JOINT SPARSE REPRESENTATION OF MONOGENIC COMPONENTS: WITH APPLICATION TO AUTOMATIC TARGET RECOGNITION IN SAR IMAGERY

Ganggang Dong, Gangyao Kuang, Linjun Zhao, Jun Lu, Min Lu

School of Electronics Science and Engineering, National University of Defense Technology, China

ABSTRACT

In this paper, classification via joint sparse representation of the monogenic signal is presented for target recognition in SAR imagery. First, the monogenic signal is performed to capture the characteristics of SAR image. Since it is infeasible to directly apply the raw component to classification due to the high data dimension and redundancy, three augmented feature vectors are defined via uniform downampling of the real part, the imagery part, and the instantaneous phase. The monogenic features are then fed into a recently developed framework, sparse representation-based classification (SRC). Rather than produce individual sparse pattern, this paper generates the similar sparsity pattern for three feature vectors by imposing a mixed norm on the representation matrix. Extensive experiments on MSTAR database demonstrate that the proposed method could significantly improve the recognition accuracy.

***Index Terms*—** Joint sparse representation, target recognition, synthetic aperture radar, classification.

1. INTRODUCTION

Synthetic aperture radar (**SAR**) automatic target recognition (**ATR**) has been widely studied in the past, yet it is still an open problem due to the extended operating conditions, in which the operational parameter is significantly different between the images used for training and those used for testing. The Typical SAR ATR system recognizes the ground tactical targets through three sequential stages [1]: detection, discrimination and classification. In the first stage, all potential targets as well as numerous false alarm, (*e.g.*, buildings, civil vehicles, *etc.*) are detected; then the natural-clutter false alarms are eliminated in the second stages, followed by a classifier to reach the inference. Only the final stage, classification, is studied in this paper.

The past several years have witnessed a resurgent development of sparse signal representation. It has been applied to many fields, such as signal processing, machine learning and computer vision. In the former works, the promising performances have been reported. In [2], the well-known work of sparse representation-based classification (**SRC**) has been proposed. The training samples are utilized as the basis vectors to encode the test sample as a sparse linear combination of them. By limiting the feasible set with the sparsity constraint (*e.g.*, ℓ_1-norm minimization), the unique representation can be obtained. The decision is made according to the characteristics of representations on reconstruction. Due to the simple implementation and great performance, SRC has been also introduced into ATR framework [3, 4]. To circumvent the preprocessing and explicit pose estimate, Ref [4] employs the sparsity model for target classification in SAR imagery. Ref [5] employ a group sparse representation technique to exploit the inherent block structure of the sparse solution induced by ℓ_1-norm minimization. Moreover, Ref [6] presents a multi-view target recognition method, in which the multiple views of the same target are represented with a joint sparsity model. However, these techniques are implemented on the raw intensity, thus it may be not effective to portray SAR images, whose characteristics vary quickly and abruptly with small changes of pose and configurations because of the specular reflections of a coherent source. To improve the performance, Ref [7] suggests a novel sparsity model, in which the classification via sparse representation of the monogenic signal is proposed. To make the data dimension suitable for classification, an augmented feature vector is defined via concatenation of the downsampled components (*i.e.*, the real part, the imagery part, and the phase). It is actually the data fusion in the feature-level. The resulting descriptor is input into SRC to make the decision. However, it fails to exploit the correlations among the multiple components.

In this paper, the classification via joint sparse representation of the monogenic signal (**JSRC**) is proposed. The monogenic signal (multiresolution) is first performed to capture the characteristics of SAR image. To make the monogenic components suitable for classification, three augmented feature vectors are defined via uniform downsampling, normalization, and concatenation of three monogenic components, respectively. Three feature vectors are then input into SRC framework simultaneously. Rather than produce individual sparse pattern, this paper generates the similar sparse pattern for three feature vectors with a mixed norm imposed on the representation matrix. Thus it is capable to share some common information for discrimination, and the inter-correlations among multiple components could be exploited accordingly. Extensive experiments on real SAR imagery demonstrate that the proposed method could effectively combine the advantage of individual monogenic component.

