IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 31, 2019, accepted November 9, 2019, date of publication November 13, 2019,
date of current version November 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953366

Joint Supervised Dictionary and Classifier Learning for Multi-View SAR Image Classification

HAOHAO REN, (Student Member, IEEE), XUELIAN YU, (Member, IEEE), LIN ZOU, (Member, IEEE), YUN ZHOU, AND XUEGANG WANG
University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Xuelian Yu (xlyu2012@uestc.edu.cn)

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61806046.

ABSTRACT A new multi-view sparse representation classification (SRC) algorithm based on joint supervised dictionary and classifier learning (MSRC-JSDC) is proposed for synthetic aperture radar (SAR) image classification. Unlike most existing sparse representation methods for SAR image classification, MSRC-JSDC learns a supervised sparse model from training samples by utilizing sample label information, rather than directly employs a predefined one. Moreover, a supervised classifier is jointly designed during dictionary learning, which can further promote the classification performance compared with unsupervised reconstruction based classifier. In the meantime, to enhance the representation capability of the sparse model, classification error is back propagated to the dictionary learning procedure to optimize dictionary atoms. In order to extract more recognition information from collected SAR images, a multi-view strategy is applied in testing stage. A new sparse constraint is introduced into multi-view sparse representation procedure so that both inner correlation and complementary information among multiple views can be extracted. This is helpful for alleviating the influence of SAR image's sensitivity on classification performance in such challenging scenarios as large depression variation and noise corruption. Extensive experiments on the moving and stationary target acquisition and recognition (MSTAR) database demonstrate that the proposed method is more robust and performs better than some state-of-the-art approaches.

INDEX TERMS Synthetic aperture radar (SAR), automatic target recognition (ATR), sparse representation based classification (SRC), dictionary learning.

I. INTRODUCTION

With the development of synthetic aperture radar (SAR) imaging technology, the potential application of high-resolution SAR images has attracted concern in many fields [1]. Among them, automatic target recognition (ATR) based on SAR images is a hot and meaningful research topic in the field of radar and remote sensing. In the past few years, many SAR ATR algorithms have been presented and most of them can achieve satisfactory recognition performance under standard operating condition (SOC) [2], [3]. These algorithms can be generally categorized into two types: template-based and model-based.

In template-based algorithms [4], the template library, which consists of a mass of samples collected from various scenarios, plays an important role for recognition tasks. The samples in the library are either SAR images or features extracted from SAR images. The identity of a query sample is decided according to a given matching criteria. Although template-based algorithms have been widely applied to SAR ATR in early days, they are somewhat inefficient due to requirement of large storage and complex computation. By contrast, model-based methods have gained increasing popularity in SAR ATR recently owing to its low storage, high efficiency and robustness. For instance, some physical and mathematic models such as scattering center [5], [6], non-negative matrix factorization [7], [8] and geometrical characteristics [9] have been studied and applied to SAR ATR. Some manifold learning algorithms are also

The associate editor coordinating the review of this manuscript and approving it for publication was Chengpeng Hao.

very popular in SAR ATR [10], [11]. Moreover, deep learning has developed rapidly in the last few years and been applied to ATR recently [12]–[15]. It is well known that deep learning and traditional machine leaning work with two different learning mechanisms. The former is a self-learning method whose learning capacity depends on network structure and parameters optimization, while the latter is to learn handcrafted features according to predetermined rule. Although deep learning has demonstrated its excellent performance in a lot of applications, it is facing with some challenges in SAR ATR such as limited training data and real-time processing requirement.

Although some remarkable advances have been achieved in SAR ATR, robust feature extraction is still an urgent and challenging issue [16], especially under various extended operating conditions (EOCs). To deal with this problem, people have studied and attempted numerous different methods. Sparse representation based classification (SRC) is among the most representative ones [17]. In [18], SRC is applied to SAR ATR for the first time and shows its effectiveness and huge potential. After that, various methods based on SRC have been proposed for SAR ATR. For example, a multi-scale monogenic signal based sparse representation method is proposed to extract different scale discriminant features from SAR images [19]. A multi-resolution based on joint sparse representation classification method is proposed in [20], and an information coupled sparse representation algorithm is proposed in [21] by fusing target image, target shadow and original image. A two-stage sparse representation algorithm is proposed by fusing manifold learning and sparse representation theory in [22].

In these SRC-based ATR algorithms, the sparse model is predefined and the dictionary is directly stacked using training samples. However, recent researches have shown that a sparse model which is learned from training samples can obtain better representation ability than a predefined one [23]–[26]. Based on this idea, some dictionary learning methods have been proposed for SAR ATR recently. For instance, label-consistent K singular value decomposition (LC-KSVD) [27] dictionary learning algorithm based on amplitude feature and scale-invariant feature transform (SIFT) descriptors fusion is proposed in [28]. Sparsity and Low-rank dictionary learning for monogenic signal is presented in [29].

More recently, several multi-view algorithms have been proposed to extract more recognition information from different views or sensors. Experimental results show their superiority over single-view ones in patter recognition areas. For example, in [30] and [31], feature learning from multiple views is integrated into dictionary learning procedure. Also, a multi-view low-rank dictionary learning algorithm is proposed to enhance the robustness in noise corruption environment [32]. However, most existing SAR ATR methods are still based on single view, yet recognition information extracted from single view is usually limited. Particularly, it is extremely challenging to extract discriminative information from single view under some EOCs. In the early days, several multi-view SAR ATR methods have been proposed. They firstly classify each view separately using traditional single-view model, and then use methods such as D-S evidence [33], Bayesian multi-view classifier [34] and principal component analysis (PCA) fusion strategy [35] to fuse all single-view classification results. These multi-view methods can achieve better performance than single view methods, however, they ignore the relationship among multiple views and mainly rely on single-view classification model.

To overcome the limitation of single-view recognition, joint sparse representation classification (JSRC) based on multi-view is presented for SAR ATR in [36]. JSRC can extract inner correlation among multiple views, which makes it perform better than previous approaches in most scenarios. However, JSRC restricts the interval of multiple views into a small range, which is difficult to realize in most real SAR scenarios. To improve this deficiency, multi-view SAR ATR with joint sparse representation over locally adaptive dictionary algorithm is proposed [37], [38], which can relax the restraint of JSRC to some extent. Nevertheless, it needs to update the locally adaptive dictionary online during testing stage, which is usually time-consuming and undesirable for practical applications. Moreover, inner correlation among randomly captured multiple views of the same target becomes weaker under EOCs. To sum up, current multi-view SAR ATR algorithms are still facing significant challenges under EOCs.

In this paper, we propose a new multi-view SRC algorithm based on joint supervised dictionary and classifier learning (MSRC-JSDC) for SAR image classification. In the training stage, a joint learning strategy is introduced to obtain a discriminative dictionary and classifier simultaneously. On one hand, a supervised dictionary is obtained by utilizing sample label information in training stage. Particularly, the dictionary learning process is optimized by making use of the back-propagation of classification error, which gives rise to a more discriminative sparse model than conventional ones. On the other hand, a supervised classifier is designed jointly during the dictionary learning, which is more superior to conventional reconstruction-based classifiers for SAR ATR. In the test stage, a new sparse constraint is introduced to multi-view sparse representation. More specifically, multiple views can flexibly select representation atoms from the learned dictionary at class-level, so that the proposed MSRC-JSDC is able to extract both inner correlation and complementary information among multiple views, which is beneficial to SAR ATR under EOCs.

The major contributions and innovations of our work are as follows:

- A new multi-view SRC algorithm based on joint dictionary and classifier learning is developed for SAR ATR.
- In our proposed algorithm, a supervised sparse model and classifier is jointly learned by using the class label information of training samples.

• A multi-view strategy is applied to extract more recognition information in testing phase. To be specific, a new sparse constraint is introduced into the multiple views sparse representation procedure to flexibly select representation atoms at class-level so that both inner correlation and complementary information among multiple views can be extracted.

• One biggest difference between our method and the better-known LC-KSVD lies in that our method utilizes classification error to adjust dictionary learning so that a more discriminative sparse model can be obtained, while LC-KSVD does not.

• Compared with the available methods designed for SAR ATR, the experimental results demonstrate that our proposed MSRC-JSDC obtains higher recognition performance on the benchmark MSTAR database and show more robustness under many challenging scenarios such as large depression angle variation, configuration recognition, noise corruption and so forth.

The remainder of this paper is organized as follows. In Section II, related work is reviewed. In Section III, the proposed MSRC-JSDC is introduced in detail. Extensive experiments are carried out on the MSTAR database in Section IV. Section V draws the conclusions.

II. RELATED WORK

A. SPARSE REPRESENTATION BASED CLASSIFICATION (SRC)

Sparse representation is based on the assumption that a sample from i-th class can be approximately represented as a linear combination of dictionary atoms from the same class. Therefore, a sample $\mathbf{x}$ belonging to i-th class can be represented as

$$\mathbf{x} = \mathbf{D}^i\mathbf{a}^i \tag{1}$$

where $\mathbf{D}^i$ is the dictionary corresponding to the i-th class, and $\mathbf{a}^i$ is the sparse representation vector of $\mathbf{x}$ over $\mathbf{D}^i$.

However, the identity of $\mathbf{x}$ is usually unknown. In this case, $\mathbf{x}$ can be approximately represented with respect to the whole dictionary $\mathbf{D}=[\mathbf{D}^1, \mathbf{D}^2, \cdots, \mathbf{D}^C]$ by enforcing a sparsity constraint on the sparse representation vector $\mathbf{a}$, where C is the number of target types. Therefore, $\mathbf{x}$ is represented as

$$\begin{aligned}\hat{\mathbf{a}} &= \arg\min_{\mathbf{a}} \|\mathbf{a}\|_0 \\ &s.t.\ ||\mathbf{x} - \mathbf{D}\mathbf{a}||_2^2 \le \epsilon\end{aligned} \tag{2}$$

where $\|\mathbf{a}\|_0$ is the number of non-zero elements in $\mathbf{a}$ and ϵ is the allowed error tolerance. (2) is a non-deterministic polynomial (N-P) hard problem due to the existence of non-convex ℓ_0 norm constraint. Some greedy algorithms such as matching pursuit [39] and relaxed reformations [40] are employed to solve it.

After obtaining the sparse representation vector $\hat{\mathbf{a}}$, the identity of $\mathbf{x}$ is decided by computing which class dictionary could result in the minimum reconstruction error. The classification rule is defined as

$$identity(\mathbf{x}) = \arg\min_i \|\mathbf{x} - \hat{\mathbf{x}}^i\|_2 = \arg\min_i \|\mathbf{x} - \mathbf{D}\delta^i(\hat{\mathbf{a}})\|_2 \tag{3}$$

where $\delta^i(\hat{\mathbf{a}})$ keeps only elements associated with i-th class while sets others to zero in $\hat{\mathbf{a}}$.

