Synthetic Aperture Radar Target Recognition Based on Multidimensional Sparse Model

Zhiqiang He*, Huaitie Xiao and Rang Liu
College of Electronic Science, National University of Defense Technology, China.

*E-mail: hzhqia@163.com

Abstract. This paper introduces a novel classification strategy based on multi-dimensional sparse model for target recognition in Synthetic Aperture Radar (SAR) image. This method constructs dictionaries for each dimension of SAR image instead of a single dictionary in traditional sparse model. In dictionary learning stage, this method adopts a dictionary-by-dictionary updating strategy. Under the premise of keeping the other dictionaries unchanged, each dictionary is updated using a traditional dictionary learning method. The sparsity coefficients of test samples in various categories of dictionaries are first calculated separately, and the test samples are classified using the sparse representation minimum error criterion. The experimental results on the MSTAR dataset show that compared with the traditional sparse models, this method can improve the target recognition rate while reducing the volume of dictionary. Therefore, in the field of SAR target identification, multidimensional sparse models have better recognition capabilities than traditional sparse models.

1. Introduction
Synthetic Aperture Radar (SAR) has been widely used in many fields such as resource exploration, marine monitoring, and target detection due to its advantages of being free from weather, time, and other factors. The application of various airborne and spaceborne SAR has rapidly increased the number of SAR data. The manual processing of SAR data is not only poor reliability but also low efficiency. Therefore, it has practical significance to study the automatic processing method of SAR data. This paper focuses on automatic target recognition (ATR) of SAR. With advances in computer technology and pattern recognition methods, machine learning methods such as support vector machine, sparse representation, and deep learning have achieved great success in target recognition field. In particular, deep learning method is widely used in applications such as face recognition and license plate recognition. However, the premise of excellent performance for deep learning is to provide a large amount of training data [1, 2], which is an insurmountable problem for SAR ATR.

Sparse representation uses the principle of orthogonal projection and can achieve high recognition performance. It has achieved great success in the areas of face recognition [3, 4] and SAR ATR [5, 6]. The classical sparse representation theory deals with vector signals. When sparse representation theory is applied to multidimensional signals such as images, the signals need to be vectorized first. This method includes the following disadvantages: (1) Signal vectorization introduces high frequency components and destroys internal structure. For SAR images, internal structure is an important means of distinguishing similar objects. Destruction of internal structural means reduced ability to identify. (2) Signal vectorization usually produces high dimensional data.
In order to overcome the shortcomings of the classic sparse representation theory when dealing with multidimensional data such as images, it can be extended to multidimensional sparse model based on tensor operation model. In [7], the Kronecker basis is used to calculate the sparse model of multidimensional signals; In [8] and [9], 2D-OMP method and 3D-OMP method are used to solve sparse coefficients of 2D and 3D signals, respectively; In [10], a 2D relaxed sparse model is achieved. Multidimensional sparse model realizes direct processing of multidimensional signal by constructing a dictionary for each dimension of signal. Although multidimensional sparse model increases the number of dictionaries, the total volume of dictionaries is greatly reduced compared to the classical sparse model. For multidimensional signals, the multidimensional sparse models have better representation capabilities than classical sparse models. In [11], multidimensional sparse synthesis model, multidimensional sparse analysis model and dictionary learning method are studied. These models are applied to image denoising, video denoising and multidimensional face recognition.

This paper studies a SAR target recognition method based on multidimensional sparse model. In our method, each type of samples includes two dictionaries. Each dictionary is updated by a dictionary-by-dictionary approach using the K-SVD algorithm. Sparse representation errors of test samples under each category of dictionaries are respectively calculated, and error minimum criteria is used to classify test samples. Experiments on MTSAR data show that multidimensional sparse model can improve the target recognition rate while significantly reducing the volume of dictionaries, whether classic sparse model using training samples to construct dictionaries directly or using dictionary learning method. Therefore, in the field of SAR ATR, multidimensional sparse models have better recognition capabilities than classic sparse models.

