estimation. On the contrary, the proposed method and our preceding works, MSRC [23], SRCR [24] do not require explicit pose estimation or any preprocessing. These methods only refer to generating the monogenic feature by simple downsampling or computing the covariance matrix, and hence are computationally attractive. Moreover, the recognition rate for TJSR is 0.9341, slightly lower than IGT with pose estimate (0.9476) and SRCR (0.9488), and much better than the remaining competitors. It is 0.45%, 2.99%, and 4.73% better than AdaBoost with pose estimation, SVM with pose estimation, and EMACH. The improvement for recognition accuracy is significant compared with those methods without pose estimation.

### *B. Target Recognition under Extended Operating Conditions*

To assure the practicability of the method, we pay attention to target recognition under extended operating conditions subsequently. It adheres to the standards set forth to train the algorithm at a specific operating condition and test them at another significantly different operating condition. To verify the performance of proposed method under extended operating conditions, several comparative experiments, including EOC difference on configuration and EOC difference on depression, are performed.

*1) EOC Difference on Configuration:* In the realistic battlefields, there are many different physical targets that can be categorized as a single general class in military. Fig. 8 provides some representative exemplars for configuration variation. To verify this property, this subsection considers target recognition under EOC difference on configuration. Four vehicle targets, BMP2, T72, BTR60, and T62 are utilized, among which BMP2 and BTR60 are armored personnel carriers, T72 and T62 are main-battle tanks. A group of exemplars for these targets are given in Fig. 9. Obviously, their scattering phenomenology are very similar in visual. Thus, it is challenging to differentiate each one from the others.

TABLE IV
THE NUMBER OF ASPECT VIEWS AVAILABLE FOR TRAINING AND TESTING IN EOC DIFFERENCE ON CONFIGURATION

| *Depr.* | BMP2 | T72 | BTR60 | T62 |
|---|---|---|---|---|
| 17° (**Train**) | **233** (*SN_9563*) | **232** (*SN_132*) | **256** | **299** |
| 15° (*Test*) | 196 (*SN_9566*) 196 (*SN_c21*) | 195 (*SN_812*) 191 (*SN_s7*) | 195 | 273 |

For BMP2 and T72, there are several variants with small structural modifications. The standards, SN_9566 for BMP2 and SN_132 for T72, collected at a 17° depression are used to train the algorithm, while the remaining (SN_9563, SN_c21, SN_812, SN_s7) taken at a 15° depression are used for testing, *i.e.*, the configurations available for testing are not contained in the training set. Thus, the configuration as well as the depression are significantly different between the image used for training and those used for testing. This is a significantly different experimental setup in comparison to the conventional works [40], [41], [48]. For BTR60 and T62, there is only one configuration, thus only the depression angle is different between the images available for training and those for testing. The detail on the experimental setting is given in TABLE IV.

TABLE V tabulates the confusion matrices and the overall recognition rates obtained using the proposed method and the reference algorithms, including SVM, SRC, SRCR, MSRC, and kNN. The proposed method achieves the performance similar to our preceding work [24], where the classification

