

# Classification on the Monogenic Scale-Space: Application to Target Recognition in SAR Image

Ganggang Dong and Gangyao Kuang, *Member, IEEE*

**Abstract**—This paper introduces a novel classification strategy based on the monogenic scale-space for target recognition in SAR image. The proposed method exploits monogenic signal theory, a multi-dimensional generalization of the analytic signal, to capture the characteristics of SAR image, *e.g.*, broad spectral information and simultaneous spatial localization. The components derived from the monogenic signal at different scales are then applied into a recently developed framework, sparse representation-based classification (SRC). Moreover, to deal with the dataset whose target classes are not linearly separable, the classification via kernel combination is proposed, where the multiple components of the monogenic signal are jointly considered into a unifying framework for target recognition. The novelty of this paper comes from (i) the development of monogenic feature via uniformly downsampling, normalization, and concatenation of the components at various scales; (ii) the development of score-level fusion for sparse representation-based classifications; (iii) the development of composite kernel learning for classification. In particular, the comparative experimental studies under nonliteral operating conditions, *e.g.*, structural modifications, random noise corruption, variations in depression angle, are performed. The comparative experimental studies of various algorithms, including linear support vector machine and the kernel version (KSVM), SRC and the variants, kernel SRC, kernel linear representation (KLR), sparse representation of monogenic signal (MSRC), are performed too. The feasibility of the proposed method has been successfully verified using MSTAR database. The experimental results demonstrate that significant improvement for recognition accuracy can be achieved by the proposed method in comparison to the baseline algorithms.

**Keywords**—The monogenic signal, sparse representation, SAR target recognition, composite kernel learning, score-level fusion, monogenic scale-space.

## I. INTRODUCTION

**D**ue to the capability to produce all-weather, 24-hour a day, high-resolution images with quality quickly approaching that of optical imaging sensors, SAR has been used in many fields. Automatic target recognition (ATR) is an essential topic for SAR image interpretation. In the past several decades, SAR ATR has been studied pervasively, yet it is still a challenging problem, especially under the extended operating conditions (EOC) [1], [2], in which a single operational parameter is significantly different between the images used for training and those for testing. The baseline SAR ATR system comprises

three separate stages: detection, discrimination, and classification [3]. The first stage roughly locates the candidate targets in a SAR image by examining the amplitude of the radar signal in each site of the image. The second stage rejects the natural-clutter false alarms with some discriminants (*e.g.*, standard deviation, fractal dimension, and weighted-rank fill ratio [4]), followed by a classifier to reject the cultural-clutter false alarms and predict the target type of the remaining detections [5]. This paper mainly focuses on classification.

SAR ATR has been once relying on the template-matching strategy [6]. It refers to defining a distance metric (*e.g.*, mean square error) to quantify the similarity between the query and templates generated by various aspect view images of the objects. The decision is made by finding which template could match the query most closely. The identity is assigned as the target type to which the matched template belong. Template matching method usually suffers the presence of coherent fading (speckle) as well as misalignment. To cover the drawback, a family of methods named correlation pattern recognition have been presented [7]. Different from the conventional strategies, these methods perform the matching procedure in the frequency domain with a certain metric (*e.g.*, peak-to-sidelobe-ratio [8]), rather than in the spatial domain. Despite of achieving shift and distortion-invariance, it is needed to estimate the pose, from which a specific classifier can be selected. However, pose estimation from SAR image is very difficult, due to the mutability of the scattering phenomenology. In addition, the computational cost is unattractive because of the repetitious convolution with the templates. Another widely used family of approach to SAR ATR is the statistical model-based method [9], [10]. It relies on representing the intensity of SAR image with a parameterized statistical distribution model (*e.g.*, conditionally Gaussian model [11]) and reaching the inference by evaluating which class of parameters could maximize a posterior probability. However, it easily fails when strong statistical relationship does not exist between the training and the query, due to the problem of parameter estimation under limited statistical samples. Furthermore, the feature-based methods have also been popularly studied [5]. Since the target chip image usually composes of three different kinds of scattering phenomena (target, shadow, and background clutter), it is typically nonstationary. Thus, it is needed to separate the target and shadow beforehand, and hence enslaved to the accuracy of image segmentation. As widely reported, SAR image segmentation is still an open problem now.

Recently, with the development of compressed sensing theory [12], a great resurgence of sparse signal representation over a redundant dictionary has been witnessed [13]–[17]. In [18], the recognition problem is cast as one of classifying

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

Dr. G. Dong and Prof G. Kuang are with School of Electronics Science and Engineering, National University of Defense Technology, Changsha, China (e-mail: dongganggang@nudt.edu.cn; kuangyeats@hotmail.com).

among multiple linear regression models, and addressed by the sparse signal representation. By sparsity constraint ( $\ell_1$ -norm minimization), the unique representation can be generated. The decision is made by evaluating which class of samples could recover the query as accurately as possible. Since SAR images of a given class lie in a manifold, whose dimension is much lower than the actual one [19], the training images can be assumed to be the samples drawn from the manifold. Then, the classification of SAR image is equivalent to finding the manifold that is closest to the query image. According to the point of local linear embedding algorithm [20], linear representation can be provided to any nonlinear manifold if only a small local region is considered. Thus, it is applicable for sparse signal representation to SAR image-based target recognition, as verified in [21]–[25].

Recently, a novel 2-D analytic signal named the monogenic signal is presented in [26], [27]. It has been built around the Riesz transform, a 2-D vector-valued extension of Hilbert transform [28]. As a generalization of the analytic signal, the monogenic signal could preserve the properties of analytic signal concerning symmetry, energy, allpass transfer function, as well as the invariance-equivariance property with respect to phase and orientation. Thus, it has been widely studied in the preceding works. In [29], [30], a 2-D monogenic wavelet transform is presented for image analysis. The main advantage of monogenic wavelet transform consists in the capability to offer a geometric representation for gray-scale image through an AM-FM model allowing invariance of coefficients to translations and rotations. In [31], a binary encoding scheme named monogenic binary coding is presented for biometric recognition. The components of the monogenic signal are encoded by various binary pattern schemes, followed by a nearest neighbor classifier to make the decision. In [32], the monogenic signal is used for optical flow estimation, where the displacement is locally computed by assuming the conservation of the monogenic phase over time. Moreover, the monogenic signal is employed for SAR target recognition in [23], [24]. To make the algorithm computationally tractable, a monogenic feature descriptor is formed via concatenating the monogenic components. Since various kinds of information are roughly combined, it may result in some information loss. To cover the shortage, this paper introduces the score-level fusion. Real part,  $i$ -imaginary part, and  $j$ -imaginary part of the monogenic signal at various scales are used to generate three different feature descriptors, from which three sparse representation-based classifications can be built. Specifically, the features of training samples are used to build a redundant dictionary to represent the counterpart of the query as a linear combination of them. The unique representations can be generated via sparsity constraint ( $\ell_1$ -norm minimization), as similar to [18], [23], [24]. Then, the results of individual classification are combined by performing score-level fusion [33]. Though the performance can be improved by score-level fusion strategy, it may be not effective if the target classes are not linearly separable. To address the problem, this paper embeds the monogenic features into a high-dimensional Hilbert space induced by a kernel function. Given the nonlinear mappings, the proposed monogenic features can be transformed into the

Hilbert spaces, where the query can be linearly represented in terms of the training samples. To jointly consider these features in a unifying system for target recognition, the component-specific Hilbert spaces are combined to form a larger global one by composite kernel learning [34]. Then, the similarity between a pair of samples can be evaluated more accurately. To produce the unique representation, this paper employs linear coding ( $\ell_2$ -norm minimization) strategy [35], [36], rather than sparse coding one ( $\ell_1$ -norm minimization) [37], [38] to limit the feasible set of the representations. This is because the data separability between different classes has been enhanced in the global Hilbert space, and a linear coding strategy could yield the performance similar to sparse coding method yet with less computational cost. Finally, the inference is reached according to the characteristics of the representation on reconstruction.