2. Multidimensional sparse model
Using tensor rule, the sparse model of N-dimensional signal \( \mathbf{x} \in \mathbb{R}^{l_1 \times l_2 \times \ldots \times l_N} \) is represented as n-mode product between N-dimensional sparse coefficient \( \mathbf{b} \in \mathbb{R}^{M_1 \times M_2 \times \ldots \times M_N} \) and N 2D dictionaries \( D_n \in \mathbb{R}^{N \times M_n} \), n = 1, 2, \ldots, N.

\[
\mathbf{x} = \mathbf{b} \times_1 D_1 \times_2 D_2 \times_n D_N, \text{s.t.} \| \mathbf{b} \|_0 = K \tag{1}
\]

Where \( K \) is sparsity, \( \| \mathbf{b} \|_0 \) is L0 norm, the number of non-zero elements in \( \mathbf{b} \). The n-mode product \( \mathbf{b} \times_n \mathbf{u} \) between \( \mathbf{b} \) and \( \mathbf{u} \in \mathbb{R}^{N \times M_n} \) is an N-dimensional matrix, whose size is \( M_1 \times \ldots \times M_n \times J_n \times M_{n+1} \times \ldots \times M_N \). Using \( B_{m_1 \ldots m_N} \) and \( U_{j m_{n+1}} \) to represent elements in corresponding matrices, the elements in \( \mathbf{b} \times_n \mathbf{u} \) are

\[
(B \times_n U)_{m_1 \ldots m_{n-1} j m_{n+1} \ldots m_N} = \sum_{m_n=1}^{M_n} B_{m_1 \ldots m_{n-1} j m_{n+1} \ldots m_N} U_{j m_{n+1}} \tag{2}
\]

According to the n-mode unfolding rule of tensor, (1) can be expressed as its equivalent form.

\[
X_{(n)} = D_n B_{(n)}(D_N \otimes \ldots \otimes D_{n+1} \otimes D_{n-1} \ldots \otimes D_1)^T \tag{3}
\]

2D matrices \( X_{(n)} \in \mathbb{R}^{l_1 \times l_2 \times \ldots \times l_N} \) and \( B_{(n)} \in \mathbb{R}^{M_1 \times M_2 \times \ldots \times M_N} \) represent the n-mode unfolding matrices of \( \mathbf{x} \) and \( \mathbf{b} \), which are defined by arranging all the n-mode vectors of \( \mathbf{x} \) and \( \mathbf{b} \) respectively. The n-mode vectors of \( \mathbf{x} \) are obtained by fixing every index but the one in the mode n. For (1), the objective function of \( \mathbf{b} \) is

\[
B = \arg \min_{\mathbf{b}} \| \mathbf{x} - \mathbf{b} \times_1 D_1 \times_2 D_2 \times_n D_N \|_0 + \lambda \| \mathbf{b} \|_0 \tag{4}
\]

Solving (4) directly involves complex tensor operations, we use vectorization of multidimensional matrices to transform (1) into a classical sparse model

\[
\text{vec}(\mathbf{x}) = D \text{vec}(\mathbf{b}), \text{s.t.} \| \text{vec}(\mathbf{b}) \|_0 = K \tag{5}
\]
\[ D = D_N \otimes D_{N-1} \otimes \cdots \otimes D_1. \] Using (5), we can obtain the objective function of \( B \) under L0 norm and L1 norm.

\[
B = \arg \min_{\text{vec}(B)} \left\| \text{vec}(X) - D\text{vec}(B) \right\|_2^2 + \lambda \left\| \text{vec}(B) \right\|_0
\]

(6)

\[
B = \arg \min_{\text{vec}(B)} \left\| \text{vec}(X) - D\text{vec}(B) \right\|_2^2 + \lambda \left\| \text{vec}(B) \right\|_1
\]

(7)

The greedy algorithm such as OMP and the convex relaxation algorithm such as Lasso can be used to solve (6) and (7), respectively. So far, the sparse model and its equivalent model of multidimensional signals have been established by using tensor operation rule and vectorization method.