B. JOINT SPARSE REPRESENTATION FOR CLASSIFICATION (JSRC)

Let $\mathbf{x}_i$ (i=1,...,M) be M views of an unknown target. These views can be reconstructed together over a predefined dictionary $\mathbf{D}$ according to the following model:

$$\begin{aligned}\{\hat{\mathbf{a}}\}_{i=1}^M &= \arg\min_{\{\mathbf{a}_i\}} \sum_{i=1}^M \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 \\ &s.t.\ ||\mathbf{a}_i||_0 \le K, \quad 1 \le i \le M\end{aligned} \tag{4}$$

Let $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M]$ and $\mathbf{A}=[\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_M]$, we can rewrite (4) as

$$\begin{aligned}\hat{\mathbf{A}} &= \arg\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \\ &s.t.\ \|\mathbf{A}\|_0 \le K\end{aligned} \tag{5}$$

where $\|\mathbf{A}\|_F$ represents the Frobenius norm of $\mathbf{A}$, $\|\mathbf{A}\|_0$ sums up the non-zero elements of $\mathbf{A}$ and K is the sparsity level. Clearly, the relationship among multiple views is not exploited in (5), since $\|\mathbf{A}\|_0$ is decomposable over each column (view).

In order to overcome the above mentioned limitation, one can assume that multiple views of the same target share a common sparse pattern at atom-level. Based on this assumption, multiple views can select the same atoms over the predefined dictionary for sparse representation while the coefficients corresponding to the same atom may be different for each view. This can be achieved by realizing the following optimization problem with $\ell_0\backslash\ell_2$ norm regularization [36] as

$$\begin{aligned}\hat{\mathbf{A}} &= \arg\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \\ &s.t.\ \|\mathbf{A}\|_{\ell_0\backslash\ell_2} \le K\end{aligned} \tag{6}$$

where $\|\mathbf{A}\|_{\ell_0\backslash\ell_2}$ is the mixed-norm of matrix $\mathbf{A}$, which can be solved by first applying ℓ_2-norm on each row of $\mathbf{A}$ and then applying ℓ_0-norm on the resulting vector. Greedy algorithms such as simultaneous orthogonal matching [41] and CoSOMP method [42] can be used to solve the problem in (6). After obtaining $\hat{\mathbf{A}}$, the identity of query target is assigned according to the minimum multi-view reconstruction error criteria, which is similar to SRC.

III. MULTI-VIEW SRC BASED ON JOINT SUPERVISED DICTIONARY AND CLASSIFIER LEARNING

In this section, we propose a new classification algorithm named multi-view SRC based on joint supervised dictionary and classifier learning (MSRC-JSDC) for SAR image classification. The whole algorithm can be split into two parts: (i) joint supervised dictionary and classifier learning in training stage; (ii) multi-view sparse representation in testing stage. The details are described as below.

A. JOINT SUPERVISED DICTIONARY AND CLASSIFIER LEARNING

Given a pair $(\mathbf{x}, y)$, where $\mathbf{x}$ is the training sample and y is the class label of $\mathbf{x}$. Suppose the total number of target types is C and $\mathbf{x}$ is from the c-th ($c = 1, \cdots, C$) class. According to the sparse representation theory, $\mathbf{x}$ can be represented as a sparse coefficient vector $\mathbf{a}^*$ over a perdefined dictionary $\mathbf{D}=[\mathbf{D}^1, \mathbf{D}^2, \cdots, \mathbf{D}^C]$ by enforcing the relaxed l_1-norm sparse constraint on $\mathbf{a}^*$ [43]

$$\mathbf{a}^*(\mathbf{x}, \mathbf{D}) = \arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1\|\mathbf{a}\|_1 + \frac{\lambda_2}{2}\|\mathbf{a}\|_2^2 \quad (7)$$

where λ_1 is a regularization parameter to balance the accuracy of reconstruction error and the sparsity level, λ_2 is also a regularization parameter that makes (7) strong convex.

As previously mentioned, a better sparse model can be obtained over a learned dictionary instead of a predefined dictionary. Besides, the label of each training sample is generally known, which can be utilized to learn a more discriminative sparse model. Also, a supervised classifier is more beneficial to classification tasks than unsupervised classifiers. Taking all these three factors into consideration, a joint supervised dictionary and classifier learning algorithm is introduced to obtain an optimal dictionary and classifier simultaneously.

Basically, we propose to solve the following minimum classification error problem:

$$\arg\min_{\mathbf{D},\mathbf{W}} f(\mathbf{D}, \mathbf{W}) + \frac{\upsilon}{2}\|\mathbf{W}\|_F^2 \quad (8)$$

where $\mathbf{D} = [\mathbf{D}^1, \cdots, \mathbf{D}^C]$ is the dictionary to be learned, $\mathbf{W} = [\mathbf{W}^1, \cdots, \mathbf{W}^C]$ is a set of classifier's parameters, υ is a regularization parameter to avoid overfitting, and $f(\bullet, \bullet)$ is a convex function defined as

$$f(\mathbf{D}, \mathbf{W}) = l(y, \mathbf{W}, \mathbf{a}^*(\mathbf{x}, \mathbf{D})) \quad (9)$$

where l is a classification loss function. Usually, l can be chosen as the square logistic function, or hinge loss function from support vector machines.

The main difficulty to solve (8) lies in the non-differentiability of $\mathbf{a}^*(\mathbf{x},\mathbf{D})$. In fact, $\mathbf{a}^*(\mathbf{x},\mathbf{D})$ can be obtained from (7). In this paper, we solve (7) by using least angle regression (LAR) [44]. Once $\mathbf{a}^*(\mathbf{x}, \mathbf{D})$ is obtained, the objective function in (8) becomes differentiable, and $\mathbf{D}$ and $\mathbf{W}$ can be computed iteratively using stochastic gradient descent (SGD) algorithm [45].

The differentiability of $f(\mathbf{D},\mathbf{W})$ on $\mathbf{D}$ and $\mathbf{W}$ is

$$\begin{aligned} \nabla_{\mathbf{W}} f(\mathbf{D}, \mathbf{W}) &= \nabla_{\mathbf{W}} l(y, \mathbf{W}, \mathbf{a}^*) \\ \nabla_{\mathbf{D}} f(\mathbf{D}, \mathbf{W}) &= -\mathbf{D}\beta^*\mathbf{a}^{*T} + (\mathbf{x} - \mathbf{D}\mathbf{a}^*)\beta^{*T} \end{aligned} \quad (10)$$

where β^* is an intermediate variable related to ($\mathbf{x}$, y), $\mathbf{W}$ and $\mathbf{D}$. With initial zero, β^* can be computed according to the following recursive formula,

$$\begin{aligned} \Lambda &\leftarrow \{j \in \{1, \cdots, p\} : \mathbf{a}^*_{|j|} \neq 0\} \\ \beta^*_\Lambda &= (\mathbf{D}_\Lambda^T\mathbf{D}_\Lambda + \lambda_2\mathbf{I})^{-1}\nabla_{\mathbf{a}_\Lambda} l(y, \mathbf{W}, \mathbf{a}^*) \end{aligned} \quad (11)$$

Algorithm 1 Joint Supervised Dictionary and Classifier Learning

Input: Initial dictionary $\mathbf{D} = [\mathbf{D}^1, \cdots, \mathbf{D}^C]$, classifier's parameters $\mathbf{W} = [\mathbf{W}^1, \cdots, \mathbf{W}^C]$, regularization parameters $\{\lambda_1, \lambda_2, \upsilon\}$, t_0, T_1, p.
Output: Learned dictionary $\mathbf{D}$ and classifier's parameters $\mathbf{W}$
for $t = 0$ *to* T_1 **do**
(1): Take a pair sample $(\mathbf{x}_t, y)$ from training set
(2): Compute $\mathbf{a}^*$ according to (7)
(3): Compute the active set: $\Lambda \leftarrow \{j \in \{1, \cdots, p\}: \mathbf{a}^*_{|j|} \neq 0\}$
(4): Compute β^*: Initialize $\beta^* = 0$
$\beta^*_\Lambda = (\mathbf{D}_\Lambda^T\mathbf{D}_\Lambda + \lambda_2\mathbf{I})^{-1}\nabla_{a_\Lambda} l(\mathbf{y}, \mathbf{W}, \mathbf{a}^*)$
(5): Choose the learning rate: $\rho \leftarrow \min(\rho, \rho\frac{t_0}{t})$
(6): Update the parameters by a projected gradient step:
$\mathbf{W} \leftarrow \mathbf{W} - \rho_t(\nabla_W l_s(\mathbf{y}, \mathbf{W}, \mathbf{a}^*) + \upsilon\mathbf{W})$
$\mathbf{D} \leftarrow \mathbf{D} - \rho_t(-\mathbf{D}\beta^*\mathbf{a}^{*T} + (\mathbf{x} - \mathbf{D}\mathbf{a}^*)\beta^{*T})$
end
Return D and **W**

where p is the dimensionality of the sparse representation vector $\mathbf{a}^*$. More detail of the derivation of (10) can be found in [24].

Now, $\mathbf{D}$ and $\mathbf{W}$ can be updated iteratively as below,

$$\begin{aligned} \mathbf{W} &\leftarrow \mathbf{W} - \rho_t(\nabla_{\mathbf{W}} l_s(\mathbf{y}, \mathbf{W}, \mathbf{a}^*) + \upsilon\mathbf{W}) \\ \mathbf{D} &\leftarrow \mathbf{D} - \rho_t(-\mathbf{D}\beta^*\mathbf{a}^{*T} + (\mathbf{x} - \mathbf{D}\mathbf{a}^*)\beta^{*T}) \end{aligned} \quad (12)$$

where ρ_t represents SGD learning rate, which can be chosen according to a heuristic rule [46].

After the iterative operations are completed on all training samples, we get the final $\mathbf{W}$ and $\mathbf{D}$, which guarantees an optimized dictionary and classifier according to (8). To accelerate the convergence of stochastic gradient descent algorithm, a classic mini-batch strategy is introduced into SGD.

An algorithm procedure for joint supervised dictionary and classifier learning is displayed in Algorithm 1, where the parameter T_1 represents the number of iterations of SGD, and t_0 is the initialization parameter associated with SGD learning rate.