2. THE PROPOSED METHOD

In the recognition problem, the goal is to infer the class of the unknown using the knowledge learned from a set of given samples for each class. Assuming that the training samples from a certain class span a linear subspace. The test can be well represented with the training samples of the same class. Since the class membership of the test is unknown initially, it is represented over all the training samples,

$$\mathbf{y} = \mathbf{x}_{1,1}\alpha_{1,1} + \cdots + \mathbf{x}_{i,j}\alpha_{i,j} + \cdots + \mathbf{x}_{K,n_K}\alpha_{K,n_K} = \mathbf{X}\alpha \quad (1)$$

where $\mathbf{X} = [\mathbf{x}_{1,1}, ..., \mathbf{x}_{K,n_K}] \in \mathbb{R}^{m\times n}$ contains all the training samples stacked by its column (called dictionary); K is the number of class, and $n = \sum_{j=1}^{K} n_j$ is the number of training samples.

2.1. Sparse representation

Considering the underdetermined system ($m < n$), the solution of (1) is not unique. To generate the unique representation, an intuitive idea is to seek the sparest representation with sparsity constraint [2],

$$\min_{\alpha} \|\alpha\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon \quad (2)$$

where $\|\cdot\|_0$ counts the number of nonzero entries; ε is the error tolerance.

However, solving (2) in underdetermined system is NP-hard and unstable. Thanks to the recent development in the theory of compressed sensing, if the representation α is sparse enough, the problem (2) can be also relaxed as ℓ_1-norm minimization,

$$\min_{\alpha} \|\alpha\|_1 \quad s.t \quad \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon \quad (3)$$

where $\|\alpha\|_1 = \sum_i \|\alpha_i\|$. The problem (3) is a convex program, and hence can be solved using standard linear programming methods in polynomial time.

With the representation $\hat{\alpha}$, the decision can be made by evaluation which class of training samples could result in the minimum reconstruction error,

$$\min_{k=1,...,K} \{\|\mathbf{y} - \mathbf{X}\delta_k(\hat{\alpha})\|_2^2\} \quad (4)$$

where $\delta_k(\cdot)$ selects the coefficients associated with the k-th class, and sets the others to zeros.

2.2. Joint sparse representation of monogenic signal

For 2-D signal $f(\mathbf{z})$, the monogenic signal [8] is defined as the combination of $f(\mathbf{z})$ and its Reisz transformed signal $f_R(\mathbf{z})$

$$f_m(\mathbf{z}) = f(\mathbf{z}) - (i,j)f_R(\mathbf{z}) \quad (5)$$

where i, j are the imagery units. As a complex number, the monogenic signal can be described in the form of local amplitude and local phase

$$\begin{aligned} A(\mathbf{z}) &= \sqrt{f(\mathbf{z})^2 + \|f_R(\mathbf{z})\|^2} \\ P(\mathbf{z}) &= atan2(f_R(\mathbf{z}), f(\mathbf{z})) \end{aligned} \quad (6)$$

With the representation (6), the signal can be represented by three distinct components, *i.e.*, the real part f, the imagery part f_R, and the instantaneous phase P. Due to the wrapping phenomena, the cosine value $\cos(P)$ is utilized instead of the raw phase coefficient.

Since the practical signal is of finite, it is necessary to perform band-pass filter. Following the former works [9], log-Gabor is used to generate multiresolution representation. Then the monogenic representation can be rewritten as

$$f_m(\mathbf{z}) = f(\mathbf{z}) * h_{lG} - (i,j)f_R(\mathbf{z}) * h_{lG}, \quad (7)$$

where h_{lG} is the log-Gabor kernel. Due to the high data dimension and redundancy, the raw components are infeasible to be applied to classification. The idea here is to generate three feature vectors via concatenation of the downsampled components (by a factor of ρ).