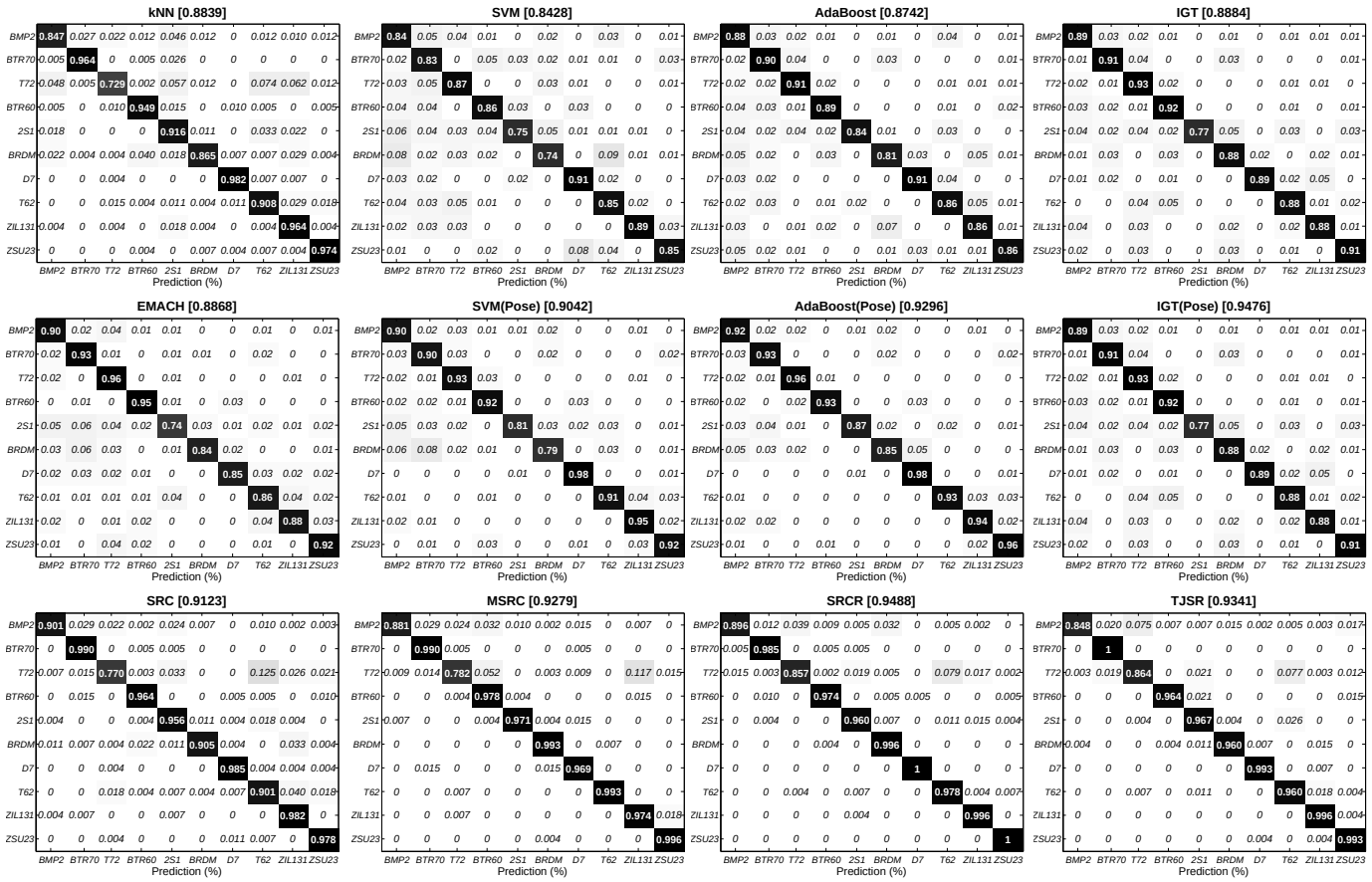



Fig. 7. The confusion matrices for the methods to be studied under standard operating conditions. In each sub-figure, the row shows the ground-truth class of the query sample, while the column displays the class membership predicted by the algorithms to be studied. The diagonal entries demonstrate the recognition rate of each class, while the non-diagonal entries represent the misclassification rate. The top entries in bracket are the overall recognition rates.

| Variability Category | Examples |
|---|---|
| Version Variant | Smoke Grenade Launchers, Side Skirts |
| Configuration Variant | Two Cables, Fuel Barrels |
| Structural Modifications | Dented Fenders, Broken Antenna Mount |

(a)

(b)

Fig. 8. Target variability. (a) Categories of target variability and the corresponding examples. (b) Three variants of T72 tank.