3. Dictionary learning model of multidimensional signal

The main difference between multidimensional signal’s sparse model and one-dimensional signal’s sparse model is that the former contains multiple 2D dictionaries, and there is no clear correspondence between the atoms of each 2D dictionary and multidimensional signals. Therefore, it is not possible to construct dictionaries by directly combining multidimensional signals, and it is necessary using a dictionary learning method to construct each dictionary. For training set \([X^1, X^2, \cdots, X^S] \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}\), the purpose of dictionary learning is to obtain \( N \) 2D dictionaries \( \{D_n\}_{n=1}^N \in \mathbb{R}^{M_n \times M_n} \) under the premise that \([B_1, B_2, \cdots, B_S] \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times M_N} \) of training samples satisfies the sparsity limit. The objective function is

\[
\left\{ \{D_n\}_{n=1}^N, \{B^j\}_{j=1}^S \right\} = \arg \min_{\{D_n\}_{n=1}^N, \{B^j\}_{j=1}^S} \left\| \text{vec}(X^j) - D_1 \cdots \otimes D_N \right\|_F^2
\]

(8)

\[
s.t. \left\| B^j \right\|_0 \leq K, (1 \leq j \leq S), \left\| D_n(:, :r) \right\|_2 = 1, (1 \leq r \leq M_n)
\]

Referring to dictionary learning method in classical sparse model, an iterative convergence method is used to solve (8). Each iterative process involves sparse coefficient solving and dictionary updating. In the stage of sparse coefficient solving, the sparse coefficient \((B^j)^{S}_{j=1}\) of each training samples is solved using (6) or (7) by fixing all dictionaries \( \{D_n\}_{n=1}^N \). In the stage of dictionary updating, the process of updating all \( N \) dictionaries at the same time is very complicated, so a dictionary-by-dictionary updating strategy is adopted. Defining \( X^{(n)} = [X_1^{(n)}, X_2^{(n)}, \cdots, X_N^{(n)}] \in \mathbb{R}^{I_1 \times H_n} \), \( A^{(n)} = [A_1^{(n)}, A_2^{(n)}, \cdots, A_N^{(n)}] \in \mathbb{R}^{M_n \times H_n} \), \( H_n = S \left( \prod_{i=1}^N I_i \right) \), \( A^{(n)} = B^{(n)} (D_N \cdots \otimes D_1, D_0) \), and \( B'_{(n)} \) are n-mode unfolding matrices of \( X^j \) and \( B^j \), respectively. We use (3) to update dictionary \( D_n \). Its objective function is

\[
D_n = \arg \min_{D_n} \left\| X^{all}_n - D_n A^{all}_n \right\|_F^2, s.t. \left\| D_n(:, :r) \right\|_2 = 1, (1 \leq r \leq M_n)
\]

(9)

Formula (9) is the objective function of classical dictionary learning, which can be solved using the K-SVD method.

4. SAR target recognition using 2D sparse model

The object of SAR target recognition is 2D target image slice. At this time, the multidimensional signal \( X \) degenerates into 2D signal \( X \), and the sparse coefficient \( B \) degenerates to \( B \). The multidimensional signal’s sparse model also degenerates into 2D signal’s sparse model. The target recognition process includes dictionary training and sparse coefficient solving. Samples are classified according to the error of sparse representation. Assume that the sparse samples have a total of
category \( L \), the number of training samples per category is \( S_l, 1 \leq l \leq L \), and the training samples are represented as \( \{X^{(1)}_{1, i}, X^{(2)}_{1, i}, \ldots, X^{(L)}_{1, i}\} \). In the stage of dictionary training, each category of training samples \( \{X^{(1)}_{s, i}\}_{i=1}^{S_l} \) obtains dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \), and the dictionaries of all categories are represented as \([D^{(1)}_{1}, D^{(1)}_{2}], \ldots, [D^{(L)}_{1}, D^{(L)}_{2}]\) . In the stage of sparse coefficient solving, test sample \( X_t \) is sparsely expressed with each category of dictionaries and the corresponding errors of sparse representation \( \{e_1, e_2, \ldots, e_L\} \) are calculated, finally \( X_t \) is classified as the category corresponding to the minimum error of sparse representation. The SAR target recognition process is shown in Table 1 (Algorithm 1).

\[
\operatorname{Label}(X_t) = \arg\min_{j} (|e_j|_{l=1}^L)
\]

Table 1. Algorithm 1.