Denote by $\mathcal{T}_r, \mathcal{T}_i, \mathcal{T}_p \in \mathbb{R}^{d\times m}, d < m$, the linear operators to produce the feature vectors from the real part, the imagery part, and the phase, the problem (3) is converted to

$$\min_{\alpha} \|\alpha\|_1 \quad s.t \quad \|\tilde{\mathbf{y}}^j - \mathbf{F}^j\alpha^j\|_2 \leq \varepsilon, \ j = 1,2,3 \quad (8)$$

where $\mathbf{F}^1 = \mathcal{T}_r\mathbf{X}$, $\mathbf{F}^2 = \mathcal{T}_i\mathbf{X}$, and $\mathbf{F}^3 = \mathcal{T}_p\mathbf{X}$ are the redundant dictionary formed by the real part, the imagery part, and the phase;$\mathcal{T}\mathbf{X} = [\mathcal{T}\mathbf{x}_1, \mathcal{T}\mathbf{x}_2, \cdots, \mathcal{T}\mathbf{x}_n]$; $\tilde{\mathbf{y}}^1 = \mathcal{T}_r\mathbf{y}$, $\tilde{\mathbf{y}}^2 = \mathcal{T}_i\mathbf{y}$, $\tilde{\mathbf{y}}^3 = \mathcal{T}_p\mathbf{y}$.

Rather than produce the sparse pattern individually with (8), this paper proposes a joint sparse representation method. Thus it is capable to handle multiple monogenic components simultaneously. With the multiscale monogenic representation, it is easy to generate three feature vectors, $\mathcal{T}_r\mathbf{y}$, $\mathcal{T}_i\mathbf{y}$ and $\mathcal{T}_p\mathbf{y}$. Instead of classifying the components individually, this paper generate three similar sparsity pattern with joint sparse representation strategy,

$$\min_{\mathbf{A}} \|\mathbf{A}\|_{l_1\backslash l_2} \quad s.t. \quad \sum_{j=1}^{3} \|\tilde{\mathbf{y}}^j - \mathbf{F}^j\alpha^j\|_2^2 \leq \varepsilon \quad (9)$$

where $\{\alpha^i\}_{j=1}^3$ is the representations recovered by three monogenic components; $\mathbf{A} = [\alpha_1, \alpha_2, \alpha_3]$ is the coefficient matrix; $\|\mathbf{A}\|_{l1\backslash l_2}$ is the mixed-norm obtained by applying l_2-norm on each row and followed by l_1-norm on the resulting vector. Since multiple monogenic components are correlated to some extent, they tend to share the same sparsity pattern. On the other hand, due to the peculiarity of each component, it is impossible to be exactly same. It is reasonable to be represented using the similar pattern but different weights.

To delineate JSRC, a simple example has been given in Fig.1. There are three classes, BMP2, BTR70 and T72. The test (BMP2) is classified. The representation obtained by individual SRC (8) are shown in the first row, and the coefficients recovered by JSRC (9) are shown in the second row, with the corresponding residuals drawn in the last row. It can be seen that the sparsity model generated by JSRC are similar from component to component, while the ones recovered by individual SRC are much different. $\mathrm{SRC}^{\mathbf{R}}$ and $\mathrm{SRC}^{\mathbf{I}}$ predict the error class label, BTR70 and T72, while $\mathrm{SRC}^{\mathbf{P}}$ and JSRC give the correct identity.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

To verify the proposed method, several experiments are conducted on the publicly released database of MSTAR program, a collection done using a 10 GHz SAR sensor in one-foot resolution spotlight mode. Three similar targets, BMP2, BTR70, and T72 are employed, among which BMP2 and T72 consist of several variants with different series number and structural modifications. The prototype SN(9563) (for BMP2) and SN(132) (for T72) taken at 17° depression are used for training, and the others, SN(9566), SN(c21), SN(812), and SN(s7) captured at 15° depression are used for testing, as listed in Table 1. The raw chips are of around 128×128-pixel in size, and cropped to 80×80-pixel to exclude the clutter background. During the generation of monogenic feature vector, the down-sampling factor is chosen from the interval, $\{1/400, 1/256, 1/100, 1/64\}$, corresponding to the feature space of $16D, 25D, 64D, 100D$.