The reminder of this paper is organized as follows. Section II briefly reviews the formulation of sparse coding and linear coding in the Hilbert space, as well as the kernel tricks. Section III devotes to the proposed classification via score-level fusion and the one via composite kernel learning. The effectiveness of the proposed methods has been evaluated in Section IV by several experiments on MSTAR database. Finally, Section V concludes this paper.

## II. RELATED WORKS

This section briefly reviews two preceding works, sparse representation-based classification and kernel linear representation-based classification. The details can be found in Ref [18], [39] and Ref [35], [36]. Hereafter, a image with  $w \times h$  pixels in size will be identified as a column vector  $\mathbb{R}^m$  ( $m = wh$ ) by rearranging its columns.

### A. Sparse representation-based classification (SRC)

Sparse and redundant representation technique has recently gained popularity in many fields, *e.g.*, pattern recognition [39]. It builds on the assumption that samples from a single class approximately lie on a linear subspace, whose dimension is much lower than the actual one. Given sufficient samples for the  $k$ -th target class,  $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,n_k}] \in \mathbb{R}^{m \times n_k}$ , for any new test sample  $\mathbf{y} \in \mathbb{R}^m$  from the same class, it can be well represented by samples of the  $k$ -th class,

$$\mathbf{y} = \mathbf{x}_{k,1}\alpha_{k,1} + \dots + \mathbf{x}_{k,n_k}\alpha_{k,n_k} \quad (1)$$

for scalars  $\alpha_k = [\alpha_{k,1}, \dots, \alpha_{k,n_k}]^T \in \mathbb{R}^{n_k}$ .

Since the class membership of the new sample is unknown initially, we usually represent it in terms of training samples from all  $K$  target classes,  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \in \mathbb{R}^{m \times n}$ , where  $n = \sum_{k=1}^K n_k$  is the total number of training samples. Then, the linear representation of  $\mathbf{y}$  can be rewritten as

$$\mathbf{y} = \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \dots + \mathbf{x}_n\alpha_n = \mathbf{X}\alpha, \quad (2)$$

where  $\alpha = [\alpha_1, \dots, \alpha_n]^T \in \mathbb{R}^n$  is the representation.

Considering the under-determined system ( $m \leq n$ ), a great many solutions to (2) can be provided, among which the most compact representation is popularly contended [18]

$$\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{X}\alpha\|_2 < \epsilon, \quad (3)$$

where  $\epsilon$  is the allowed error tolerance. In (3), the objective function  $\|\cdot\|_0$  measures the sparsity level, while the constraint term reflects the reconstruction error. Since the objective function is nonconvex, a direct solution requires a combinational search of exponential size for taking all possible supports into account, and hence is prohibitive. Then, various alternative algorithms for approximating (3) have been presented, among which a convex relaxation that replaces  $\|\alpha\|_0$  by  $\|\alpha\|_1$  is widely utilized [12],

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{X}\alpha\|_2 < \epsilon. \quad (4)$$

If considering  $\alpha$  as a set of random variates, the formulation of (4) can be interpreted as attributing a Laplace-prior  $P(\mathbf{x}) \propto \exp\{-\lambda\|\mathbf{x}\|_1\}$  on the representation coefficients [40].

With the optimal representation  $\hat{\alpha}$ , the identity of the query can be retrieved by evaluating which class of samples could result in the minimum reconstruction error

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

### B. Kernel linear representation-based classification (KLR)

If the classes are not linearly separable, the prior assumption that samples from a single class approximately lie on a linear subspace is not founded. Then, an intuitive idea is to embed the samples into the Hilbert space ( $\mathcal{F}$ ) with a nonlinear mapping.

Let  $\mathbf{x}$  be the sample and  $\phi(\mathbf{x})$  be its imagery in  $\mathcal{F}$ . The kernel function,  $\kappa(\cdot, \cdot) : \mathcal{F} \times \mathcal{F} \mapsto \mathbb{R}$  is defined as the inner produce in the Hilbert space  $\mathcal{F}$

$$\kappa(\mathbf{x}_p, \mathbf{x}_q) = \langle \phi(\mathbf{x}_p), \phi(\mathbf{x}_q) \rangle_{\mathcal{F}}. \quad (6)$$

A valid kernel satisfies the Mercer's condition, *i.e.*, the kernel Gram matrix formed by restricting  $\kappa$  to any finite subset of  $\mathcal{F}$  is positive semi-definite [41]. A popularly used kernel is the Gaussian radial basis function (RBF)  $\kappa(\mathbf{x}_p, \mathbf{x}_q) = \exp(-\gamma\|\mathbf{x}_p - \mathbf{x}_q\|_2^2)$  with  $\gamma$  controlling the width.

Since the classes are linearly separable in  $\mathcal{F}$ , the test sample can be linearly represented by the training samples

$$\phi(\mathbf{y}) = \phi(\mathbf{x}_{1,1})\alpha_{1,1} + \phi(\mathbf{x}_{1,2})\alpha_{1,2} + \dots + \phi(\mathbf{x}_{K,n_K})\alpha_n = \phi(\mathbf{X})\alpha,$$

where  $\phi(\mathbf{X}) = [\phi(\mathbf{X}_1), \dots, \phi(\mathbf{X}_K)]$  is the dictionary composes of the imageries of the training samples in the Hilbert space. To produce the unique representation, linear coding strategy

$$\min_{\alpha} \|\alpha\|_2 \text{ subject to } \|\phi(\mathbf{y}) - \phi(\mathbf{X})\alpha\|_2 < \epsilon \quad (7)$$

is employed [36]. Following the thought of sparse coding, the formulation of (7) can be also understood as attributing a Gibbs distribution prior  $P(\mathbf{x}) \propto \exp\{-\lambda\|\mathbf{x}\|_2\}$  on the representation coefficients. For computational convenience, (7) can be converted to  $\min_{\alpha} \{g(\alpha) = \|\phi(\mathbf{y}) - \phi(\mathbf{X})\alpha\|_2^2 + \lambda\|\alpha\|_2^2\}$ , where  $\lambda$  is a tradoff parameter to balance the fidelity and sparsity [42]. By conducting the derivative and set the result to be zero,  $\frac{\partial g(\alpha)}{\partial \alpha} = 2(\phi(\mathbf{X})^T \phi(\mathbf{X})\alpha - \phi(\mathbf{X})^T \phi(\mathbf{y}) + \lambda\alpha) = 0$ , it is easy to obtain the representation,

$$\hat{\alpha} = (\phi(\mathbf{X})^T \phi(\mathbf{X}) + \lambda \mathbf{I})^{-1} \phi(\mathbf{X})^T \phi(\mathbf{y}), \quad (8)$$

where  $\mathbf{I}$  is an identity matrix. By taking the kernel (6) into (8), the solver can be simplified as  $\hat{\alpha} = (\langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{F}} + \lambda \mathbf{I})^{-1} \langle \mathbf{X}, \mathbf{y} \rangle_{\mathcal{F}}$ , where  $\mathbf{K} = \langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{F}} \in \mathbb{R}^{n \times n}$  is the kernel Gram matrix;  $\mathcal{K}_{\mathbf{y}} = \langle \mathbf{X}, \mathbf{y} \rangle_{\mathcal{F}} \in \mathbb{R}^n$ . Since  $\mathbf{K}$  is positive semi-definite,  $(\mathbf{K} + \lambda \mathbf{I})$  is positive definite, and hence the matrix inverse in (8) makes sense. The solver (8) is in close form, thus it is computationally attractive compared to sparse coding (4).

With the optimal representation  $\hat{\alpha}$ , the inference can be reached by evaluating which class of samples could recover the query as accurately as possible,  $\min_k \|\phi(\mathbf{y}) - \phi(\mathbf{X}_k) \hat{\alpha}_k\|_2$ .

### III. CLASSIFICATION ON THE MONOGENIC SCALE-SPACE

The monogenic signal is a vector-valued generalization of the analytic signal. It could capture the broad spectral information and simultaneous spatial localization with compact support [31]. Thus, it is used to characterize SAR image in this paper. Points in  $\mathbb{R}^2$  is notated as  $\mathbf{z} = (x, y)^T$  in the spatial domain, and  $\mathbf{u} = (u, v)^T$  in the frequency domain.