B. MULTI-VIEW SPARSE REPRESENTATION

Let $\mathbf{X}= [\mathbf{x}_1,\mathbf{x}_2, \cdots,\mathbf{x}_M]$ be M views of the same target. In JSRC, it is assumed that all views select the same atoms for sparse representation. In other words, all these M views share a common sparse pattern at atom-level. This can be illustrated in Fig. 1(a), where the number of views is $M = 3$, each column is a sparse coefficient vector corresponding to a single view, and each block represents a coefficient value. A white block means zero while others mean non-zero, and the darker the block, the larger the sparsity value

As mentioned before, JSRC can extract inner correlation among multiple views, thus obtain better recognition

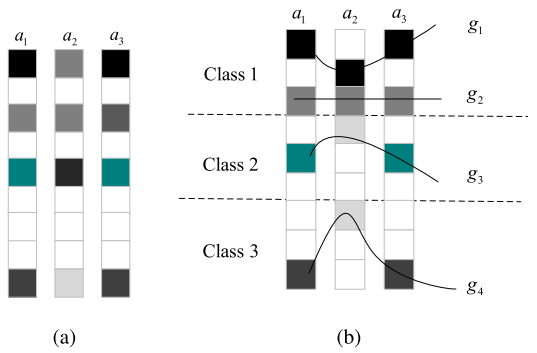


FIGURE 1. (a) sparse representation vectors share common sparse pattern at atom-level, (b) sparse representation vectors share similar sparse pattern at class-level.

performance than single view counterparts under SOC scenarios. However, inner correlation among multiple views of the same target becomes weaker under EOCs scenarios, which will degrade the performance of JSRC. In fact, multiple views of the same target contain some complementary information which may be useful for image classification but not utilized in JSRC.

In order to simultaneously extract inner correlation and complementary information among multiple views of the same target, we assume in MSRC-JSDC that multiple views share a similar sparse pattern at class-level. More specifically, multiple views can flexibly select representation atoms at class-level for sparse representation, which is depicted in Fig. 1(b) .

First of all, we define a new term called dynamic active set, $\mathbf{g}_s$=[$\mathbf{g}_s(1), \mathbf{g}_s(2), \cdots, \mathbf{g}_s(M)$], to represent the row indices of a set of coefficients corresponding to the same class in the coefficient matrix $\mathbf{A}$=[$\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_M$], where $s = 1, \cdots, K$, $\mathbf{g}_s(m)$ is for m-th column of $\mathbf{A}$ and K is the sparsity level.

Secondly, a new vector is formed by coefficients which are selected from $\mathbf{A}$ by $\mathbf{g}_s$, denoted as

$$\mathbf{b}_s = [\mathbf{A}(\mathbf{g}_s(1), 1), \mathbf{A}(\mathbf{g}_s(2), 2), \cdots, \mathbf{A}(\mathbf{g}_s(M), M)] \quad (13)$$

Then, for the sparse representation of M views, a new mixed norm is defined for $\mathbf{A}$ as below

$$||\mathbf{A}||_G = ||[||\mathbf{b}_1||_2, ||\mathbf{b}_2||_2, \cdots]||_0 \quad (14)$$

As one can see from (13) and (14), ℓ_2 norm is applied over $\mathbf{g}_s$ to sum up the strength of all atoms within a dynamic active set and then ℓ_0 norm across the ℓ_2 norm is employed to promote sparsity. Finally, the multi-view sparse representation used in MSRC-JSDC can be formulated as

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \quad s.t.\ ||\mathbf{A}||_G \leq K \quad (15)$$

Solving (15) is also a challenging task due to the existence of mixed-norm constraint. Herein, an iterative algorithm [47] is employed to solve problem (15), which is listed in Algorithm 2. In Algorithm 2, joint dynamic sparsity mapping (JDSM) is a core operation to implement atom selection. It is summarized in Algorithm 3.

Algorithm 2 Multi-View Joint Sparse Representation

Input: Muitlple views $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_M]$, the learned dictionary $\mathbf{D} = [\mathbf{D}^1, \cdots, \mathbf{D}^C]$, sparsity level K, number of views M.
Output: sparse coefficient vector $\mathbf{A}$=[$\mathbf{a}_1, \cdots, \mathbf{a}_M$]
Initialization: $\mathbf{R} = \mathbf{D}$ % Initialize the reconstruction error
while stopping criteria false **do**
  $\mathbf{E}$=$\mathbf{D}^T\mathbf{R}$;
  % (**i**) atom selection via joint dynamic sparsity mapping representation(JDSM)
  $\mathbf{I}_{new}$ = JDSM($\mathbf{E}$,2K)
  $\mathbf{I}$←[$\mathbf{I}^T$,$\mathbf{I}_{new}^T$] % (**ii**) index matrix updating
  %(**iii**) representation coefficients updating
  for $m = 1$ *to* M **do**
    $\mathbf{i} \leftarrow \mathbf{I}(:,m)$;
    $\mathbf{C}(\mathbf{i},m)$←$(\mathbf{D}(:,\mathbf{i})^T\mathbf{D}(:,\mathbf{i}))^{-1}\mathbf{D}(:,\mathbf{i})^T\mathbf{x}_m$;
  end
  %(**iv**) atom pruning via joint dynamic sparsity mapping
  $\mathbf{I}$← JDSM($\mathbf{C}$,K); $\mathbf{A}$←$\mathbf{0}$;
  for $m = 1$ *to* M **do**
    $\mathbf{i}$ ←$\mathbf{I}(:, m)$, $\mathbf{A}(\mathbf{i}, m)$=$\mathbf{C}(\mathbf{i}, m)$;
  end
  $\mathbf{R}$=$\mathbf{D}\mathbf{A}$-$\mathbf{X}$;
end
for $m = 1$ *to* M **do**
  $\mathbf{i}$ ←$\mathbf{I}(:, m)$, $\mathbf{A}(\mathbf{i}, m)$←$(\mathbf{D}(:,\mathbf{i})^T\mathbf{D}(:,\mathbf{i}))^{-1}\mathbf{D}(:,\mathbf{i})^T\mathbf{x}_m$;
end

C. CLASSIFICATION RULE

In section 3.1, a supervised classifier is jointly learned during the dictionary learning. Here, a weighted regression strategy is designed to assign the identity of M views $\mathbf{X}$= [$\mathbf{x}_1$,$\mathbf{x}_2$, $\cdots$,$\mathbf{x}_M$],

$$identity(\mathbf{X}) = \arg\max_c \sum_{i=1}^{M} \omega_i \mathbf{W}^c \delta^c(\mathbf{a}_i) \quad (16)$$

where the operation $\delta^c(\bullet)$ selects the coefficients corresponding to c-th class from a_i, and ω_i is the weight of i-th view. For simplicity, the parameter $\omega_i (i = 1, \cdots, M)$ is set to 1 in our experiments.

D. ALGORITHM FLOWCHART

The whole flowchart of MSRC-JSDC can be drawn in Fig. 2, where the upper is joint supervised dictionary learning in training stage, and the lower is multi-view sparse representation and classification in testing stage.

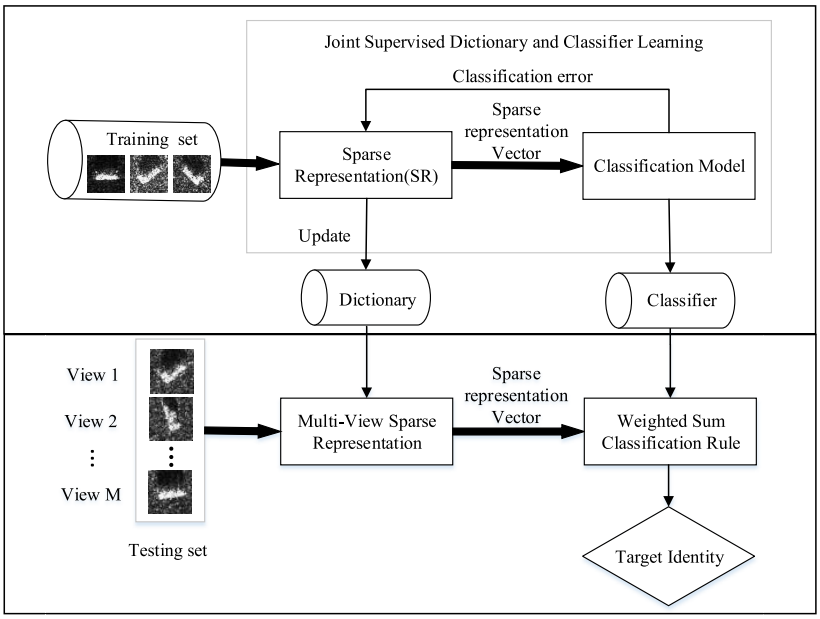


FIGURE 2. Flowchart of MSRC-JSDC.

Algorithm 3 Joint Dynamic Sparsity Mapping (JDSM)

Input: Sparse coefficient matrix $\mathbf{A}$, the number of dynamic active sets K, label vector $\mathbf{L}$ for atoms in the dictionary, number of classes C, number of views M

Output: Index matrix $\mathbf{I}_K$ for the top L dynamic active sets

Initialize: $\mathbf{I}_K \leftarrow \phi$ % initialize the index matrix as empty

for $k = 1$ *to* K **do**

  for $c = 1$ *to* C **do**

    $\mathbf{c} \leftarrow$ find($\mathbf{L}$,c) % get the index vector for c-th class

    for $m = 1$ *to* M **do**

      % (i) find the maximum absolute value v and it index t for the c-th class, m-th view

      $[v,t\ ] \leftarrow \max(|\mathbf{A}(\mathbf{c},m)|)$;

      $\mathbf{V}(c,m) \leftarrow v$, $\hat{\mathbf{I}}(c,m) \leftarrow \mathbf{c}(t)$;

      % (ii) combine the max coefficients for each class

      $\mathbf{s}(\mathrm{c}) \leftarrow \sqrt{\sum_{m=1}^{M} \mathbf{V}(c,k)^2}$;

    end

    % find the best cluster of atoms belonging to the same class across all the classes

    $[\hat{v}, \hat{t}] = \max(\mathbf{s})$;

  end

  $\mathbf{I}_K(k,:) = \hat{\mathbf{I}}(\hat{t},:)$, $\mathbf{A}(\hat{\mathbf{I}}(\hat{t},:)) \leftarrow \mathbf{0}$

end

E. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, the computational complexity of MSRC-JSDC is analyzed. Suppose the dimensionality of samples is greater than the number of samples, i.e., $d > N$.