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

Fig. 9. Exemplars of four targets to be recognized. (a), (b), (c), and (d) show SAR images of BMP2, BTR60, T72, and T72 with a 45° aspect view taken at a 15° depression angle.

are constructed on Riemannian manifolds, rather than an Euclidean space. The overall recognition rate for TJSR is 0.9117, compared to 0.8748 for MSRC, 0.8692 for SRC, 0.8459 for SVM, and 0.8587 for kNN. It is 3.69%, 4.25%, 6.58%, and 5.30% better than the competitors, MSRC, SRC, SVM, and kNN. From the experimental results, we can see that the proposed method could effectively deal with the problem of target recognition under configuration variation.

*2) EOC Difference on Depression:* This subsection handles target recognition under different depression angles. Three vehicle targets, 2S1, BRDM2, and ZSU23/4 are utilized, among which BRDM2 and ZSU23/4 have the articulated variants. The articulated refers to a physical variation, such as open hatch or rotated gun turret. Fig. 10 displays the standard and the articulated of ZSU23/4 taken at the different depressions.

(a) (b)

Fig. 10. Illustration of the standard and the articulated; (a) and (b) provide SAR image of ZSU23/4 with turret straight and articulated taken at 30° and 45° depression.

To assess the performance under different depression angles, images captured at an operating condition of a 17° depression are used to train the algorithm, while the ones collected at an operating condition of 30° and 45° depressions are used for testing. The details on images available for training and

TABLE V
THE CONFUSION MATRICES FOR THE METHODS TO BE STUDIED UNDER EOC DIFFERENCE ON CONFIGURATION.

| | kNN (0.8587) | | | | SVM (0.8459) | | | | SRC (0.8692) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | BMP2 | T72 | BTR60 | T62 | BMP2 | T72 | BTR60 | T62 | BMP2 | T72 | BTR60 | T62 |
| BMP2 | **0.8929** | 0.0485 | 0.0281 | 0.0306 | **0.7985** | 0.1352 | 0.0434 | 0.0230 | **0.9490** | 0.0306 | 0.0077 | 0.0128 |
| T72 | 0.0984 | **0.7021** | 0.0130 | 0.1865 | 0.0466 | **0.7694** | 0 | 0.1839 | 0.0596 | **0.6658** | 0.0052 | 0.2694 |
| BTR60 | 0.0154 | 0.0154 | **0.9641** | 0.0051 | 0.0564 | 0.0154 | **0.8821** | 0.0462 | 0.0051 | 0 | **0.9846** | 0.0103 |
| T62 | 0 | 0.0330 | 0.0110 | **0.9560** | 0.0037 | 0 | 0 | **0.9963** | 0.0073 | 0.0183 | 0.0147 | **0.9597** |
| | MSRC (0.8748) | | | | SRCR (0.9117) | | | | TJSR (0.9117) | | | |
| | BMP2 | T72 | BTR60 | T62 | BMP2 | T72 | BTR60 | T62 | BMP2 | T72 | BTR60 | T62 |
| BMP2 | **0.9286** | 0.0230 | 0.0255 | 0.0230 | **0.9311** | 0.0281 | 0.0306 | 0.0102 | **0.9107** | 0.0510 | 0.0230 | 0.0153 |
| T72 | 0.0415 | **0.6736** | 0.0259 | 0.2591 | 0.0389 | **0.8005** | 0.0052 | 0.1554 | 0.0181 | **0.8394** | 0 | 0.1425 |
| BTR60 | 0 | 0 | **0.9897** | 0.0103 | 0.0103 | 0 | **0.9846** | 0.0051 | 0.0256 | 0.0103 | **0.9385** | 0.0256 |
| T62 | 0 | 0 | 0 | **1** | 0 | 0.0073 | 0.0037 | **0.9890** | 0 | 0.0037 | 0 | **0.9963** |

the ones for testing are summarized in TABLE VI, where the entries in parentheses denote the number of aspect views available for the articulated. The confusion matrices and the overall recognition rates obtained using the methods to be studied are tabulated in TABLE VII.