#### A. The Monogenic Scale-Space

The analytic signal is defined as a combination of the signal and its Hilbert transformed one. Denote by  $\mathcal{H}\{\cdot\}$  the Hilbert transform. For a 1-D real valued signal  $f(x)$ , the analytic signal is defined as

$$f_A(x) = f(x) + j\mathcal{H}\{f\}(x) = A(x) \exp\{j\varphi(x)\}, \quad (9)$$

where  $A(x)$  and  $\varphi(x)$  denote local amplitude and local phase. In the polar-coordinate representation, local amplitude refers to a local quantitative measure of a signal, while local phase provides a local measure for the qualitative information of a signal.

The monogenic signal is a multi-dimensional generalization of analytic signal [26]. It has been built around the Riesz transform, a scalar-to-vector extension of Hilbert transform [28]. For signal  $f(\mathbf{z})$  and its frequency spectra  $F(\mathbf{u})$ , the Riesz transform can be expressed as

$$f_R(\mathbf{z}) = h_R(\mathbf{z}) * f(\mathbf{z}) = [h_R^x(\mathbf{z}) * f(\mathbf{z}), h_R^y(\mathbf{z}) * f(\mathbf{z})]^T, \quad (10)$$

where  $h_R(\mathbf{z}) = -\frac{\mathbf{z}}{2\pi|\mathbf{z}|^3}$  is the Riesz kernel;  $h_R^x(\mathbf{z}), h_R^y(\mathbf{z})$  are the first and second-order Riesz kernel. Correspondingly, the Riesz transform in the frequency domain can be shown as

$$F_R(\mathbf{u}) = \frac{i\mathbf{u}}{|\mathbf{u}|} F(\mathbf{u}) = \left[ \frac{i u}{|\mathbf{u}|} F(\mathbf{u}), \frac{i v}{|\mathbf{u}|} F(\mathbf{u}) \right]^T. \quad (11)$$

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

$$f_M(\mathbf{z}) = f(\mathbf{z}) - (i, j) f_R(\mathbf{z}) \quad (12)$$

where  $i, j$  are the imagery units and  $\{i, j, 1\}$  is an orthonormal basis of  $\mathbb{R}^3$ . Then, it is capable to decompose the original signal into three components, real part,  $i$ -imaginary part, and  $j$ -imaginary part.

As an extension of analytic signal, the monogenic signal could preserve the properties of the analytic signal concerning

symmetry, doubling energy of the original signal, allpass transfer function, as well as the invariance-equivariance property of signal decomposition [26].

Since the practical signal is of finite length, it has broad Fourier spectra [43], [44], and hence needs infinite extension with a band-pass filter before applying the Riesz transform. Then, the monogenic representation of a given signal is the process of computing multiple components of a bandpass version of the signal, whose expression will be expressed as

$$f_M(\mathbf{z}) = (h_{bp} * f)(\mathbf{z}) - (i, j)h_{bp} * f_R(\mathbf{z}) \quad (13)$$

where  $h_{bp}$  denotes the kernel of bandpass filter.

The property of Riesz transform requires that  $h_{bp}$  should be symmetric to maintain the odd character of Riesz kernels and with null DC component [45]. Considering the mathematical, computational, and empirical reasons, this paper employs log-Gabor filter, whose expression in the frequency domain is

$$G(\omega) = \exp \left\{ - \frac{[\log(\omega/\omega_0)]^2}{[\log(\sigma/\omega_0)]^2} \right\}, \quad (14)$$

where  $\omega_0$  is the center frequency, and  $\sigma$  is the scaling factor of the bandwidth. Usually, the ratio  $\sigma_r = \frac{\sigma}{\omega_0}$  is kept as a constant. The log-Gabor filter could capture the broad spectral information of the signal with compact spatial filter, as recommended in the preceding works [46]–[48].

To fully exploit the characteristics of a signal, the multi-scale representation strategy is required. The popularly used approach to realize multiscale log-Gabor filters is to tune the scaling factor  $\sigma$  and center frequency  $\omega_0$ ,

$$\sigma = \sigma_r \omega_0, \quad \omega_0 = (\lambda_{\min} \mu^{s-1})^{-1}, \quad (15)$$

where  $\lambda_{\min}$  is the minimal wavelength;  $\mu$  is the multiplicative factor of wavelength;  $s$  is the scale index. Then, the monogenic signal is embedded in the log-Gabor scale-space by convolving its components with the bandpass filter  $h_{bp}$ , and hence the monogenic scale-space can be formed [27].

Suppose  $S$ -scale log-Gabor filters are built, the monogenic signal at different scales can be shown as  $f_M^1, f_M^2, \dots, f_M^S$ , with the corresponding components expressed as,

$$\left[ \underbrace{\mathcal{R}_1, \mathcal{I}_1, \mathcal{J}_1}_{f_M^1}, \dots, \underbrace{\mathcal{R}_k, \mathcal{I}_k, \mathcal{J}_k}_{f_M^k}, \dots, \underbrace{\mathcal{R}_S, \mathcal{I}_S, \mathcal{J}_S}_{f_M^S} \right], \quad (16)$$

where  $\mathcal{R}_k, \mathcal{I}_k, \mathcal{J}_k$  denote real part,  $i$ -imaginary part, and  $j$ -imaginary part of the monogenic signal at the  $k$ th scale.

To illustrate the multiscale monogenic signal representation, an example for SAR image is given in Fig.1. Obviously, it is able to generate a set of functions that gives a good coverage of the frequency domain with the log-Gabor filters.

### B. Classification via Score-level Fusion

As pointed in Section II-A, sparse signal representation builds on the assumption that samples from a single class approximately lie on a linear subspace. In addition, sufficient training samples should be available for each class. Unfortunately, these prior assumptions are easily violated for SAR target recognition. SAR image reflects the fine target structure

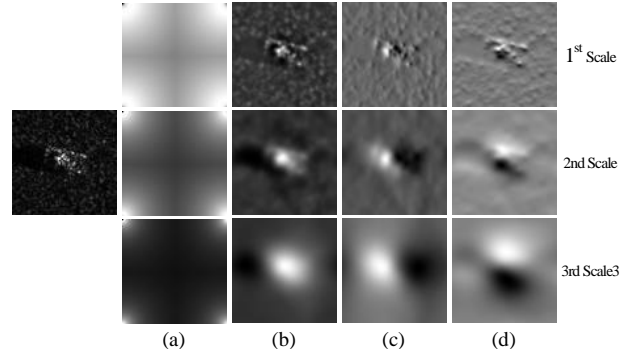


Fig. 1. Illustration of 3-scale monogenic representation for SAR image; (a) log-Gabor filters; (b) real part; (c)  $i$ -imaginary part; (d)  $j$ -imaginary part.

(point scattering distribution on the target surface) at a certain pose. Parts of the target structure will be occluded when illuminated by the radar from another pose, and hence results in dramatic differences of characteristics from image to image taken with angular increments of only a few degrees. In other words, SAR images of the same target taken at different aspect views show great difference of scattering phenomenology. Furthermore, the training resource available for radar target recognition is scarce. To cover these shortages, this paper adopts the monogenic signal, which is capable to capture broad spectral information and simultaneous spatial localization, into the framework of sparse signal representation.

Since the components of the monogenic signal are of high-dimension, it is unrealistic to be directly utilized. To address the problem, the preceding works [23], [24] derive a single feature descriptor from the components of the monogenic signal at different red scales. This method can be viewed as the data fusion in feature-level. Since various kinds of information are roughly concatenated, it may result in some information loss. To improve the performance further, this paper introduces the score-level fusion to combine various kinds of information contained in the monogenic signal for target recognition. The details on the proposed method are described as follows.