In Algorithm 1 (joint supervised dictionary and classifier learning), the main computational complexity comes from the computation of $\mathbf{a}^*$ and the update of $\mathbf{D}$ and $\mathbf{W}$, which are realized by LAR and SGD, respectively. In each iteration of Algorithm 1, $\mathbf{a}^*$ solved by LAR takes at most $O(N^3)$ [44], and the process of updating dictionary and classifier with SGD [46] method costs $O(dN)$ respectively. If Algorithm 1 converges within T_1 iteration steps, the upper bound of the total computational complexity of Algorithm 1 is at most $O(T_1N^3)$. That is to say, the computation complexity of the proposed method takes at most $O(T_1N^3)$ in training phase. However, considering that training procedure is generally offline learning in ATR system, so the computational complexity in training phase has less impact on ATR system compared with the testing phase. In what follows, we will analyze the computational complexity of test phase in detail.

In testing phase, the proposed method is divided into two steps, i.e. multiple views sparse coding and classification based on coding coefficients. The major computational complexity of multiple views sparse coding focuses on JSDM in Algorithm2 . As for M views, if sparse level is pre-set to K, JSDM costs at most $O(MKCd^2)$ in each iteration [47]. If Algorithm 2 converges within T_2 iteration steps, sparse coding takes at most $O(T_2MKCd^2)$. In classification stage, if the size of dictionary is set to S, single view classification based on coding coefficients costs $O(S)$. The multi-views classification takes $O(MS)$. Therefore, the computational complexity of the proposed method is about $O(T_2MKCd^2)$ during testing. Referring to the relevant literatures [36], [37], the computational complexity of several multi-view methods during testing is summarized in Table 1 .

TABLE 1. Computational complexity of various methods during testing phase.

Method	JSRC	LAD-JSR	MSRC-JSDC
complexity	$O(T_2dlogd + dM)$	$O(T_2Mdlog^2d)$	$O(T_2MKCd^2)$

From Table 1, it is clear that the computation complexity of our proposed method has no advantage from the view of recognition performance improvement compared with other two multi-view method. This is due to the fact that the recognition performance and the computational complexity are in equilibrium with each other. To demonstrate this point in numerical testing, relevant experiment is conducted in section IV C.

IV. EXPERIMENTS AND DISCUSSION

To verify the validity of the proposed method, extensive experiments are conducted on MSTAR database, which is a gallery captured using a 10GHz SAR sensor with 0.3m × 0.3m resolution in range and azimuth. All data sets in the MSTAR database are released by the U.S. Defense Advanced Research Projects Agency and the U.S. Air Force Research Laboratory, and have been made publicly available [48]. SAR images in the MSTAR database are collected at various depression angles over the full aspect angles from 0°- 360°. Some SAR images and their corresponding optical images are shown in Fig. 3. The chip images are of around 128 × 128 pixels in size, and cropped to 64 × 64 pixels region of interest.

The MSTAR database consists of two sub-data sets, i.e., the standard operating condition (SOC) and various extent operating conditions (EOCs). Among them, EOCs includes some challenging scenarios, such as large depression variation, configuration variation and noise corruption scenarios and so forth. In the rest of this section, experiments are carried out under these two different conditions to test the effectiveness of the proposed method. Several state-of-the-art approaches, listed in Table 2, are also tested for performance comparison.

As for multi-view methods, including JSR, JSR-LAD, MSRC-FF and MSRC-JSDC, multiple views are randomly selected between 0° and 360° azimuth to form the test samples and experimental results are averaged on 10 repetitions. In the following experiments, the batch size is set to 100, $t_0 = T_1/10$, $\lambda_1 = 0.1$, $\lambda_2 = 0$, $\upsilon = 0.01$ and $p = 0.01$.

A. VISUALIZATION OF SPARSE REPRESENTATION RESULTS AND RECONSTRUCTION ERRORS

In this section, we visually display the sparse representation results and classification error of MSRC-JSDC, compared with single-view SRC and JSRC. Considering a three-class problem and the targets include 2S1, BRDM2 and ZSU234. A query image (chip number: 50) is taken from ZSU234, and another two views are also randomly sampled from the same target. Supposing that the chip number of randomly captured other two images are 10 and 300 in the MSTAR database respectively. In this test, the size of class dictionary of the proposed method is set to 100, and the sparse level is set to 10.

The visualization results are shown in Fig. 4, where the top row is sparse representation result and the bottom is the reconstruction error of each method. In Fig. 4 top, the position of each matchstick indicates that the atom located in this position is selected to represent the query sample, and the height of matchstick represents the sparse representation coefficient of testing sample on this atom. In Fig. 4 bottom, the height of each bar represents the reconstruction or classification error of testing sample on that target's dictionary, where red, green and blue bars correspond to 2S1, BRDM2 and ZSU234, respectively. The testing sample is classified to the target type whose error is minimum.

It can be seen from Fig. 4 top that most presentation atoms selected by JSRC and MSRC-JSDC come from ZSU234 while most atoms selected by SRC come from BRDM2. Consequently, SRC assigns a wrong identity (BRDM2) to the testing sample (which is from ZSU234), while the two multi-view methods, JSRC and MSRC-JSDC, can classify it correctly, as can be seen from Fig. 4 bottom. The results demonstrate that the two multi-view algorithms perform better than the single-view one. Moreover, it is worth noting that in MSRC-JSDC the differences among three classification errors are more distinct and larger than those in JSRC, which indicates more discriminative ability of MSRC-JSDC.

B. PERFORMANCE TEST UNDER SMALL DEPRESSION ANGLE VARIATION

In this section, we test the effectiveness of the proposed method under SOC scenario. In this scenario, images at depression angle 17° are selected as training set, while the ones at depression angle 15° are used for testing. Clearly, the depression angle variation between training set and testing set is small. The detail of experimental dataset is listed in Table 3. Following previous studies [36]–[38], the number of view is set to 3, and the size of the class dictionary of each target is set to 100 in MSRC-JSDC. Other parameters are set experimentally. The average recognition rate of each method is given in Table 4. Fig. 5 shows the confusion matrices of each method.

As can be seen from Table 4, all methods achieve an average recognition rate over 90% under SOC scenario. The reason is that most ATR algorithms can extract robust features under small target depression angle variation scenario. Nevertheless, it is clear that all multi-view approaches outperform single view methods, which indicates that multi-view joint recognition strategy is benefit for SAR ATR. Moreover, one can see that MSRC-JSDC and JSR-LAD get comparable recognition results. This is because that they can obtain more recognition information by using special dictionary construction strategy. In terms of average recognition performance, the proposed method is 7.34%, 5.28%, 4.08%, 3.13%, 0.62% better than SVM, SRC, LC-KSVD, JSRC, JSR-LAD, respectively.

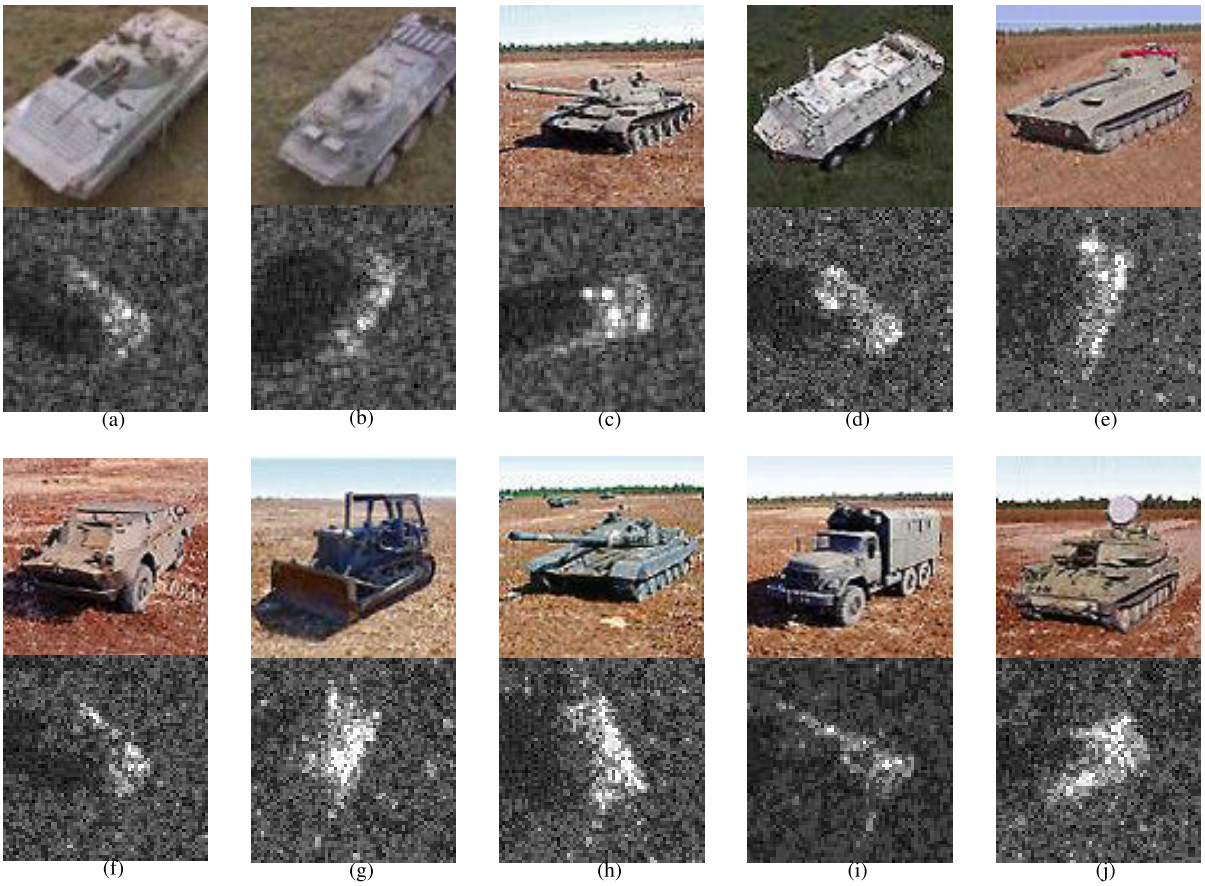


FIGURE 3. SAR images (top) and optical images (bottom), from (a) to (j): BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62, ZIL131, ZSU234.

TABLE 2. Methods to be tested for performance comparison.

Abbreviation.	description	Reference
SVM	Support Vector Machine	[49]
SRC	Sparse Representation-Based Classification	[18]
MSRC-FF	Multi-views SRC Based on PCA Feature Fusion	[35]
JSRC	Joint Sparse Representation Classification	[36]
JSR-LAD	Joint Sparse Representation over Locally Adaptive Dictionary	[37]
LC-KSVD	Label Consistent K-singular Value Decomposition	[27]

TABLE 3. The detail of dataset under SOC.