Table 1. The number of aspect views available for 3 objects

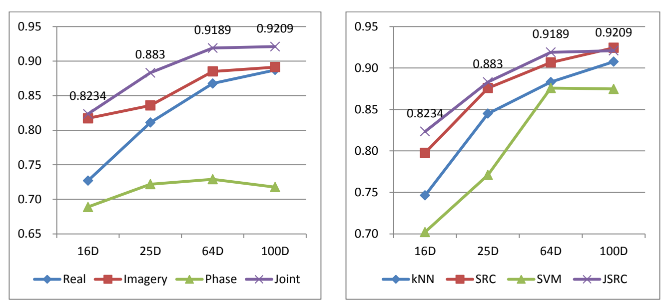
Depr.	BMP2	T72	BTR70	SUM
17°	**233**(9563)	**232**(132)	**233**	**698**
15°	196(9566) 196(c21)	195(812) 191(s7)	196	974

Table 2 lists the recognition accuracies with the feature space of $16D$, $25D$, $64D$, and $100D$. The corresponding performance is pictorially displayed in Fig.2, where the comparison of joint sparse representation model with the individual sparse representation model is shown in Fig.2 (a), and the comparison with the reference methods is illustrated in Fig.2 (b). The table and the picture are self-explanatory. With the feature space increased from $16D$ to $100D$, the significant improvement in accuracy has been achieved because much more information has been extracted for classification. From Fig.2 (a), it can be seen that the performance of JSRC with 4 feature space is much better than three individual SRC, *i.e.*, $\mathrm{SRC}^{\mathbf{R}}$, $\mathrm{SRC}^{\mathbf{I}}$, and $\mathrm{SRC}^{\mathbf{P}}$. The results demonstrate that joint sparse representation model could combine the advantage of individual monogenic component effectively.

Table 2. The recognition rates with the feature space of $16D$, $25D$, $64D$, and $100D$.

Dim.	$\mathrm{SRC}^{\mathbf{R}}$	$\mathrm{SRC}^{\mathbf{I}}$	$\mathrm{SRC}^{\mathbf{P}}$	JSRC	SRC	SVM	kNN
$16D$	0.7269	0.8172	0.6889	0.8234	0.7977	0.7022	0.7464
$25D$	0.8111	0.8358	0.7218	0.8830	0.8758	0.7710	0.8450
$64D$	0.8676	0.8850	0.7290	0.9189	0.9066	0.8758	0.8830
$100D$	0.8871	0.8912	0.7177	0.9209	0.9243	0.8747	0.9076

From Fig.2 (b), it can be seen that the accuracies for JSRC with the feature space of $16D$, $25D$, $64D$, and $100D$ are 0.8234, 0.8830. 0.9189, and 0.9209, 1.05%, 8.06%, 4.11% better in average than SRC, SVM, and kNN. The better performance obtained using JSRC mainly results from twofold. First, rich information has been extracted with the monogenic signal representation, and it could characterize SAR image effectively. Secondly, the correlations among multiple monogenic components have been exploited with the joint sparse representation model. Thus it is not surprising that JSRC improves the recognition accuracy with different feature space.



(a) (b)

Fig. 2. The comparison of recognition accuracy.

4. CONCLUSION

The classification via joint sparse representation of monogenic signal is presented for SAR image-based target recognition in this paper. The monogenic signal of multiresolution are used to characterize SAR image, and the joint sparse representation strategy is utilized to reach the inference. Several experimental comparison of the proposed method and the individual SRC, as well as other popular methods are conducted on MSTAR database. From the results, it can be come the conclusion the advantage of monogenic components are combined effectively with joint sparse representation scheme.