First, three monogenic features are derived from real part,  $i$ -imaginary part, and  $j$ -imaginary part of the monogenic signal at different scales. Specifically, the components of the monogenic signal at various scales are first uniformly down-sampled by a factor of  $\rho$ , followed by normalization and concatenation to form the feature vectors, as expressed by

$$\begin{aligned} \chi_{\mathcal{R}} &= [\mathcal{R}_1^{(\rho)}(:); \mathcal{R}_2^{(\rho)}(:); \dots; \mathcal{R}_S^{(\rho)}(:)] \\ \chi_{\mathcal{I}} &= [\mathcal{I}_1^{(\rho)}(:); \mathcal{I}_2^{(\rho)}(:); \dots; \mathcal{I}_S^{(\rho)}(:)], \\ \chi_{\mathcal{J}} &= [\mathcal{J}_1^{(\rho)}(:); \mathcal{J}_2^{(\rho)}(:); \dots; \mathcal{J}_S^{(\rho)}(:)] \end{aligned} \quad (17)$$

where  $(:)$  reshapes a matrix to be a single vector. To make the subsequent optimization problem ( $\ell_1$ -norm minimization) computationally tractable, the data dimensionality of the monogenic features are further reduced using principal component analysis transform.

Have defined the feature descriptors, the remaining problem is how to reach the decision based on these features.

Different from the preceding works [23], [24], this paper constructs three component-specific classifications via sparse signal representation. Specifically, the monogenic features of training samples are used to build a redundant dictionary, for example,  $\mathcal{D}_{\mathcal{R}} = [\chi_{\mathcal{R}}(\mathbf{x}_{1,1}), \chi_{\mathcal{R}}(\mathbf{x}_{1,2}), \dots, \chi_{\mathcal{R}}(\mathbf{x}_{K,n_K})]$ . Then, the counterpart of the new query sample  $\chi_{\mathcal{R}}(\mathbf{y})$  can be represented as a linear combination of atoms in  $\mathcal{D}_{\mathcal{R}}$ ,

$$\chi_{\mathcal{R}}(\mathbf{y}) = \chi_{\mathcal{R}}(\mathbf{x}_{1,1})\alpha_{1,1} + \chi_{\mathcal{R}}(\mathbf{x}_{1,2})\alpha_{1,2} + \dots + \chi_{\mathcal{R}}(\mathbf{x}_{K,n_K})\alpha_{K,n_K}$$

Similar to (4), the most compact representation can be obtained by,

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|\chi_{\mathcal{R}}(\mathbf{y}) - \mathcal{D}_{\mathcal{R}}\alpha\|_2 < \epsilon, \quad (18)$$

from which the residual  $\mathbf{e}^{\mathcal{R}} = \{\|\chi_{\mathcal{R}}(\mathbf{y}) - \chi_{\mathcal{R}}(\mathbf{x}_k)\hat{\alpha}_k\|\}_{k=1}^K$  can be computed. Correspondingly, it is able to build the other two classifications by features  $\chi_{\mathcal{I}}$  and  $\chi_{\mathcal{J}}$ , from which the residuals  $\mathbf{e}^{\mathcal{I}}$  and  $\mathbf{e}^{\mathcal{J}}$  can be obtained accordingly.

In the framework of sparse signal modeling, the entry of residual reflects the distance from the query to the manifold formed by a certain class, and hence is used to decide the class membership of the query. The smaller the minimal residual is, the more reliable the decision is, and the more believable the classifier is. Therefore, it is reasonable to quantify the classification power of each classifier according to its residual. To make fair comparison, the residuals derived from each classifiers are normalized into a certain range, e.g., (0,1). Though various approaches to normalization have been presented in the preceding works [33], [49], the detail comparison on which one outperforms the others beyonds the scope of this paper. Here, the summation normalization,  $\aleph(\mathbf{e}) = \mathbf{e} / \sum_{k=1}^K \mathbf{e}_k$ , is used to make the residual energy to be unit.

To combine the benefit of the individual classifier and tackle their drawback, two simple yet effective strategies, summation rule and Bayes decision, are proposed as follows. The flowchart of the proposed classification via score-level fusion is pictorially demonstrated in Fig.2.

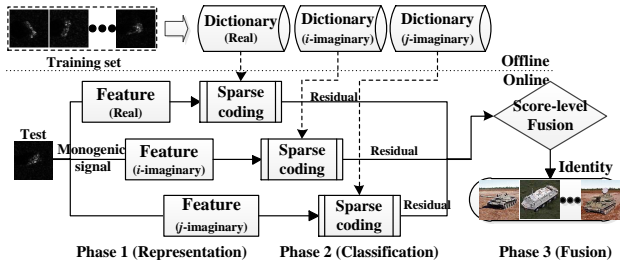


Fig. 2. The block diagram of classification via score-level fusion. Three classifications are formed by real part,  $i$ -imaginary part, and  $j$ -imaginary part of the monogenic signal. The results are combined by performing score-level fusion (Bayesian inference).

**[Summation]** An intuitive idea to combine the residuals from the individual classifiers is to produce a global residual by summing their values,

$$\tilde{\mathbf{e}} = \aleph(\mathbf{e}^{\mathcal{R}}) + \aleph(\mathbf{e}^{\mathcal{I}}) + \aleph(\mathbf{e}^{\mathcal{J}}). \quad (19)$$

The decision is then produced according to the global residual,  $\min_k \tilde{\mathbf{e}}_k$ , as similar to (5).

**[MAP]** Another approach for score-level fusion derives from the Bayesian theory. Let  $\{TG_1, TG_2, \dots, TG_K\}$  be  $K$  target classes. Since the entry of residual provides the distance from the query to the manifold built by a certain class, this paper defines the conditional probability associated with hypothesis  $H_k : \mathbf{y} \in TG_k$  to be inversely proportional to the residual committed to the  $k$ -th class,

$$\mathbb{P}(\mathbf{y}|H_k) = \frac{1/e_k}{\sum_{l=1}^K 1/e_l}. \quad (20)$$

According to the Bayes theorem, a posterior probability associated with the hypothesis  $H_k$  can be computed by

$$\mathbb{P}(H_k|\mathbf{y}) = \frac{\mathbb{P}(\mathbf{y}|H_k)\mathbb{P}(H_k)}{\mathbb{P}(\mathbf{y})}, \quad (21)$$

where  $\mathbb{P}(H_k)$  is the prior probability. Since  $\mathbb{P}(\mathbf{y})$  is of constant, the maximum of (21) can be approximated to  $\max\{\mathbb{P}(H_k|\mathbf{y})\} \propto \max\{\mathbb{P}(\mathbf{y}|H_k)\mathbb{P}(H_k)\}$ . By assuming the prior probability  $\mathbb{P}(H_k)$  to be equiprobable,  $\mathbb{P}(H_k) = \frac{1}{K}$  for  $k = 1, \dots, K$ , the maximum of a posterior probability can be further simplified as  $\max\{\mathbb{P}(H_k|\mathbf{y})\} \propto \max\{\mathbb{P}(\mathbf{y}|H_k)\}$ . It should be noted that the prior probability  $\mathbb{P}(H_k)$  can be also task-specific if some prior knowledges are available.

Here, it is assumed that the decision derived from a monogenic features are conditionally statistically independent to the one derived from the others [33]. Then, the maximum of a joint posterior probability can be approximated as

$$\begin{aligned} & \max\{\mathbb{P}(H_k|\chi_{\mathcal{R}}(\mathbf{y}), \chi_{\mathcal{I}}(\mathbf{y}), \chi_{\mathcal{J}}(\mathbf{y}))\} \\ & \propto \max\{\mathbb{P}(\chi_{\mathcal{R}}(\mathbf{y})|H_k)\mathbb{P}(\chi_{\mathcal{I}}(\mathbf{y})|H_k)\mathbb{P}(\chi_{\mathcal{J}}(\mathbf{y})|H_k)\} \end{aligned} \quad (22)$$

According to the maximum a posterior probability criteria, the final inference can be reached by

$$\max_k (\mathbb{P}(\chi_{\mathcal{R}}(\mathbf{y})|H_k)\mathbb{P}(\chi_{\mathcal{I}}(\mathbf{y})|H_k)\mathbb{P}(\chi_{\mathcal{J}}(\mathbf{y})|H_k)). \quad (23)$$

To delineate the proposed method, an example is given in Fig. 3. There is a total of four target classes ( $TG_1, TG_3, TG_5$ , and  $TG_7$ ). The detail description on the dataset can be found in Section IV. The query sample is chosen from  $TG_1$  set. Though real and  $j$ -imaginary-specific classifications recognize the query as  $TG_7$  and  $TG_3$ , the proposed method correctly infers the class membership of the query. The classification via summation rule generates the minimum global residual to  $TG_1$  class, while the classification via MAP criteria produces the maximum posterior probability associated with the hypothesis that the query belongs to  $TG_1$ .