The first experimental scenario copes with a medium EOC difference on depression. Images taken at an operating condition of 30° are used to test the algorithm, *i.e.*, a change of 13° from 17° to 30° exists between the images available for training and those for testing. The performances are satisfied on the whole. The overall recognition rate for TJSR is 0.9524, compared to 0.9515 for SRCR, 0.9497 for MSRC, 0.9488 for SRC, and 0.9237 for kNN. The performance of the proposed method is slightly better than the baseline algorithms. The algorithms are further evaluated using the images captured at an operating condition of 45°, *i.e.*, there is a change of 28° from 17° to 45° between the training samples and the testing samples. Then, the performances of all the methods are drastically degraded. The recognition rate for SRC even drops from 0.9488 to 0.5366. The proposed method, TJSR, still achieves the highest recognition rate, 0.7073. It is 7.14%, 6.79%, 17.07%, 11.84% better than SRCR, MSRC, SRC, and kNN. The improvement for recognition accuracy is significant. The results demonstrate that the proposed method are much more robust towards depression variation than the reference methods.

TABLE VI
THE NUMBER OF ASPECT VIEWS AVAILABLE FOR TRAINING AND TESTING IN EOC DIFFERENCE ON DEPRESSION.

| *Depression* | | 2S1 | BRDM_2 | ZSU23 | Total |
|---|---|---|---|---|---|
| **Training** (17°) | | **299** | **298** | **299** | **896** |
| *Testing* | (30°) | 288 | 287 (133) | 288 (118) | 1114 |
| | (45°) | 303 | 303 (120) | 303 (119) | 1148 |

### C. Outlier Rejection Performance

This subsection deals with the problem of outlier rejection in target recognition. The outlier includes the confusers (*e.g.*, civil vehicle) and the clutters (natural clutter or man-made clutter). Following the preceding works [40], two targets, D7 and 2S1 are treated as the confusers. To evaluate clutter rejection performance, this paper also employs the clutter chip images generated from the clutter scene SAR image as the invalid samples to be rejected. Fig. 11 provides some clutter exemplars, including trees, buildings, streetlight, *etc*. For the conventional sparse representation technique and its variants, MSRC, and TJSR, the decision is made in terms of the minimum reconstruction error. The smaller the minimum residual, the believable the decision. Thus, we reject the outliers according to their minimum residual. A valid target sample contributes to the probability of detection $P_d$ if the minimum residual is lower than a specific threshold $\tau$. On the contrary, a clutter sample contributes to the probability of false alarm $P_{fa}$ if the minimum residual is lower than the threshold $\tau$. The probability of detection $P_d$ is defined as the percentage of valid samples correctly recognized as targets, while the probability of false alarm is defined as the percentage of clutter images incorrectly classified as targets, correspondingly. The total number of the confuser and the clutter are 548 and 1616. Thus, the probability of false alarm for the confuser and the one for the clutter are a percentage out of 548 and 1616 respectively.

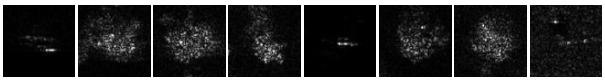


Fig. 11. The clutter exemplars.

By varying the threshold $\tau$ in a certain range, we are able to generate the receiver operating characteristics (**ROC**) curves for $P_d$ vs. $P_{fa}$. It is one of the most widely used approach to assess the capability of outlier rejection. The ROC curves for the proposed method, as well as the reference methods are drawn in Fig. 12. As can be seen the propose method achieves some improvement in the confuser rejection, as well as in the clutter rejection. In Fig 12 (b), when the probability of false alarm is lower than 0.1 (threshold), the probability of detection obtained using the proposed method is much higher than the ones obtained using the remaining competitor algorithms. The similar performance can be observed in Fig 12 (a). The experimental results demonstrate that the proposed method is capable to detect the invalid samples. Moreover, it performs much better than the baseline algorithms.