Depr.	BMP2	BTR70	T72	BTR60	2S1	BRDM2	D7	T62	ZIL	ZSU234
17°	233(SN9563)	233	232	256	299	298	299	299	299	299
15°	195(SN9563) 196(SN9566) 196(SNC21)	196	231(SN812) 195(SN812) 191(SNS7)	195	274	274	274	273	274	274

TABLE 4. Average recognition rate of each method under SOC.

Method	SVM	SRC	LC-KSVD	JSRC	JSR-LAD	MSRC-JSDC
recognition rate(%)	90.31	92.37	93.57	94.52	97.03	97.65

We can see from Fig. 5 that all methods are prone to confuse BMP2 and 72, perhaps because the two targets are somewhat similar in appearance, as shown in Fig. 3 . Also, it can be observed that single view methods are also difficult to distinguish BTR60 and BRDM2 correctly while multi-view methods improve this deficiency. By comparing the diagonal elements of each confusion matrix, one can see that MSRC-JSDC can obtain over 90% recognition rate for

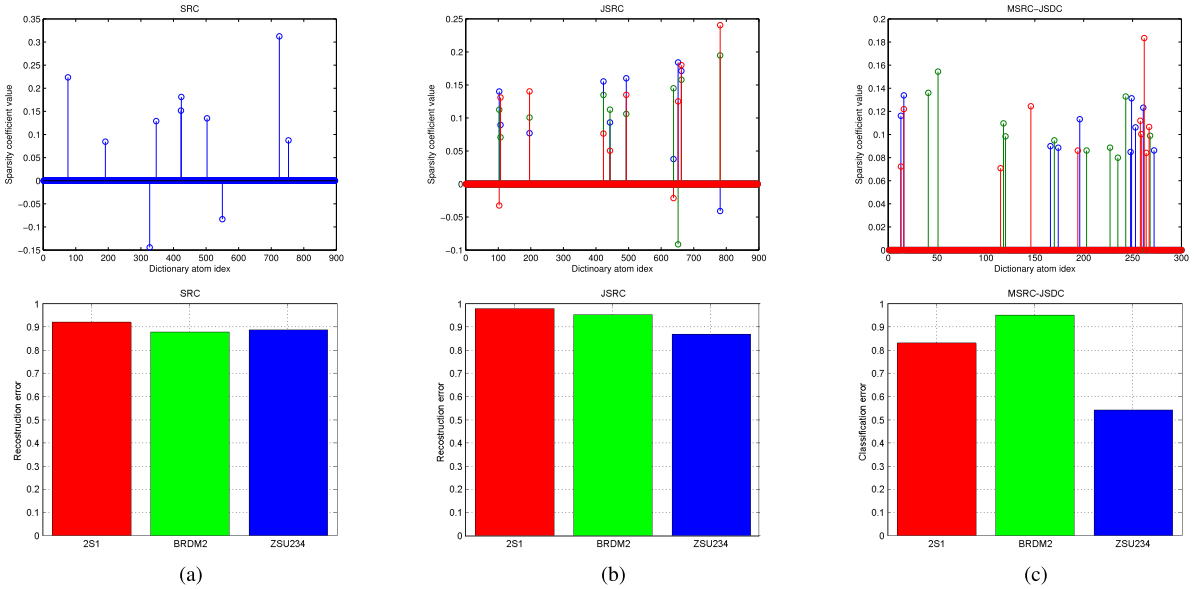


FIGURE 4. Top: sparse representation results, bottom: reconstruction errors. (a) SRC, (b) JSRC, (c) MSRC-JSDC.

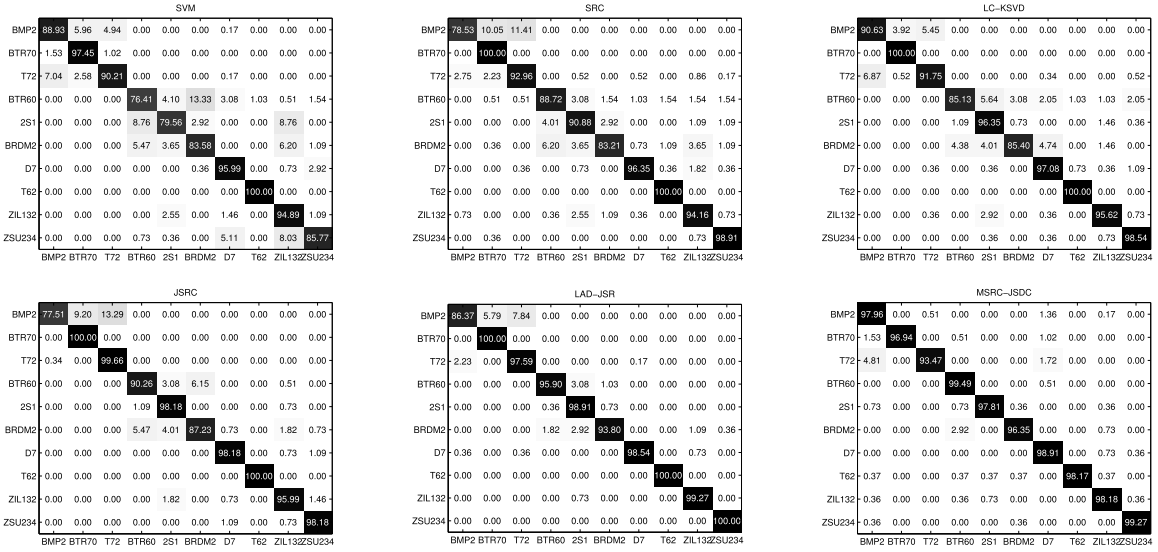


FIGURE 5. Confusion matrices obtained by the methods to be studied under SOC database. The vertical axis provides the ground-truth label, while the horizontal axis gives the predicted label, and the diagonal element in each graph shows the correct recognition accuracy of each target.

each target while other methods cannot, which indicates the proposed method is promising in SAR ATR.

C. PERFORMANCE TEST UNDER LARGE DEPRESSION VARIATION

It is well known that SAR images are sensitive to a variety of extended operating conditions (EOCs) scenarios including large depression variation, configuration variants and noise corruption. Performances of most SAR ATR algorithms degrade seriously under these EOCs scenarios. Consequently, SAR image classification under EOCs scenarios is a challenging issue.

In this section, a set of experiments are performed under large depression angle variation scenarios. Three military targets, 2S1, BRDM2, and ZSU234 are employed for experiments, among which BRDM2 and ZSU234 have some articulated variants in different scenarios. These target articulations and occlusions such as target open hatch or rotated gun turret are universal phenomena in real scenarios. Fig. 6 displays the standard and the articulated ZSU234 taken at the different depression angles.

In this scenario, SAR images are captured at different depression angles. Images at depression angle 17° are selected for training, while the ones at depression angles

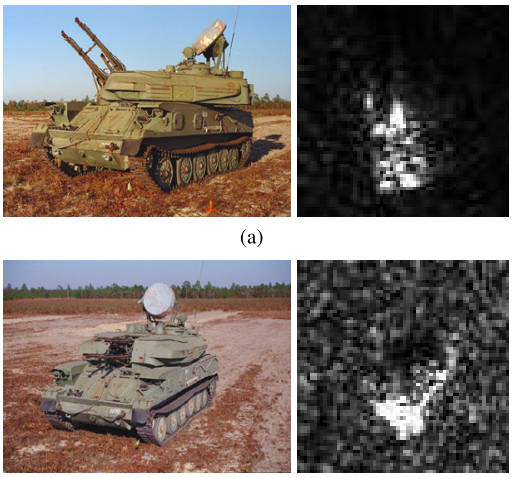


FIGURE 6. Exampler on the standard and the articulated scenario, (a) and (b) contain SAR images and optical image of ZSU234 with turret straight and articulated taken at different depression angle.

TABLE 5. The database under various depression angles.

Target type	2S1	BRDM2		ZSU234	
		scene1	scene2	scene1	scene2
train (17°)	299	298	298	299	299
test (30°)	288	287	133	288	118
test (45°)	303	303	120	303	118

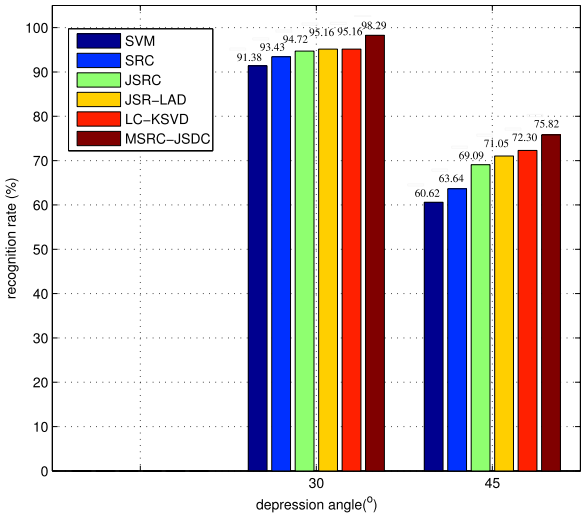


FIGURE 7. Recognition rate of various methods at different depression angles.

30° and 45° are for testing respectively. A summary of experimental dataset is listed in Table 5. The experimental setups are same as section 4.2. Fig. 7 plots the average recognition rate of each algorithm in the form of a histogram, where horizontal axis represents the depression angle of testing images, and vertical axis denotes recognition rate.

As can be seen from Fig. 7, when images at 30° depression angle are used for test, i.e., the depression angle changes between training set and testing set is 13°, all methods can get relatively satisfactory recognition results. This is due to that the global and local features of the SAR image can be still preserved at such a moderate depression angle change. Moreover, it can be observed that all SRC-based methods can achieve better performance than SVM. This is because that SAR images are sensitive to target pose and thus pose estimation procedure is usually necessary to traditional ATR algorithms, yet we didn't conduct any pose estimation in this paper. So, our experimental results further indicate the merits of SRC-based method in SAR ATR. We can also see that the proposed MSRC-JSDC performs better than competitors and achieves a recognition rate up to 98.29% with this moderate depression angle change.

TABLE 6. CPU running time of single view for SRC-based classification methods.