### C. Classification via Kernel Combination

Previously, three individual classifications are formed on the monogenic scale-space, and the results of which are combined by score-level fusion scheme. Though the performance can be improved, it may be not effective if the classes are not linear separable. To address this problem, a popular idea is to cast the samples into the Hilbert space whose dimension is high or even infinite. Then, the similarity between a pair of samples can be estimated in a sufficiently rich feature space ( $\mathcal{F}$ ).

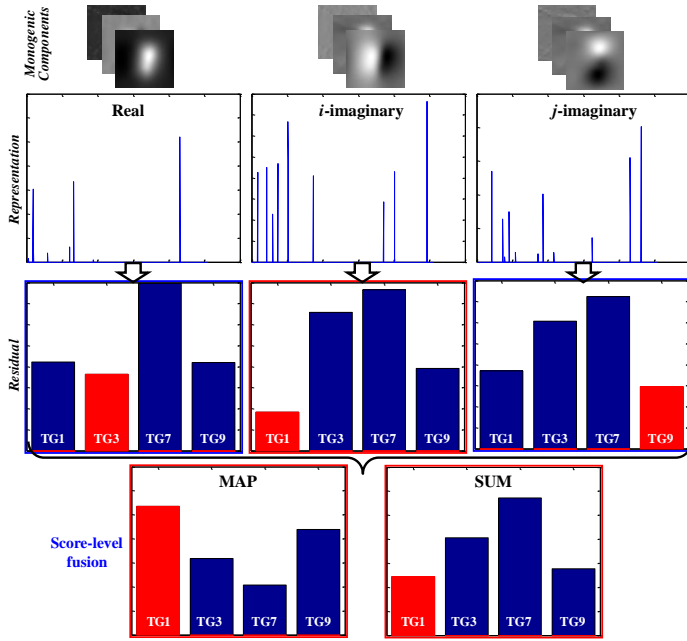


Fig. 3. Example of classification via score-level fusion. The first row shows the components of the monogenic signal at different scales. The second and third rows draw the representations and the corresponding residuals resulting from sparse signal modeling. The last row paints the resulting global residual (or posterior probability).

Given a nonlinear mapping  $\phi(\cdot) : \mathbb{R}^m \mapsto \mathcal{F}$ , by which the monogenic features (17) can be cast into the Hilbert space. Inspired by the preceding work [34], where the spatial and spectral information are combined by constructing a family of composite kernels, this paper proposes to combine the information contained in the monogenic signal by composite kernel learning. We first integrate three component-specific Hilbert spaces into a larger global space, in which the similarity between a pair of samples could be evaluated more accurately. The proposed method can be summarized as three sequential steps, as pictorially shown in Fig.4.

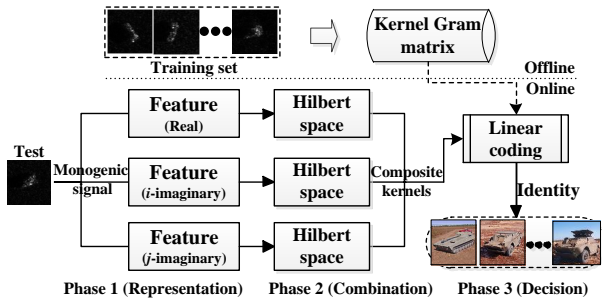


Fig. 4. The flowchart of classification via kernel combination. The monogenic features generated on the monogenic scale-space are embedded into the Hilbert spaces. The component-specific Hilbert spaces are combined to form a larger one, in which linear coding strategy is used to make the decision.

- Monogenic representation. For any new test sample  $y$ , it is represented by three monogenic feature descriptors,

$\chi_{\mathcal{R}}(y), \chi_{\mathcal{I}}(y), \chi_{\mathcal{J}}(y)$ , formed on log-Gabor filters induced monogenic scale-space.

- Kernel combination. The monogenic features are embedded into the Hilbert spaces with the nonlinear mappings. By stacked features or summation kernel, the component-specific Hilbert spaces are combined to form a larger global space.
- Make decision. In the global Hilbert space, the query is represented as a linear combination of training samples. The unique representation is generated by linear coding strategy, and the decision is made according to the characteristics of the representation on reconstruction.

Since the first step has been stated previously, the following section will pay more attention to kernel combination.

In Section III-B, three monogenic features are generated on the monogenic scale-space. To combine the information contained in the monogenic features, this paper jointly considers these features into a unifying framework for target recognition. It is realized by a recently developed learning skill, kernel combination. Inspired by the preceding works [34], two schemes are contended.

**Scheme 1: Stacked Features.** The first scheme refers to forming an augmented feature  $\chi_c \in \mathbb{R}^d$  by stacking the monogenic features  $\chi_{\mathcal{R}}, \chi_{\mathcal{I}}$  and  $\chi_{\mathcal{J}}$  to be an single vector, as similar to the preceding work [23]. Then, we cast the resulting augmented feature into the Hilbert space  $\mathcal{F}_c$ . For any two samples  $\{\mathbf{x}_p, \mathbf{x}_q\}$ , if denote by  $\phi_c(\cdot) : \mathbb{R}^d \mapsto \mathcal{F}_c$  the mapping that projects an augmented feature into the Hilbert space  $\mathcal{F}_c$ , the kernel can be formed as

$$\kappa(\mathbf{x}_p, \mathbf{x}_q) = \langle \phi_c(\chi_c\{\mathbf{x}_p\}), \phi_c(\chi_c\{\mathbf{x}_q\}) \rangle_{\mathcal{F}_c}, \quad (24)$$

where  $\langle \cdot, \cdot \rangle_{\mathcal{F}_c}$  is defined in the combined Hilbert space  $\mathcal{F}_c$ .

**Scheme 2: Summation Kernel.** Unlike the first scheme, the second one transforms three monogenic features into three distinct Hilbert spaces. Given three nonlinear mappings,  $\phi_{\mathcal{R}}(\cdot)$ ,  $\phi_{\mathcal{I}}(\cdot)$ , and  $\phi_{\mathcal{J}}(\cdot)$ , by which the monogenic features can be cast into three different Hilbert spaces  $\mathcal{F}_{\mathcal{R}}, \mathcal{F}_{\mathcal{I}}$ , and  $\mathcal{F}_{\mathcal{J}}$ . Then, the global mapping can be expressed as

$$\phi_s(\mathbf{x}_p) = \{\phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_p\}), \phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_p\}), \phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_p\})\}. \quad (25)$$

Correspondingly, the kernel in the global Hilbert space can be formed as (26). Here,  $\kappa_{\mathcal{R}}(\cdot, \cdot)$ ,  $\kappa_{\mathcal{I}}(\cdot, \cdot)$ , and  $\kappa_{\mathcal{J}}(\cdot, \cdot)$  are the kernel functions to induce three component-specific Hilbert spaces  $\mathcal{F}_{\mathcal{R}}, \mathcal{F}_{\mathcal{I}}$ , and  $\mathcal{F}_{\mathcal{J}}$ . This paper specifies them as the Gaussian radial basis function parameterized by the width  $\gamma_{\mathcal{R}}, \gamma_{\mathcal{I}}$ , and  $\gamma_{\mathcal{J}}$ . The inner products,  $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{R}}}$ ,  $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{I}}}$ , and  $\langle \cdot, \cdot \rangle_{\mathcal{F}_{\mathcal{J}}}$  are defined in the Hilbert spaces  $\mathcal{F}_{\mathcal{R}}, \mathcal{F}_{\mathcal{I}}$ , and  $\mathcal{F}_{\mathcal{J}}$ , respectively. It should be noted that this paper equally deals with three monogenic features  $\chi_{\mathcal{R}}, \chi_{\mathcal{I}}$ , and  $\chi_{\mathcal{J}}$ , *i.e.*, no special emphasis (importance) is considered for each feature. Give some prior knowledge, these kernel functions can be also weighted, *e.g.*,  $\phi_s(\mathbf{x}_p) = \{\beta_{\mathcal{R}}\phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_p\}), \beta_{\mathcal{I}}\phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_p\}), \beta_{\mathcal{J}}\phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_p\})\}$ , where  $\beta_{\mathcal{R}}, \beta_{\mathcal{I}}$ , and  $\beta_{\mathcal{J}}$  are the weights for three monogenic features.