TABLE VII
THE CONFUSION MATRICES AND OVERALL RECOGNITION RATES OBTAINED USING THE METHODS TO BE STUDIED UNDER EOC DIFFERENCE ON DEPRESSION. ($TG_1$ →2S1, $TG_2$ →BRDM2, $TG_3$ →ZSU23/4).

| | | kNN (0.9237) | | | SRC (0.9488) | | | MSRC (0.9497) | | | SRCR (0.9515) | | | TJSR (0.9524) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $TG_1$ | $TG_2$ | $TG_3$ | $TG_1$ | $TG_2$ | $TG_3$ | $TG_1$ | $TG_2$ | $TG_3$ | $TG_1$ | $TG_2$ | $TG_3$ | $TG_1$ | $TG_2$ | $TG_3$ |
| 30° | $TG_1$ | **0.9479** | 0.0451 | 0.0069 | **0.9757** | 0.0243 | 0 | **0.9618** | 0 | 0.0382 | **0.9132** | 0.0868 | 0 | **0.9861** | 0.0069 | 0.0069 |
| | $TG_2$ | 0.0667 | **0.9048** | 0.0286 | 0.0595 | **0.9190** | 0.0214 | 0.0429 | **0.9262** | 0.0310 | 0.0119 | **0.9810** | 0.0071 | 0.0214 | **0.9000** | 0.0786 |
| | $TG_3$ | 0.0320 | 0.0419 | **0.9261** | 0.0172 | 0.0222 | **0.9606** | 0.0123 | 0.0222 | **0.9655** | 0.0148 | 0.0369 | **0.9483** | 0.0025 | 0.0148 | **0.9828** |
| | | kNN (0.5889) | | | SRC (0.5366) | | | MSRC (0.6394) | | | SRCR (0.6359) | | | TJSR (0.7073) | | |
| 45° | $TG_1$ | **0.3333** | 0.6568 | 0.0099 | **0.7789** | 0.2013 | 0.0198 | **0.8647** | 0.0825 | 0.0528 | **0.4620** | 0.5182 | 0.0198 | **0.5644** | 0.3366 | 0.0990 |
| | $TG_2$ | 0.0615 | **0.9149** | 0.0236 | 0.4232 | **0.5650** | 0.0118 | 0.2648 | **0.4610** | 0.2742 | 0.0591 | **0.9267** | 0.0142 | 0.0757 | **0.6738** | 0.2506 |
| | $TG_3$ | 0.1137 | 0.4929 | **0.3934** | 0.6706 | 0.0237 | **0.3057** | 0.1848 | 0.1588 | **0.6564** | 0.1872 | 0.3436 | **0.4692** | 0.0095 | 0.1469 | **0.8436** |

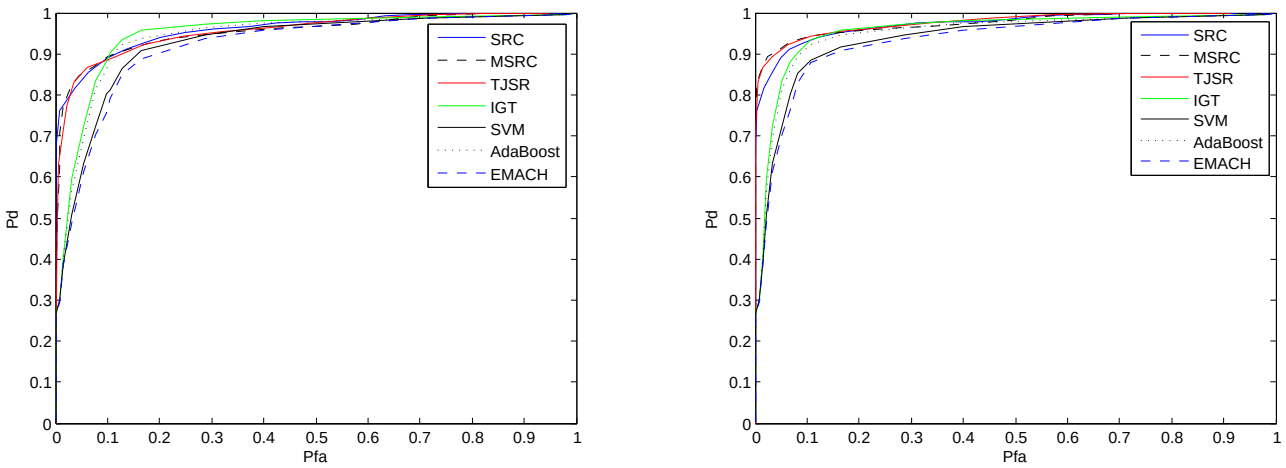



Fig. 12. The experimental results of outlier rejection. (a) ROC curves for the confuser rejection; (b) ROC curves for the clutter rejection.