Method	SRC	MSRC-FF	JSRC	JSR-LAD	MSRC-JSDC
Time(s)	0.023	0.028	0.031	0.038	0.16

However, when images at 45° depression angle are used for test, i.e., the depression angle changes between training and testing set is 28°, all methods suffer dramatic performance degradation, as shown in Fig. 7 . This is due to the fact that it is very difficult to extract robust features from SAR images under such a large depression angle variation. In contrast, the proposed method can still achieve a recognition rate of 75.82% and perform better than the reference methods. Furthermore, it can be found that the recognition rate of MSRC-JSDC is 5.09% and 4.77% higher than those of JSRC and JSR-LAD, respectively. This is due to the fact that MSRC-JSDC can not only extract correlation information, but also mine complementary information between multiple views. When depression angle changes from 30° to 45°, the recognition rate degradation is 22.47% for MSRC-JSDC, compared to 22.86% for LD_KSVD, 24.10%for JSR-LAD, 25.63% for JSRC, 29.79% for SRC and 30.76% for SVM. Therefore, we can say that the proposed method is more robust to large depression angle variation than other methods.

SAR target recognition consists of two phases, i.e., offline training and online testing. Among them, the real-time performance of online testing phase is the most important for SAR target recognition system. Table 6 lists the average running time taken by each method to recognize single SAR image or a set of multi-view SAR images. All experiments are performed on personal computer (CPU: i7, RAM = 32G, Matlab2014b). It can be seen that MSRC-JSDC has no advantage over other SRC-based methods in time computational complexity. The main reason is that sparse coding based on new sparse constraint need to take more time to extract inner correlation and complementary information among multiple views. We all know the fact that time computation of algorithm generally increases as performance improves. Therefore, this testing result is also predictable. In the future work, we will try to make it run fast.

FIGURE 8. Three different configurations of T72 tank.

TABLE 7. Dataset of different configurations.

Target type	BMP2	T72	BTR60	T62
Train (17°)	233 (SN_9563)	232 (SN_132)	256	299
Test (15°)	196 (SN_9566)	195 (SN_812)	256	273
	196 (SN_C21)	191 (SN_S7)		

D. CONFIGURATION RECOGNITION TEST

In many real application scenarios, it is common to load different configurations such as gun, battery and oiltank to the same target. As is shown in Fig. 8, different configurations (gun, antenna and oiltank) are assembled on tank T72. In this section, we test the performance of the proposed method for different target configuration recognition. Four military targets, BMP2, T72, BTR60 and T62 are employed in this experiment, among which BMP2 and BTR60 are armored personnel carrier, while T72 and T62 are tank. BMP2 and T72 have several configurations with small structural modifications respectively. SAR images taken at 17° depression angle are selected as training set, while the other collected at 15° depression angle are used for testing. The experimental dataset is summarized in Table 7. Noted that SN_9563 for BMP2 and SN_132 for T72 are used for training while the other configurations (SN_9566, SN_C21, SN_812, SN_S7) are utilized for testing, so target configurations for testing are not contained in the training set. The experimental setups are same as section 4.2. The performance comparison of different approaches is shown in Fig. 9, where both the confusion matrices and the average recognition rate are given.

From Fig. 9, it can be seen that all methods other than SVM can achieve higher than 90% recognition rate. This further shows that SRC-based algorithms has some advantages in SAR configuration recognition task compared to conventional classification methods. The reason may be that the target to be recognized has only tiny local structural differences compared to the training dataset, while some global features such as target contour and shadow are still retained in this scenario. Furthermore, there is a similar phenomenon to Fig. 5 in IV B, it is still easy for all competitors to confuse the configurations of BMP2 and T72, yet MSRC-JSDC still performs well.

As for multi-view methods including MSRC-JSDC, JSRC and JSR-LAD, they can extract the correlation information embedded in multiple views, thus these approaches perform better than other single views methods, as can be seen in Fig. 9. MSRC-JSDC performs best and achieves a recognition rate of 99.33% on average, which is 2.66% higher than JSRC and 1.68% higher than JSR-LAD. Consequently, the experimental results indicate that the proposed MSRC-JSDC is effective for SAR target recognition under configuration variation.

TABLE 8. Type and sample number of training and testing sets.

Target type	2S1	BRDM2		ZSU234	
		scene1	scene2	scene1	scene2
train (17°)	299	298	298	299	299
test (30°)	288	287	133	288	118

E. EFFECT OF DICTIONARY SIZE

As we all known, collecting a large amount of SAR images is a challenging task. SAR ATR algorithms must remain considerably robust in case of limited training samples. In this experiment, we consider to investigate the effects of dictionary size on the recognition performance of MSRC-JSDC. A summary of the experimental dataset is given in Table 8. For comparison, other SRC-based algorithms including SRC, MSRC-FF, JSRC and JSR-LAD are also tested. The class dictionary size is set to 10, 20, 40, 80, 100, 120 and 150, respectively.

The recognition rate of each method versus dictionary size are plotted in Fig. 10. As can be seen, the performance of MSRC-JSDC performs better than the reference methods under all dictionary sizes. These results demonstrate that our proposed MSRC-JSDC is more robust with limited training samples. SRC performs the worst especially with small training samples, this is reasonable because that it has proved SRC needs a complete sample space to get good classification results [17].

As compared with LC-KSVD, which is also a dictionary learning method, one can see that MSRC-JSDC performs much better than LC-KSVD at all time. Here are two possible reasons. First, MSRC-JSDC updates the dictionary atoms iteratively according to the classification error, which helps to improve the representation ability of the learned dictionary. Second, making use of multiple views also plays an important role in MSRC-JSDC.

F. PERFORMANCE TEST UNDER DIFFERENT VIEW NUMBERS

In this section, we assess the effects of different view numbers on the performance of the proposed algorithm. Three other multiple views algorithms including MSRC-FF, JSRC and JSR-LAD are also tested for comparison. The experimental dataset is same as IV B. The number of views ranges from 2 to 10 and other parameters are set experimentally. The average recognition rate obtained by each method under different view numbers are summarized in Table 9. It can be seen clearly that MSRC-JSDC performs better than other reference methods under each view number. Moreover, the performance of MSRC-JSDC fluctuates slightly with the change of view numbers, while the recognition rate of JSRC fluctuates greatly. One possible reason is that the correlation

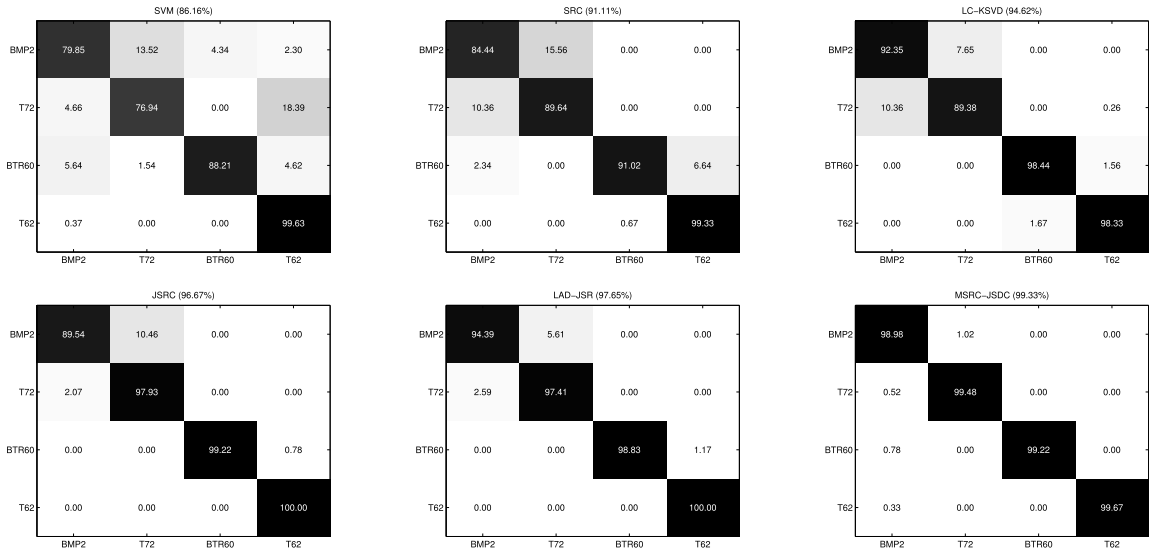


FIGURE 9. Confusion matrices and average recognition rate of various methods under different configurations.

TABLE 9. The average recognition rate (%) under different view numbers.

View(M)	2	3	4	5	6	7	8	9	10
JSRC	95.10	95.53	93.46	92.23	91.19	91.70	89.87	90.91	92.96
MSRC-FF	94.10	94.81	94.70	95.26	94.11	93.75	93.88	93.39	93.35
JSR-LAD	95.74	97.03	97.28	97.32	97.29	96.54	96.31	96.41	96.44
MSRC-JSDC	97.15	97.65	98.03	97.97	97.96	98.82	97.12	96.61	96.89

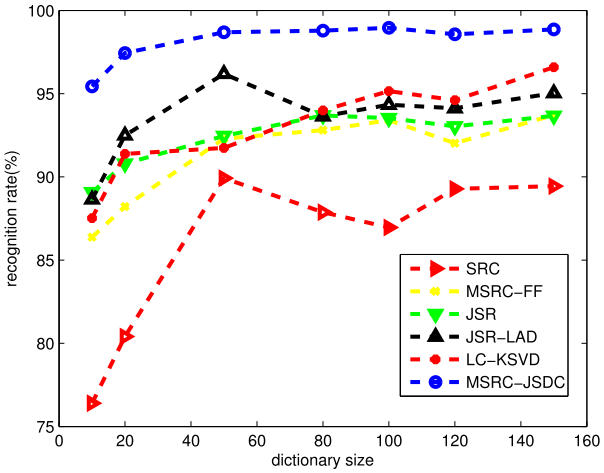


FIGURE 10. The average recognition rates for different dictionary size.

of randomly captured multiple views in JSRC is uncertain. If multiple views randomly captured have weak inner correlation, less useful information can be extracted from multiple views. Additionally, JSRC only extracts inner correlation while the proposed MSRC-JSDC utilizes both complementary information and inner correlation. MSRC-FF fuses multiple views by using PCA, which only extracts global features, so MSRC-FF has some limitations for multiple views recognition tasks.

It is worth noting that JSR-LAD can achieve comparable recognition results to our proposed MSRC-JSDC. This is expected since JSR-LAD can extract some differences and inner correlation information among multiple views through a two-level dictionary construction. However, MSRC-JSDC is still superior to JSR-LAD. One reason may be that MSRC-JSDC learns a more discriminative sparse model from training set while JSR-LAD does not. The experimental results demonstrate that MSRC-JSDC is more robust to view number than other reference methods for SAR ATR.