Previously, two schemes are presented for kernel combination learning. The remaining problem is how to make the decision in the resulting Hilbert space. Since the classes are linear separable in the global Hilbert space, it is able



$$\begin{aligned}
\kappa(\mathbf{x}_p, \mathbf{x}_q) &= \langle \phi_s(\mathbf{x}_p), \phi_s(\mathbf{x}_q) \rangle \\
&= \langle \{\phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_p\}), \phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_p\}), \phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_p\})\}, \{\phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_q\}), \phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_q\}), \phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_q\})\} \rangle \\
&= \langle \phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_p\}), \phi_{\mathcal{R}}(\chi_{\mathcal{R}}\{\mathbf{x}_q\}) \rangle_{\mathcal{F}_{\mathcal{R}}} + \langle \phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_p\}), \phi_{\mathcal{I}}(\chi_{\mathcal{I}}\{\mathbf{x}_q\}) \rangle_{\mathcal{F}_{\mathcal{I}}} + \langle \phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_p\}), \phi_{\mathcal{J}}(\chi_{\mathcal{J}}\{\mathbf{x}_q\}) \rangle_{\mathcal{F}_{\mathcal{J}}}
\end{aligned} \tag{26}$$

to represent the query as a linear combination of training samples [35], [36]. By limiting the feasible set with  $\ell_2$ -norm minimization, it is easy to generate the unique representation  $\hat{\alpha} = (\langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{F}} + \lambda \mathbf{I})^{-1} \langle \mathbf{X}, \mathbf{y} \rangle_{\mathcal{F}}$  by least square fitting method. The details can be found in Section II-B or [36]. Here,

$$\langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{F}} = \{\kappa(\mathbf{x}_p, \mathbf{x}_q)\}_{p,q=1}^n \tag{27}$$

is the kernel Gram matrix, and

$$\langle \mathbf{X}, \mathbf{y} \rangle_{\mathcal{F}} = \{\kappa(\mathbf{x}_p, \mathbf{y})\}_{p=1}^n \tag{28}$$

is a  $n$ -dimensional vector whose entries are the similarity measurements between the query and each training sample. By incorporating (24) or (26) into (27) and (28), the optimal representation  $\hat{\alpha}$  (in  $\ell_2$ -minimization sense) with respect to two different kernel combination schemes can be obtained. Then, the inference can be reached by evaluating which class of samples could recover the query as accurately as possibly.

To demonstrate the proposed classification via kernel combination, an example is given in Fig. 5. The experimental setup is similar to the one in Fig. 3. The query is from  $TG_1$  set. As can be seen, both stacked features and summation kernel correctly predict the class membership of the query.

#### IV. EXPERIMENTS AND DISCUSSIONS

To verify the proposed method, extensive experiments are conducted on MSTAR database, a gallery collected using a 10 GHz SAR sensor in one-foot resolution spotlight mode. Images are captured at various depressions over a full  $0 \sim 359^\circ$  range of aspect view. The depression refers to the angle between the line of the sight from the radar to the target and the horizontal plane at the radar. The chip images are of around  $128 \times 128$  pixels in size, and cropped to  $80 \times 80$  pixels region of interest. The number of aspect views available for these targets is listed in TABLE I, where the entries in parentheses are the series number of variants with structural modifications, and the first row gives their generic labels.

TABLE I. THE NUMBER OF ASPECT VIEWS FOR DIFFERENT TARGETS

	$TG_1$	$TG_3$	$TG_2$	$TG_4$	$TG_5$	$TG_6$	$TG_7$	$TG_8$	$TG_9$	$TG_{10}$
Depr.	BMP2	T72	BTR70	2S1	BRDM2	ZSU23	BTR60	D7	T62	ZIL131
$17^\circ$	233(SN_9563)	232(SN_132)	233	299	298	299	256	299	299	299
$15^\circ$	196(SN_9566) 196(SN_c21)	195(SN_812) 191(SN_s7)	196	274	274	274	195	274	273	274

For log-Gabor filter, the parameters in (15) are set as  $\lambda_{\min} = 12$ ,  $\mu = 3$ ,  $\sigma_r = 0.28$ . To balance the computational cost and recognition accuracy, 3-scale-space is built for the monogenic signal. Following [37], the width parameter of Gaussian radial basis function is set as the median value of  $\frac{1}{\|\mathbf{f}_i - \bar{\mathbf{f}}\|}$ , where  $\mathbf{f}$  denote the feature descriptor (17), and  $\bar{\mathbf{f}}$  is the mean. The methods to be studied, including the proposed methods and the baseline algorithms, are described in TABLE II.

TABLE II. THE METHODS TO BE STUDIED IN THIS PAPER.

Abbre.	Full name (description)	Ref.	Monogenic (Y/N)	Kernelized (Y/N)
k-NN	k-Nearest-Neighbor classifier	[50]	N	N
SVM	Linear support vector machine learning	[51]	N	N
SRC	Sparse representation-based classifier	[18]	N	N
MSRC	Sparse representation of monogenic signal	[23]	Y	N
SUM	Score-level fusion via summation rule		Y	N
MAP	Score-level fusion via MAP criteria		Y	N
KSVM	Kernel support vector machine	[52]	N	Y
KSRC	Kernel sparse representation-based classifier	[37]	N	Y
KLR	Kernel linear representation-based classifier	[36]	N	Y
KLSF	Least square fitting for monogenic covariance in the reproducing kernel Hilbert space	[24]	Y	Y
CKLR1	Kernel combination via stacked features		Y	Y
CKLR2	Kernel combination via summation kernel		Y	Y

#### A. Preliminary performance verification

We first verify the preliminary performance of the proposed method. Three targets,  $TG_1$ ,  $TG_2$ , and  $TG_3$  are utilized, among which  $TG_1$  and  $TG_3$  have three variants with structural modifications, as illustrated in Fig. 6.

The standards (SN\_9563 for  $TG_1$  and SN\_132 for  $TG_3$ ) taken at  $17^\circ$  depression are used to train the algorithm, while the remaining collected at  $15^\circ$  depression are used for testing, i.e., the variants used for testing are not contained in the training set. The details can be found in TABLE I. To assess the performance under different feature spaces, four down-sampling factors  $\rho = \{\frac{1}{400}, \frac{1}{256}, \frac{1}{100}, \frac{1}{64}\}$  are tested. The results are shown in TABLE III, where the three component-specific classifiers are notated as  $\text{SRC}_{\mathcal{R}}$ ,  $\text{SRC}_{\mathcal{I}}$ ,  $\text{SRC}_{\mathcal{J}}$ .

TABLE III. THE COMPARISON OF THE PROPOSED METHODS WITH INDIVIDUAL CLASSIFIERS.

Dim.	$\text{SRC}_{\mathcal{R}}$	$\text{SRC}_{\mathcal{I}}$	$\text{SRC}_{\mathcal{J}}$	MSRC	SUM	MAP	CKLR1	CKLR2
16	0.6961	0.7669	0.8183	0.8368	0.8429	0.8542	0.8901	0.8942
25	0.8080	0.8234	0.8676	0.8932	0.8943	0.8963	0.9250	0.9271
64	0.9004	0.8522	0.9086	0.9179	0.9281	0.9302	0.9599	0.9681
100	0.9261	0.8747	0.9312	0.9322	0.9466	0.9487	0.9661	0.9722

From TABLE III, some conclusions can be drawn.