### D. Computational Complexity

The dominant computational cost of the proposed method consists in the calculation of sparse representation matrix $\mathcal{A}$ via (18). Specifically, the alternative iteration of $\alpha$ and $\lambda$ shown in (23) consumes the most computational resource. The costs of two terms in (23) are $\mathcal{O}(3 \times d \times n)$ and $\mathcal{O}(2 \times 3 \times T \times d \times n)$ floating-point operations, respectively, where $d$ donates the dimensionality of monogenic features, $card(\chi_A), card(\chi_\theta), card(\chi_\phi)$, $n$ is the total number of training samples, and $T$ is the number of iterations for running (23). Then, the total operations for the proposed method can be approximated as $\mathcal{O}(3dn + 6Tdn)$. Correspondingly, the computation complexity for sparse representation [18] and joint sparse representation in Ref [49] are $\mathcal{O}(dnt)$ (by OMP) and $\mathcal{O}(d \log d + dn)$. The computational cost of the proposed method is moderately enhanced than the reference methods. However, it is still acceptable under the realistic applications. Moreover, it would be computationally attractive provided more efficient optimization algorithms, *e.g.*, [50].

### E. Performance under Small Training Size

For the problem of target recognition in SAR image, the training resource is scarce to access. Thus, it is important to evaluate the performance in the difficult regime of limited training samples. This subsection considers target recognition with small training size. Following the experiment prototype in Section IV-B1, four-class target recognition are dealt with. Unlike the experimental setting in Section IV-B1, the number of training samples are gradually decreased to form limited training resource. The training samples preserved are randomly selected from the original training set.

The experimental results are shown in Fig. 13, where the recognition rates across the number of training samples are drawn. With the number of training samples decreased from 1020 to 340, the performances obtained using these methods are gradually degraded. SRCR even generates a drop of 7.54% from 0.9117 to 0.8363 for recognition accuracy. When 340 training samples (nearly one third of the original training set) are preserved, the recognition rate for the proposed method is 0.8595, compared to 0.8363 for SRCR, 0.8539 for SRC, 0.8210 for MSRC, 0.8315 for kNN, and 0.8194 for SVM. The experimental result demonstrates that the proposed method is capable to deal with the limited training resource.

## V. CONCLUSION

This paper presents a multi-task joint sparse representation model for target recognition in SAR image. A recently developed multi-dimensional analytic signal namely monogenic signal is used to capture the broad spectral information yet spatial localization. The components of monogenic signal at different scales are used to generate three component-specific features, from which three linear regression systems can be

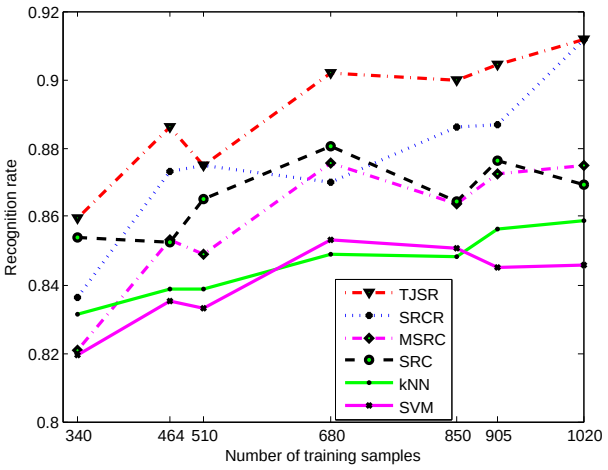



Fig. 13. The recognition rates across the number of training samples.

formed. To transfer some knowledge from one task to another, the multi-task learning is used to produce the representations that share the similar sparse pattern with respect to their own dictionary. The inference is reached in terms of the total reconstruction error accumulated from all tasks.