G. RECOGNITION UNDER NOISE CORRUPTION

In real situation, SAR images may be corrupted by complex noise coming from environment and radar system itself. Considering that the SNR (signal-to-noise ratio) of original SAR images in the MSTAR database is more than 30dB, it is necessary to add more noise to further evaluate the performance of the proposed method. Thus, complex additive white Gaussian noise (AWGN) is add to original SAR images from frequency domain according to the SNR which is defined below [36], [38].

$$SNR(dB) = 10\log_{10}\frac{\sum_{i=1}^{W}\sum_{j=1}^{H}|f(i,j)|^2}{HW\sigma^2} \tag{17}$$

where $f(i,j)$ is the complex frequency data corresponding to the original SAR image, H and W represent the size of the SAR image in frequency domain, and σ^2 is the variance of complex Gaussian noise.

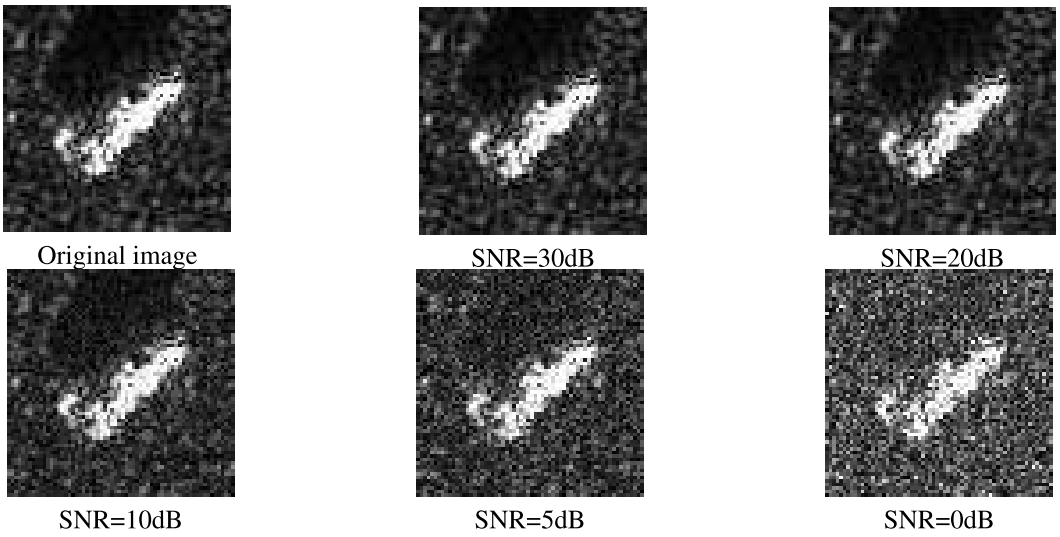


FIGURE 11. The corrupted SAR images at different SNR levels.

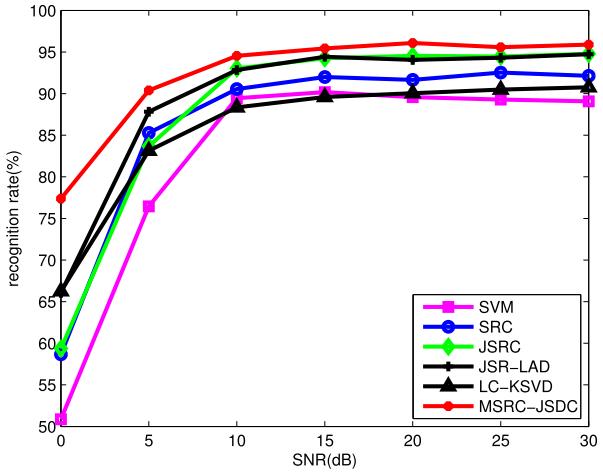


FIGURE 12. Recognition rates of different methods at different SNR levels.

Experimental dataset is same as IV B. Images for testing are corrupted by different levels of noise, while training images are not corrupted. A set of corrupted SAR images is displayed in Fig. 11.

Fig. 12 draws the recognition rate of various methods at different SNR levels. Apparently, all SRC based ATR algorithms outperform SVM at different SNR levels. This point has been proven that SRC based methods are more robust than conventional methods toward noise corruption [38]. Broadly speaking, each method suffers performance degradation as SNR reduces from 30dB to 0dB. However, both two dictionary learning methods, MSRC-JSDC and LCKSVD, show slower performance degradation than other SRC-based methods, which further shows that it is necessary to learn a discriminant dictionary from training samples for sparse representation based ATR algorithm. Even at 0dB, MSRC-JSDC still achieves a recognition rate higher than 75%, while other reference methods suffer from dramatic performance degradation. Especially, SVM only achieves about 50% at 0dB. One can see that the performance of MSRC-JSDC consistently performs better than the reference methods at different SNR levels. Therefore, the experimental results demonstrate that our proposed algorithm is more robust to Gaussian white noise corruption.

Random noise is also common in SAR image. To add random noise to original SAR images, we randomly select a percent of pixels from each of testing image, and then replace their pixel values (intensity) with independent and identically distributed samples drawn from a uniform distribution $\mathcal{U}[0, u_{max}]$, where $u_{max}$is the largest possible pixel value. The percentage of corrupted pixels vary from 5%-30%. A set of images contaminated with different level of random noise is shown in Fig. 13.

Fig. 14 plots the recognition rate of each method versus different levels of noise corruption. Again, the proposed MSRC-JSDC is consistently superior to all reference methods. And as we except, all SRC-based methods perform better than SVM. As the percentage of noise corruption increases, all methods suffer performance degradation.

However, the degradation trends of the two dictionary learning methods, LC-KSVD and MSRC-JSDC, are obviously slower than others. This again demonstrates the importance of dictionary learning for SAR ATR.

In a word, experimental results in this section demonstrate that our proposed MSRC-JSDC is effective in noise corruption environment and more robust compared with other reference methods.

H. RECOGNITION PERFORMANCE COMPARISON WITH DEEP LEARNING ALGORITHMS

In this section, we compare the proposed method with three deep learning based methods applied to SAR

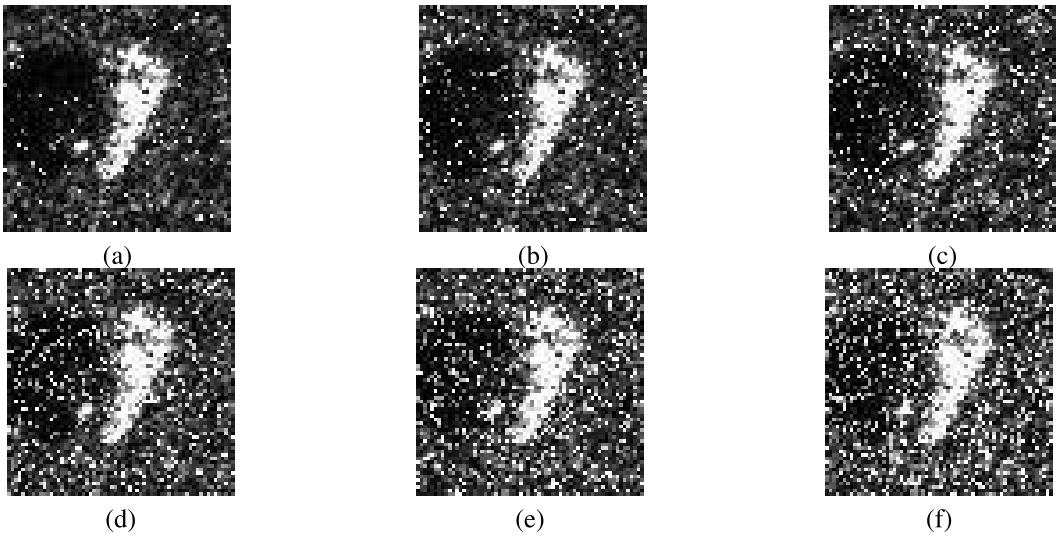


FIGURE 13. The corrupted SAR images at different levels of random noise: (a):5%, (b):10%, (c):15%, (d):20%, (e):25%, (f):30%.

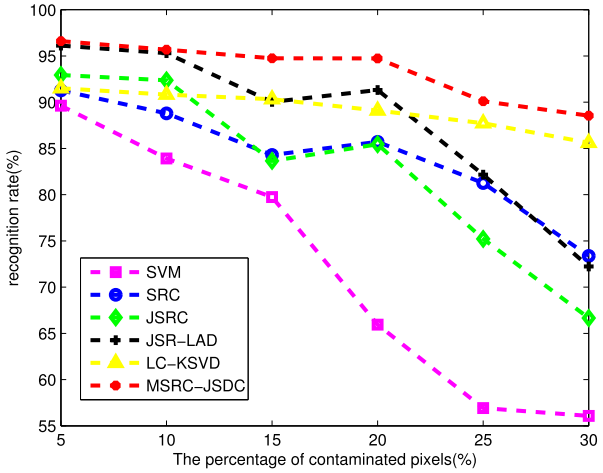


FIGURE 14. Recognition rates of different methods at different corruption levels.

target recognition. These methods are multi-view deep convolution neural network (VDCNN) [50] and two currently published deep convolution neural networks (DCNNs) [13], [51]. The experimental setup is same as section IV B. Recognition rate for each method under SOC scenario are listed in Table 10, In Table 10, the results are cited from the related literature [13], [50].

As can be seen from Table 10, the recognition performance of our proposed MSRC-JSDC is comparative to 3-VDCNN, and is superior to the other two DCNN methods. The corresponding analysis is summarized as follows.

(1) As for multi-view methods, i.e., MSRC-JSDC and 3-VDCNN, the performance of MSRC-JSDC is slightly inferior to 3-VDCNN. One reason is in 3-VDCNN that the aspect angle interval between two adjacent SAR views is limited to a certain range so that more discriminant features can be

TABLE 10. Recognition performance of various methods.

Methods	recognition rate(%)
DCNN [51]	92.30
DCNN [13]	94.10
3-VDCNN [50]	98.17
MSRC-JSDC	97.65

extracted easily while MSRC-JSDC does not. However, this strict constraint of 3-VDCNN on view interval is impractical especially in non-cooperative SAR target recognition scenarios. On the contrary, our proposed MSRC-JSDC can jointly recognize a set of randomly collected multiple views, which is very flexible for SAR target recognition.

(2) Compared with other two deep learning methods, MSRC-JSDC obviously outperforms two DCNN. This may be attributed to that MSRC-JSDC can extract more recognition information from multiple views, while DCNN can only obtain limited information from single SAR image at a time. Therefore, our proposed method is superior to some popular deep learning based approaches in SAR target recognition field.