(i) The classification via data fusion significantly outperforms the individual classifications. Under 16D feature space, the recognition rates for MSRC, SUM, and MAP are 0.8368, 0.8429, and 0.8542, compared to 0.6961 for  $\text{SRC}_{\mathcal{R}}$ , 0.7669 for  $\text{SRC}_{\mathcal{I}}$ , 0.8183 for  $\text{SRC}_{\mathcal{J}}$ . The similar performance can be achieved under the other feature spaces.

(ii) The performance of score-level fusion always exceeds the one of feature-level fusion. Under 16D, 25D, 64D, and 100D feature spaces, the recognition rates for MAP are 0.8542, 0.8963, 0.9302, and 0.9487, compared to 0.8368, 0.8932, 0.9179, and 0.9322 for MSRC. It is 1.74%, 0.31%, 1.23%, and 1.65% better than the competitor. The result proves that the combination by concatenating monogenic components may results in some information loss.

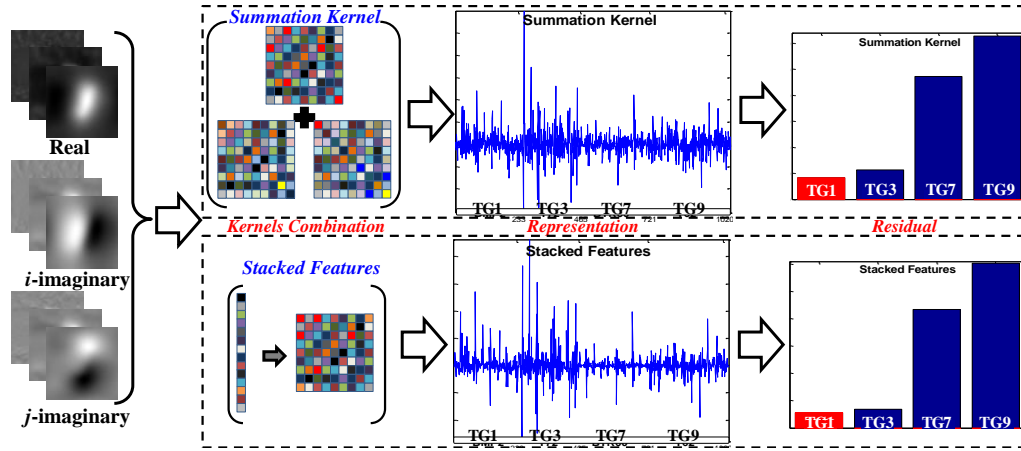


Fig. 5. Example of classification via kernel combination. The first column shows the components of the monogenic signal at different scales. The second column draws two schemes for composite kernels learning. The third and fourth columns demonstrate the representations and the corresponding residuals.

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




Fig. 6. Example target variability categories.

(iii) The performance of kernel combination consistently exceeds the one of score-level fusion. Under 100D feature space, the recognition rates for CKLR1 and CKLR2 are 0.9661 and 0.9722, compared to 0.9487 for MAP, 0.9466 for SUM. The result corroborates that the classification via score-level fusion may be not effective if the classes are not linearly separable.

(iv) The classification via summation kernel always exceeds the one via stacked features. Under 16D, 25D, 64D, and 100D feature spaces, the accuracies for CKLR2 are 0.8942, 0.9271, 0.9681, and 0.9722, compared to 0.8901, 0.9250, 0.9559, and 0.9661 for CKLR1.

#### B. Comparison with the conventional methods

This subsection devotes to the comparison with the state-of-the-art methods. Since the proposed methods are free-training, it is needed to choose the competitors likewise. Thus, k-NN, SVM, SRC [18], [21], MSRC [23], KSRC [37], [38] and KLR [36] are employed as the baseline algorithms. Four targets,  $TG_1$ ,  $TG_3$ ,  $TG_7$ , and  $TG_9$  are chosen from TABLE I. Samples of the targets are shown in Fig. 7.

To assess the accuracies of individual class, the confusion matrices as well as the overall recognition rates are shown in Fig. 8, from which the following conclusions can be reached. First, it is more difficult to differentiate two main-battle tanks ( $TG_3$  and  $TG_9$ ), whose scattering characteristics are very similar. For  $TG_3$ , even 109 of 386 query samples are recognized as



Fig. 7. Examples of  $TG_1$ ,  $TG_7$ ,  $TG_3$ , and  $TG_9$  with the aspect view of  $15^\circ$ . The first pair of targets ( $TG_1$  and  $TG_3$ ) are armored personnel carriers, while the second pair of targets ( $TG_2$  and  $TG_4$ ) are main-battle tanks.

$TG_9$  by k-NN, while the recognition rates obtained by MSRC, SRC, SUM, and MAP are 0.6736, 0.7668, 0.6476, 0.6476, and 0.6606, much lower than the accuracies of remaining targets. This is because both configuration and depression are significantly different between the images used for training and those used for testing for  $TG_1$  and  $TG_3$ . Second, the classifications in the Hilbert space always outperform the ones in the spatial domain. Compared the first row with the second row, we find the accuracy difference between two families of methods (classification in the spatial domain and the one in the Hilbert space) mainly comes from the recognition rate of  $TG_3$ . For  $TG_3$ , the recognition rates obtained by KSVM, KSRC, KLR, CKLR1, and CKLR2 are 0.7538, 0.9611, 0.9559, 0.9870, and 0.9689, compared to 0.6191 for kNN, 0.6476 for SRC, 0.6736 for MSRC, 0.6476 for SUM, and 0.6606 for MAP. The classifications via kernel combination achieve significant improvement in overall accuracy as well as individual recognition rate in comparison to the baseline methods. The results can be explained from the perspective of dataset. Since the classes are not linearly separable, the prior assumption that samples of a certain class lie on a linear subspace is not founded. Fortunately, the shortage can be covered by embedding samples into the Hilbert space with a nonlinear mapping. Then, the similarities between a pair of samples can be measured in a sufficiently rich feature space. Last but not least, the monogenic signal representation, which is capable to capture the broad spectral information and simultaneous spatial



localization with compact support, also contributes to the better performance.

Truth	k-NN [0.8339]				SRC [0.8708]				MSRC [0.8748]				SUM [0.8756]				MAP [0.8804]			
	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$
$TG_1$	349	22	16	5	377	10	4	1	364	9	10	9	379	7	3	3	380	8	2	2
$TG_2$	26	239	12	109	18	250	3	115	16	260	10	100	21	250	1	114	22	255	1	108
$TG_3$	2	0	189	4	3	0	189	3	0	0	193	2	1	0	192	2	2	0	191	2
$TG_4$	0	8	3	262	0	4	0	269	0	0	0	273	0	3	0	270	0	3	0	270
	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$
Truth	KSVM [0.8796]				KSRC [0.9045]				KLR [0.9061]				CKLR1 [0.9325]				CKLR2 [0.9349]			
	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$
$TG_1$	349	32	9	2	371	15	4	2	369	16	5	2	381	11	0	0	374	16	2	0
$TG_2$	13	291	0	82	14	299	2	71	14	305	1	66	10	324	1	51	12	327	1	46
$TG_3$	5	4	184	2	5	0	188	2	7	1	185	2	5	0	188	2	4	0	189	2
$TG_4$	1	0	0	272	0	2	2	269	0	2	1	270	0	1	0	272	0	1	0	272
	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$	$TG_1$	$TG_2$	$TG_3$	$TG_4$

Fig. 8. The results of four-target-recognition. The diagonal entries give the numbers of correct classified targets, while the others show the misclassified ones. The overall accuracies are in bracket.

### C. Target Recognition under Random Noise Corruption

For this experiment, we evaluate the performance of proposed methods in the case of strong noise. The experimental setup is similar to the one in Section IV-B. To simulate the noise, we corrupt a percentage of randomly chosen pixels from each of the query images, replacing their intensity with independent and identically distributed samples from a uniform distribution. The corrupted pixels are randomly chosen from the image, *i.e.*, the locations are randomly decided. We vary the percentage of corrupted pixels from 0 percent to 20 percent. A set of example images are given in Fig. 9.