To verify the performance, target recognition under standard operating conditions is performed. The results are compared with state-or-the-art algorithms, including AdaBoost, EMACH, and IGT. To diversify the experiment setting, target recognition under extended operating conditions is conducted, where both the depression angle and configuration are different between the images used for training and those used for testing. We also consider the limited training resource, and the capability of outliers rejection. From the experiment results, we are able to come the following conclusions.

- The monogenic signal representation could effectively capture the characteristics of SAR image.
- The proposed method could exploit the inter-correlation among the component-specific linear regression models.
- The proposed method could deal with the non-literal experimental setup, *e.g.*, depression and configuration variation. It performs much better than the reference algorithms, especially under extended operating conditions.
- The proposed method is capable to detect the outliers, for example, the confusers and clutters. Given the same false alarm rate, the proposed method could achieve better detection performance than the reference methods.

In the future, more attention will be paid to extend the multi-task joint sparse representation technique into some related fields, *e.g.*, target detection, super-resolution, SAR imaging, image restoration.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61201338 and 61401477. The authors would like to thank the handling Associate Editor and the anonymous reviewers for their valuable comments and suggestions for this paper.

## REFERENCES

[1] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, 2001.
[2] ——, "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.
[3] ——, "The monogenic scale space on a rectangular domain and its features," *Int'l J. Comput. Vis.*, vol. 64, no. 2, pp. 187–201, 2005.
[4] L. Wietzke and G. Sommer, "The conformal monogenic signal," in *DAGM 2008, Lecture Notes in Computer Science, 5096*, pp. 527–536.
[5] G. Demarcq, L. Mascarilla, M. Berthier, and P. Courtellemont, "The color monogenic signal: Application to color edge detection and color optical flow," *J. Math. Imag. Vis.*, vol. 40, pp. 269–284, 2011.
[6] S. C. Olhede and G. Metikas, "The monogenic wavelet transform," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3426–3441, Sep. 2009.
[7] M. Unser, D. Sage, and D. V. 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.
[8] M. Unser and D. V. D. Ville, "Wavelet steerability and the higher-order riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 2042–2058, Mar. 2010.
[9] L. Zhang and H. Li, "Encoding local image patterns using Riesz transforms: With applications to palmprint and finger-knuckle-print recognition," *Image and Vision Computing*, vol. 30, pp. 1043–1051, Sep. 2012.
[10] M. Yang, L. Zhang, S. Shiu, and et al., "Monogenic binary coding: An efficient local feature extraction approach to face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1738–1751, Dec. 2012.
[11] X. Huang, G. Zhao, W. Zheng, and M. Pietikinen, "Spatiotemporal local monogenic binary patterns for facial expression recognition," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 476–480, May 2012.
[12] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.
[13] D. Geiger, T.-L. Liu, and M. J. Donahue, "Sparse representations for image decompositions," *Int'l J. Comput. Vis.*, vol. 33, no. 2, pp. 1–24, Sep. 1999.
[14] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 609–616.
[15] 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.
[16] J. Gemmeke and B. Cranen, "Using sparse representations for missing data imputation in noise robust speech recognition," in *Proc. Eur. Signal Processing Conf.*, Aug. 2008.
[17] K. Varshney, M. etin, J. Fisher, and A. Willsky, "Sparse representation in structured dictionaries with application to synthetic aperture radar," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3548–3561, Aug. 2008.
[18] J. Wright, A. Yang, A. Ganesh, S. 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.
[19] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
[20] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.
[21] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1043, Jun. 2010.
[22] H. Zhang, N. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2495, Jul. 2011.
[23] 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.
[24] G. Dong and G. Kuang, "Target recognition in SAR images via classification on Riemannian manifolds," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 199–203, Jan. 2015.
[25] G. Dong, G. Kuang, L. Zhao, J. Lu, and M. Lu, "Joint sparse representation of monogenic components: with application to automatic target recognition in SAR imagery," in *Proc. IEEE Int'l Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 549–552.
[26] R. Caruana, "Multi-task learning," *Machine Learning*, vol. 28, pp. 41–75, 1997.