**Dictionary training**

**Input:** Training samples set \( \{X^{(1)}_{1, i}, X^{(2)}_{1, i}, \ldots, X^{(L)}_{1, i}\} \), maximum number of iterations num, dictionary length \( M_1, M_2 \), sparsity \( K \).

**Initialization:** Randomly initialize all categories of dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \) according to the dimension \( L \) of training samples and the length \( M_1 \) and \( M_2 \) of the dictionaries.

\[
\text{for } i = 1: \text{num}
\]

**Sparse coefficient solving:** For all categories of training samples, compute (6) by using greedy algorithm such as OMP or compute (7) by using convex relaxation algorithm such as Lasso, and obtain sparse coefficient \( \{[B^{(1)}_{1, i}]_{i=1}^{S_l}, [B^{(2)}_{1, i}]_{i=1}^{S_l}, \ldots, [B^{(L)}_{1, i}]_{i=1}^{S_l}\} \) of training samples.

**Dictionary updating:** Update dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \) by using (9) and \( \{[B^{(1)}_{1, i}]_{i=1}^{S_l}, [B^{(2)}_{1, i}]_{i=1}^{S_l}, \ldots, [B^{(L)}_{1, i}]_{i=1}^{S_l}\} \).

\[
\text{end for}
\]

**Output:** Dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \).

**Sparse Coefficient Solving and Classification**

**Input:** Testing samples set \( X_r \), dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \), sparsity \( K \).

**Sparse coefficient solving:** Use (6) or (7) to obtain the sparse representation coefficients of test sample \( X_r \) in dictionaries \( [D^{(1)}_{1}, D^{(1)}_{2}] \), respectively.

**Sample classification:** Calculate sparse representation errors \( \{e_1, e_2, \ldots, e_L\} \) of test sample \( X_r \),

\[
e_i = \|X_r - B_i \times D^{(1)}_1 \times D^{(2)}_2\|_p ,
\]

and \( X_r \) is classified as the category corresponding to the minimum sparse representation error.

**Output:** Recognition results of test samples.

5. Experiments and discussion

The experiments validate the performance of SAR target recognition using multi-dimensional sparse model and classic sparse model through the MSTAR data set. The MSTAR dataset includes multiple categories SAR image slices captured at various depressions over a full \( 0 \sim 360 \) range of aspect view.
The detailed information of image slices in the data set is shown in Table 2, where the entries in parentheses are the series number of variants with structural modifications. For ease of operation, all slice sizes were cut to $60 \times 60$ in the experiment. The slices taken at $17^\circ$ depression are used as training samples, and slices taken at $15^\circ$ depression are used as test samples. The experiments test the recognition capabilities of 2D sparse model and classic sparse model by using T1, T2 and T3 objects and all ten objects.

Table 2. Information of MTSAR data set.

| Depr | T1       | T2       | T3       | T4       | T5       | T6       | T7       | T8       | T9       | T10      |
|------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 17°  | 233(9563)| 232(132) | 299      | 298      | 256      | 299      | 299      | 299      | 299      | 299      |
| 15°  | 195(9563)| 196      | 195(812) | 195      | 274      | 274      | 273      | 274      | 274      |          |