V. CONCLUSION

This paper proposes a new multi-view SRC algorithm based on joint supervised dictionary and classifier learning, called MSRC-JSDC, for SAR image classification. MSRC-JSDC can learn a supervised dictionary from training samples by utilizing label information, which helps to improve the classification performance compared with the predefined dictionary. Moreover, the dictionary learning procedure is adjusted by the back-propagation of classification error so that the representation capability of the dictionary can be further enhanced. In the meantime, a supervised classifier is jointly designed during dictionary learning procedure,

which can again enhance the classification performance. Besides, by flexibly selecting dictionary atoms at class-level for multi-view sparse representation, MSRC-JSDC can mine both inner correlation and complementary information among multiple views of the same target, which significantly improves recognition performance, especially in various EOCs scenarios. Extensive experiments are performed on the MSTAR database, and some state-of-the-art methods such as SVM, SRC, JSRC, JSR-LAD, MSRC-FF and LC-KSVD are employed for comparison. Experimental results demonstrate that the proposed method has better performance than the reference methods and is more robust to depression variation, configuration variants, view number, dictionary size, and noise corruption.

ACKNOWLEDGMENT

The authors would like to appreciate the editor and all reviewers for their valuable suggestions and constructive comments.

REFERENCES

[1] S. Chen, Y. Yuan, S. Zhang, H. Zhao, and Y. Chen, "A new imaging algorithm for forward-looking missile-borne bistatic SAR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1543–1552, Apr. 2016.

[2] B. Bhanu, D. E. Dudgeon, E. G. Zelnio, A. Rosenfeld, D. Casasent, and I. S. Reed, "Guest editorial introduction to the special issue on automatic target detection and recognition," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 1–6, Jan. 1997.

[3] P. Zhao, K. Liu, H. Zou, and X. Zhen, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sens.*, vol. 10, no. 9, p. 1473, Sep. 2018.

[4] Q. H. Pham, A. Ezekiel, M. T. Campbell, and M. J. T. Smith, "New end-to-end SAR ATR system," *Proc. SPIE*, vol. 3721, pp. 292–302, Aug. 1999.

[5] Z. Jianxiong, S. Zhiguang, C. Xiao, and F. Qiang, "Automatic target recognition of SAR images based on global scattering center model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3713–3729, Oct. 2011.

[6] Y. Cong, B. Chen, H. Liu, and B. Jiu, "Nonparametric Bayesian attributed scattering center extraction for synthetic aperture radar targets," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4723–4736, Sep. 2016.

[7] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar Sonar Navigat.*, vol. 9, no. 9, pp. 1376–1385, 2015.

[8] S. Dang, Z. Cui, Z. Cao, and N. Liu, "SAR target recognition via incremental nonnegative matrix factorization," *Remote Sens.*, vol. 10, no. 3, p. 374, 2018.

[9] K. Tang, X. Sun, H. Sun, and H. Wang, "A geometrical-based simulator for target recognition in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 958–962, Sep. 2012.

[10] J. Pei, Y. Huang, W. Huo, J. Wu, J. Yang, and H. Yang, "SAR imagery feature extraction using 2DPCA-based two-dimensional neighborhood virtual points discriminant embedding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2206–2214, Jun. 2016.

[11] M. Liu, Y. Wu, Q. Zhang, F. Wang, and M. Li, "Synthetic aperture radar target configuration recognition using locality-preserving property and the Gamma distribution," *IET Radar, Sonar Navigat.*, vol. 10, no. 2, pp. 256–263, 2016.

[12] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[13] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

[14] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," *IEEE Trans. Geosci. Remote. Sens. Lett.*, vol. 14, no. 7, pp. 1091–1095, Jul. 2017.

[15] Y. Zhou, J. Shi, X. Yang, C. Wang, S. Wei, and X. Zhang, "Rotational objects recognition and angle estimation via kernel-mapping CNN," *IEEE Access*, vol. 7, pp. 116505–116518, 2019.

[16] K. El-Darymli, E. W. Gill, P. Mcguire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[17] J. Wright, A. Y. 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.

[18] J. J. Thiagarajan, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proc. 4th Int. Symp. Commun., Control Signal Process. (ISCCSP)*, Mar. 2010, pp. 1–4.

[19] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3316–3328, Jul. 2015.

[20] Z. Zhang and S. Liu, "Joint sparse representation for multi-resolution representations of SAR images with application to target recognition," *J. Electromagn. Waves Appl.*, vol. 32, no. 11, pp. 1342–1353, 2018.

[21] M. Chang and X. You, "Target recognition in SAR images based on information-decoupled representation," *Remote Sens.*, vol. 10, no. 1, p. 138, 2018.

[22] M. Liu, S. Chen, J. Wu, F. Lu, X. Wang, and M. Xing, "SAR target configuration recognition via two-stage sparse structure representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2220–2232, Apr. 2018.

[23] R. Jiang, H. Qiao, and B. Zhang, "Efficient Fisher discrimination dictionary learning," *Signal Process.*, vol. 128, no. 1, pp. 28–39, 2016.

[24] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[25] Y. Quan, Y. Xu, Y. Sun, and Y. Huang, "Supervised dictionary learning with multiple classifier integration," *Pattern Recognit.*, vol. 55, pp. 247–260, Jul. 2016.

[26] Y. Quan, Y. Xu, Y. Sun, Y. Huang, and H. Ji, "Sparse coding for classification via discrimination ensemble," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5839–5847.

[27] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[28] Y. Sun, L. Du, Y. Wang, Y. Wang, and J. Hu, "SAR automatic target recognition based on dictionary learning and joint dynamic sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1777–1781, Dec. 2016.

[29] G. Dong, N. Wang, G. Kuang, and H. Qiu, "Sparsity and low-rank dictionary learning for sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 141–153, Jan. 2018.

[30] F. Wu, X. Dong, L. Han, X.-Y. Jing, and Y.-M. Ji, "Multi-view synthesis and analysis dictionaries learning for classification," *IEICE Trans. Inf. Syst.*, vol. E102.D, no. 3, pp. 659–662, 2019.

[31] F. Wu, X.-Y. Jing, and D. Yue, "Multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries for image classification," *Neural Process. Lett.*, vol. 45, no. 2, pp. 649–666, 2017.

[32] F. Wu, X.-Y. Jing, X. You, D. Yue, R. Hu, and J.-Y. Yang, "Multi-view low-rank dictionary learning for image classification," *Pattern Recognit.*, vol. 50, pp. 143–154, Feb. 2016.

[33] R. Huan and Y. Pan, "Decision fusion strategies for SAR image target recognition," *IET Radar, Sonar Navigat.*, vol. 5, no. 7, pp. 747–755, Aug. 2011.

[34] M. Z. Brown, "Analysis of multiple-view Bayesian classification for SAR ATR," *Proc. SPIE*, vol. 5095, pp. 265–275, Sep. 2003.

[35] R. Huan and Y. Pan, "Target recognition for multi-aspect SAR images with fusion strategies," *Prog. Electromagn. Res.*, vol. 134, no. 134, pp. 267–288, Mar. 2013.

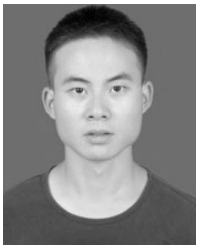
[36] H. Zhang, M. Nasser, and Y. Zhang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2497, Jul. 2012.

[37] Z. Cao, L. Xu, and J. Feng, "Automatic target recognition with joint sparse representation of heterogeneous multi-view SAR images over a locally adaptive dictionary," *Signal Process.*, vol. 126, pp. 27–34, Sep. 2016.

[38] B. Ding and G. Wen, "Exploiting multi-view SAR images for robust target recognition," *Remote Sens.*, vol. 9, no. 11, p. 1150, 2017.

[39] P. Knee, J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "SAR target classification using sparse representations and spatial pyramids," in *Proc. IEEE RadarCon (RADAR)*, May 2011, pp. 294–298.

[40] 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.
[41] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.
[42] M. F. Duarte, V. Cevher, and R. G. Baraniuk, "Model-based compressive sensing for signal ensembles," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep./Oct. 2009, pp. 244–250.
[43] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc., B (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.
[44] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.
[45] S. Bahrampour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins, "Multimodal task-driven dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 24–38, Jan. 2016.
[46] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[47] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-observation visual recognition via joint dynamic sparse representation," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 595–602.
[48] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant, "Standard SAR ATR evaluation experiments using the MSTAR public release data set," *Proc. SPIE*, vol. 3370, pp. 566–573, Sep. 1998.
[49] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 643–654, Apr. 2001.
[50] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.
[51] D. A. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," *Proc. SPIE*, vol. 9475, May 2015, Art. no. 94750F.



HAOHAO REN (S'16) was born in Gansu, China. He received the B.S. degree from the College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, China, in 2016. He is currently pursuing the Ph.D. degree in signal and information processing with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. His research interests include radar signal processing, synthetic aperture radar, target recognition, pattern recognition, and deep learning.

XUELIAN YU (M'10) was born in Henan, China. She received the B.S. and M.S. degrees in signal and information processing from the Xi'an University of Science and Technology, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2008. From 2008 to 2010, she held a postdoctoral position at the Department of Electronic Engineering, UESTC. From 2014 to 2015, she was a Visiting Scholar with the Department of Electronic Engineering, The Ohio State University. Since 2010, she has been an Associate Professor with the School of Information and Communication Engineering, UESTC. Her research interests include radar signal processing, synthetic aperture radar, target recognition, target tracking, and manifold learning.

LIN ZOU (M'16) was born in Sichuan, China. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1999, 2002, and 2013, respectively. From 2014 to 2015, he was a Visiting Scholar with the Department of Electronic Engineering, Linkoping University, Sweden. He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include radar signal processing, radar systems, and radar echo simulator.

YUN ZHOU was born in Sichuan, China. He received the Ph.D. degree in signal and information processing from the School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016. He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include radar signal processing, target recognition, and target tracking.

XUEGANG WANG was born in Hunan, China. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the School of Electronic Engineering, Xidian University, Xi'an, China, in 1984, 1987, and 1992, respectively. From 1992 to 1994, he held a postdoctoral position at the Department of Electronic Engineering, University of Electronic Science and Technology of China (UESTC). Since 2000, he has been a Professor with the Department of Electronic Engineering, UESTC. He is currently the Director of the Lab of Radar System and Digital Technology. His research interests include radar signal processing, target recognition and detection, radar echo simulator, and statistical signal processing.