[27] X. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.

[28] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[29] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[30] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.

[31] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[32] J. Thiagarajan, N. Karthikeyan, K. Peter, A. Spanias, and V. Berisha, "Sparse representation for automatic target classification in SAR images," in *Int'l Sym. Communcitaion, Control and Signal Processing*, 2010, pp. 1–4.

[33] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.

[34] V. Berisha, N. Shah, D. Waagen, H. Schmitt, S. Bellofiore, A. Spanias, and D. Cochran, "Sparse manifold learning with applications to SAR image classification," in *Proc. IEEE Conf. Acoustic, Sonar and Signal Processing (ICASSP)*, 2007, pp. 1089–1092.

[35] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

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

[37] A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[38] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

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

[40] U. Srinivas, V. Monga, and R. G. Raf, "Adaptive boosting for SAR automatic target recognition." *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 591–606, Jan. 2014.

[41] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition." *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 112–125, Jan. 2007.

[42] V. Sierra-Vzquez and I. Serrano-Pedraza, "Application of Riesz transforms to the isotropic AM-PM decomposition of geometrical-optical illusion images," *J. Opt. Soc. Am. A*, vol. 27, no. 4, pp. 781–796, 2010.

[43] G. Obozinski, B. Taskar, and M. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *J. Stat. Comput.*, vol. 20, no. 2, pp. 231–252, Jan. 2009.

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

[45] P. Zhao and B. Yu, "Stagewise Lasso," *J. Machine Learning Research*, vol. 8, pp. 2701–2726, 2007.

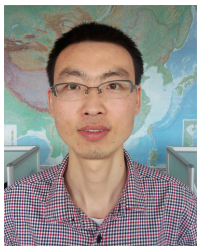
[46] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[47] Q. Zhao and J. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 643–654, Apr. 2001.

[48] R. Singh and B. Kumar, "Performance of the extended maximum average correlation height filter and the polynomial distance classifier correlation filter for multiclass SAR detection and classification," in *Algorithms for SAR Imagery IX, Proc. SPIE*, vol. 4727, Aug. 2002.

[49] L. Zhang, W. Zhou, P. Chang, and *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.

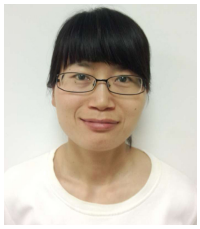
[50] A. Y. Yang, Z. Zhou, A. Ganesh, S. S. Sastry, and Y. Ma, "Fast $\ell_1$-minimization algorithms for robust face recognition," *arXiv:1007.3753*, Aug. 2012.

**Ganggang Dong** received the B. Eng. degree in UAV application engineering from the Artillery Academy, Hefei, China, in 2004 and the M.A.Eng. degree in information and communciation engineering from the National University of Defense Technology, Changsha, China, in 2012. He is currently working toward the Ph.D. degree in the National University of Defense Technology.

His research interests include the applications of compressed sensing and sparse representations, SAR image interpretation, data fusion and filter banks.

**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.



**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.

**Lingjun Zhao** received the B.S. degree in information engineering and the M.S. degree in circuits and system and the Ph.D. degree in communication engineering from the National University of Defense Technology, Changsha, China, in 2003, 2004, and 2007 respectively. She has been with the Remote Sensing Information Processing Lab in National University of Defense Technology, Changsha, since 2007. Her research interests include SAR image classification and urban SAR image interpretation, particularly automated extraction of urban objects.

**Jun Lu** received the B.S. degree in marine engineering from Northwestern Polytechnic University, Xian, in 1993, and the M.S. and Ph.D. degrees in electronic engineering from the National University of Defense Technology, Changsha, Hunan, China, in 1996 and 2000 respectively. He is an associate professor of the Remote Sensing Information Processing Lab, National University of Defense Technology, Changsha, China. His research interests are in the areas of SAR image processing, change detection, and classification with polarimetric SAR images.