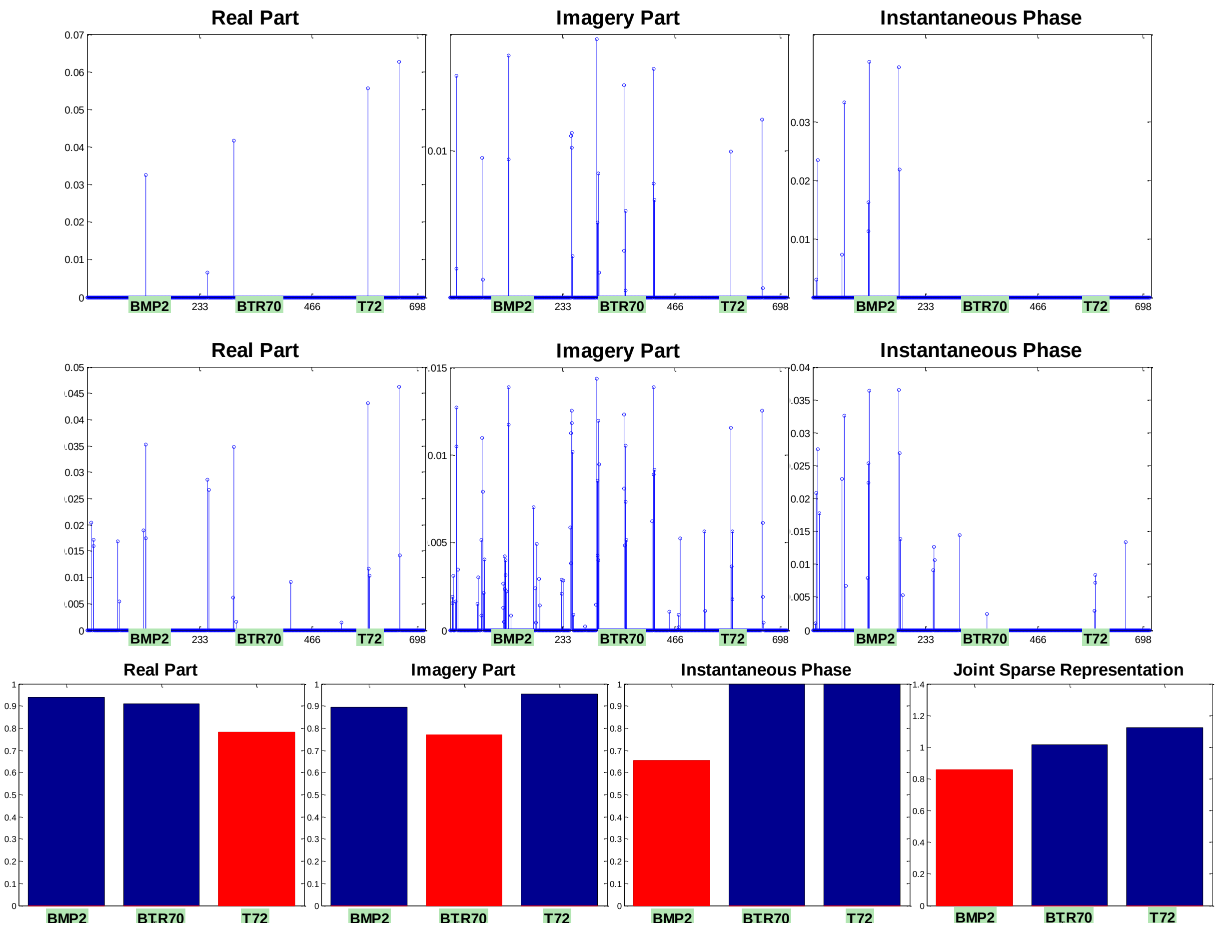


Fig. 1. Comparison of JSRMC with individual classifiers.

5. REFERENCES

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

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

[3] Y. Chen, N. Nasrabadi, and D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, 2011.

[4] J. Thiagarajan, K. Ramamurthy, P. Knee, and *et al.*, "Sparse representation for automatic target classification in SAR images," in *Int'l Sym. Communcitaion, Control and Signal Process.*, Mar. 2010, pp. 1–4.

[5] V. Patel, N. Nasrabadi, and R. Chellappa, "Sparsity-motivated automatic target recognition.," *Applied Optics*, vol. 50, no. 10, pp. 1425–1433, 2011.

[6] H. Zhang, N. Nasrabadi, Y. Zhang, and *et al.*, "Multi-view automatic target recognition using joint sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2495, 2011.

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

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

[9] M. Yang, L. Zhang, S. Shiu, and *et al.*, "Monogenic binary coding: An efficient local feature extraction approach to face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1738–1751, Dec. 2012.