5.1. 3-Target Recognition
Classic sparse model firstly vectorizes data when processing image signal. Since the number of target slices per category in the MSTAR dataset is between 200 and 300, directly vectorizing slices of size $60 \times 60$ results in vector dimension greater than the number of training samples. It is necessary to perform downsampling or dimensionality reduction while vectorizing, and this paper directly uses downsampling. The dictionary of classic sparse model can be composed directly from the training samples or by using dictionary learning method. (I) Directly using training samples to form dictionary $D \in \mathbb{R}^{I \times M}$: In order to guarantee $I \leq M$ in dictionary, the target slices are downsampled at sampling intervals of 3 and 4, respectively, and their sizes are changed to $20 \times 20$ and $15 \times 15$. These two methods are represented by SR1 and SR2, respectively. (II) Dictionary learning method. Each sub-dictionary is obtained by using dictionary learning method respectively, and total dictionary is formed by splicing each sub-dictionary. In order to guarantee $I \leq M$, in each sub-dictionary, the target slices are downsampled at sampling intervals of 4 and 5, respectively, and their sizes are changed to $15 \times 15$ and $12 \times 12$. The number of atoms in each sub-dictionary is 200. These two methods are represented by SR3 and SR4, respectively. 2D sparse model directly processes the target slices of size $60 \times 60$, and the size of dictionary is $60 \times 90$. This method is represented by TSR. The recognition results and the dictionaries volume of this five processing methods are shown in Table 3 and Table 4.

Table 3. Recognition results for three categories of targets

|        | SR1 (89.82) | SR2 (84.81) | SR3 (86.74) | SR4 (81.15) | TSR (93.74) |
|--------|-------------|-------------|-------------|-------------|-------------|
| T1     | 76.49       | 10.39       | 13.12       | 69.17       | 14.65       | 16.18       | 75.47       | 8.69       | 15.84       | 70.53       | 13.46       | 16.01       | 86.88       | 6.98       | 6.13       |
| T2     | 0.00        | 99.49       | 0.51        | 2.55        | 95.41       | 2.04        | 1.53        | 96.94       | 1.53        | 4.08        | 92.35       | 3.57        | 0.00        | 100.00      | 0.00       |
| T3     | 4.64        | 1.89        | 93.47       | 7.22        | 2.92        | 89.86       | 9.11        | 3.09        | 87.80       | 13.06       | 6.36        | 80.58       | 2.75        | 2.92        | 94.33       |

Table 4. Volume of dictionary for five methods

| Method | SR1   | SR2   | SR3   | SR4   | TSR   |
|--------|-------|-------|-------|-------|-------|
| Volume | 279200| 157050| 135000| 86400 | 32400 |

The percent of correct classification (PCC) with this five methods from high to low are TSR>SR1>SR3>SR2>SR4, TSR is 3.9%, 8.9%, 7.0%, and 12.6% higher than SR1, SR2, SR3, and SR4, respectively. The PCC of 2D sparse model is better than classic sparse model. In addition, the dictionary volume of TSR is the smallest, and the dictionary volume of the other four methods is much larger than TSR, which is 8.6 times, 4.8 times, 4.2 times, and 2.7 times the TSR, respectively. Therefore, TSR can achieve better recognition results under the premise of reducing the volume of
dictionary, and has obvious advantages over classic sparse models in the fields of two-dimensional and multi-dimensional signal recognition. The reason is that multidimensional sparse model processes the multidimensional signal directly, and retains signal structure information in recognition process. Classic sparse model requires vectorization of multidimensional signals, and this operation destroys signal structure. In addition, due to the limited number of training samples, downsampling and other dimensionality reduction operations can result in the loss of some data, which can lead to a decrease in recognition performance.

5.2. 10-Target Recognition
Use the methods in 5.1 to classify the ten targets in Table 1. The parameter setting for each method is the same as in 5.1. The recognition results and the volume of dictionaries are shown in Table 5.

Table 5. Ten categories of target recognition results

| Method | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | Average | Volume |
|--------|----|----|----|----|----|----|----|----|----|----|---------|--------|
| SR1    | 68.31 | 93.37 | 78.18 | 87.96 | 83.94 | **97.95** | 98.91 | 85.71 | 91.61 | **98.91** | 88.48 | 1098800 |
| SR2    | 54.68 | 83.67 | 68.73 | 77.01 | 75.55 | 87.18 | 97.08 | 75.09 | 84.31 | 96.35 | 79.96 | 618075  |
| SR3    | 62.18 | 89.29 | **69.24** | 77.74 | 66.06 | 88.72 | 94.53 | 80.95 | 82.12 | 95.95 | 80.68 | 450000  |
| SR4    | 52.47 | 79.08 | 69.07 | 49.64 | 54.38 | 74.36 | 88.69 | 60.07 | 63.50 | 92.70 | 68.40 | 288000  |
| TSR    | **87.73** | 95.92 | 67.18 | **93.43** | 86.50 | 94.36 | 98.18 | **91.21** | 93.80 | 97.45 | 90.57 | **108000** |

The PCC with this five methods from high to low are TSR>SR1>SR3>SR2>SR4, TSR is 2.1%, 10.6%, 9.9%, and 22.2% higher than SR1, SR2, SR3, and SR4, respectively. The corresponding dictionary volume SR1>SR2>SR3>SR4>TSR. The dictionary volume of the previous four methods is 10.2 times, 5.7 times, 4.2 times, and 2.7 times the TSR, respectively. Regardless of ten types of target recognition or three types of target recognition, the experimental results all show the same rule. That is, the recognition effect using the 2D sparse model is better than that of the classic sparse model, and the dictionary volume of the 2D sparse model is smaller than the classic sparse model. This is because the multidimensional sparse model can make full use of image local information and can improve the recognition rate while reducing the dictionary volume. Therefore, the multidimensional sparse model is more advantageous than the classic sparse model in identifying multidimensional data such as images.

6. Conclusion
This paper proposes a method for SAR ATR using multidimensional sparse models. This method combines sparse representation of multidimensional signals with tensor theory, and realizes the direct processing of images by constructing dictionaries for each dimension of the image respectively. It avoids the internal structure damage caused by vectorization of the classic sparse model and the downsampling operation that must be performed when the training sample is insufficient. Experiments on the MSTAR data set show that the multidimensional sparse model can improve target recognition rate while significantly reducing the volume of dictionary, and has better recognition performance than classic sparse model in the field of SAR ATR.

References
[1] Song Q, Xu F. 2017. Zero-Shot Learning of SAR Target Feature Space With Deep Generative Neural Networks[J]. IEEE Geoscience & Remote Sensing Letters. 14(12), 2245-2249.
[2] Li Y, Wang J, Xu Y, et al. 2017. DeepSAR-Net: Deep convolutional neural networks for SAR target recognition[C]. IEEE, International Conference on Big Data Analysis. 740-743.
[3] Wang J, Lu C, Wang M, et al. 2014. Robust Face Recognition via Adaptive Sparse Representation[J]. IEEE Transactions on Cybernetics. 44(12):2368-2378.
[4] John W, Allen Y, Arvind G, et al. 2009. Robust Face Recognition via Sparse Representation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 31(2):210-226.
[5] Karine A, Toumi A, Khenchaf A, et al. 2017. Target Recognition in Radar Images Using Weighted Statistical Dictionary-Based Sparse Representation[J]. IEEE Geoscience & Remote Sensing Letters. 14(12):2403-2407.

[6] Li M, Guo Y, Li M, et al. 2017. Coupled Dictionary Learning for Target Recognition in SAR Images[J]. IEEE Geoscience & Remote Sensing Letters. 14(6):791-795.

[7] Caiafa C F, Cichocki A. 2013. Computing sparse representations of multidimensional signals using kronecker bases[M]. MIT Press.

[8] Fang Y, Wu J J, Huang B M. 2012. 2D sparse signal recovery via 2D orthogonal matching pursuit[J]. Science China(Information Sciences). 55(4):889-897.

[9] Huo Y, Fang Y, Huang L. 2016. 3D sparse signal recovery via 3D orthogonal matching pursuit[J]. Journal of Systems Architecture. 64:3-10.

[10] Dong Q. 2013. Two-dimensional relaxed representation[J]. Neurocomputing. 121(18):248-253.

[11] Qi N, Shi Y, Sun X, et al. 2017. Multi-Dimensional Sparse Models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 40(1):163-178.