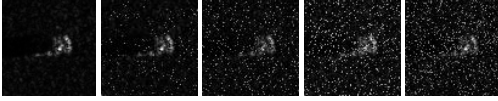


Fig. 9. Illustration of random noise corruption. The set of images demonstrate that 0%, 5%, 10%, 15%, and 20% of pixels are corrupted.

Fig. 10 draws the recognition performance as a function of the level of noise corruption. The corresponding recognition rates are tabulated in TABLE IV. Here, the robust version of SRC, which solves the extended  $\ell_1$ -norm minimization problem [18], is utilized to deal with the occlusion.

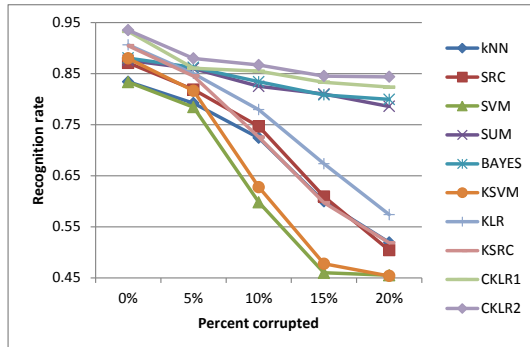


Fig. 10. The recognition rate across the range of noise corruption.

TABLE IV. THE RECOGNITION RATES ACROSS THE LEVEL OF NOISE CORRUPTION.

	0%	5%	10%	15%	20%	Drop
k-NN	0.8339	0.7921	0.7239	0.5995	0.5193	31.46%
SRC	0.8708	0.8186	0.7464	0.6091	0.5040	36.68%
SVM	0.8331	0.7841	0.5979	0.4598	0.4550	37.81%
SUM	0.8756	0.8596	0.8250	0.8098	0.7865	8.91%
MAP	0.8804	0.8620	0.8339	0.8082	0.7994	8.10%
KSVM	0.8796	0.8162	0.6276	0.4775	0.4536	42.60%
KSRC	0.9045	0.8451	0.7255	0.5971	0.5185	38.60%
KLR	0.9061	0.8499	0.7793	0.6734	0.5738	33.23%
CKLR1	0.9325	0.8603	0.8547	0.8330	0.8234	10.91%
CKLR2	0.9349	0.8876	0.8667	0.8451	0.8435	9.14%

From Fig. 10 and TABLE IV, we could see that the performances are degraded with the range of corruption increased. From 0 percent upto 20 percent corruption, the recognition rates for MAP are 0.8804, 0.8602, 0.8339, 0.8082, and 0.7994, compared to 0.8708, 0.8186, 0.7464, 0.6091, and 0.5193 for the nearest competitor. A drop of 8.10% and 8.91% for recognition accuracy are obtained by MAP and SUM, compared to the drop of 37.81%, 36.68%, and 51.93% for SVM, SRC, and k-NN. The similar performances are obtained using two proposed kernel combination methods. At 20% corruption, none of the competitors achieves the recognition rate higher than 60%, while CKLR1 and CKLR2 achieve the recognition rate of 0.8234, and 0.8435. CKLR1 and CKLR2 are 24.96% and 26.97% better than the nearest competitor. From 0 percent to 20 percent, CKLR2 generates a drop of 9.14% for recognition rate, compared to the drop of 42.60%, 38.60%, 33.23% for accuracy by KSVM, KSRC, and KLR. The experimental results prove that the proposed methods are valid in the case of strong noise. The stronger the noise, the better the proposed methods perform than the competitors.

### D. Target Recognition under Depression Variations

This subsection considers target recognition under different depressions. Three similar targets,  $TG_4$ ,  $TG_5$ , and  $TG_6$  are chosen from TABLE I, among which  $TG_5$  and  $TG_6$  have the articulated variants. The articulated refers to a physical change such as open hatch or rotated gun turret. An example of  $TG_6$  and the articulated variants are shown in Fig. 11.

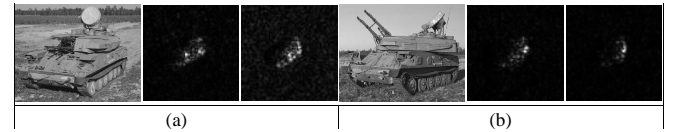


Fig. 11. Illustration of the articulated variant; (a) and (b) show SAR images of  $TG_6$  with turret straight and articulated taken at 30° and 45° depressions.

Images taken at 17° depression are used to train the algorithms, while the ones captured at 15°, 30°, and 45° are used for testing. The number of aspect views available for three targets at different depressions are listed in TABLE V, where the entries in parentheses denote the articulated variants. Example of three targets collected at different depressions are demonstrated in Fig 12.

TABLE V. ASPECT VIEWS FOR 3 TARGETS AT VARIOUS DEPRESSIONS.

Depression		$TG_4$	$TG_5$	$TG_6$	total
Training	(17°)	299	298	299	896
Testing	(15°)	274	274	274	822
	(30°)	288	287(133)	288(118)	1114
	(45°)	303	303(120)	303(119)	1148

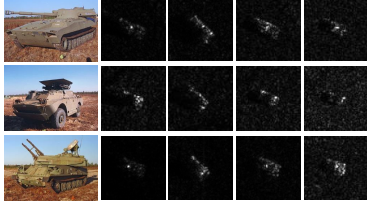
Fig. 12. Example of three vehicle targets. The first, second, and third rows demonstrate SAR images of  $TG_4$ ,  $TG_5$ , and  $TG_6$  collected at 15°, 17°, 30° and 45° depression.

Fig. 13 tabulates the overall recognition rates (in parentheses), as well as the confusion matrices obtained using various methods. The first scenario deals with a minor EOC difference on depression. It adheres to the standards set forth to train the algorithms at an operating condition of 17° depression and test them at an operating condition of 15° depression, *i.e.*, a change of 2° from 17° to 15° exists between the images used for training and those used for testing. As can be seen, all of the algorithms produce a satisfied result. The overall recognition rates for SRC, KSRC, KLR, MSRC, SUM, MAP, CKLR1, and CKLR2 are 0.9818, 0.9903, 0.9866, 0.9964, 0.9988, 0.9988, 0.9963, and 0.9975. Two score-level fusion methods, SUM and MAP, even correctly recognize 821 of 822 query samples. The second scenario evaluates a medium EOC difference on depression. The algorithms are tested at an operating condition of 30° depression, *i.e.*, there is a change of 13° from 17° to 30° between the training samples and query samples. The performance is slightly degraded. The overall recognition rates obtained by SRC, KSRC, KLR, MSRC, SUM, MAP, CKLR1, and CKLR2 are 2.04%, 1.81%, 1.80%, 1.35%, 3.30%, 2.66%, 2.51%, and 1.46% lower than the rates at 15° depression. CKLR2 and MSRC produce the best performance, 0.9829. The third scenario considers a major EOC difference on depression. The algorithms are further tested at an operating condition of 45° depression, *i.e.*, a drastic change of 28° from 17° to 45° exists between the training samples and testing samples. Obviously, the performance is sharply degraded. Two targets,  $TG_4$  and  $TG_5$  are easily to be confused due to the abrupt change in the signatures of the same target from different depression angles. The overall recognition rate for KSRC even declines by 42.78% and 44.59% compared to the ones at 15° and 30° depressions. However, the classifications via score-level fusion still achieve the overall recognition rates of 0.7186 and 0.7221, compared to 0.6394 for MSRC, 0.5453 for KLR, 0.5444 for KSRC, and 0.5366 for SRC. CKLR2 achieves the best recognition rate, 0.7456. It is 10.62%, 20.03%, 20.12%, and 20.90% better than MSRC, KLR, KSRC, and SRC. Though the recognition rates are significantly decreased compared to

the ones at 30° and 15° depressions, the proposed methods are much more robust towards the EOC difference on depression than the other baseline methods.

From the experimental results, we come the following conclusions. The proposed method achieves the performance similar to the baseline algorithms under minor and medium EOC difference on depression (or standard operating condition). However, when the major EOC difference on depression is considered, it demonstrates better advantage than the reference methods. On the other hand, the results also prove that the tolerance towards variation in depression could be improved with the monogenic signal representation.

### E. 10-Target Recognition

A more challenging problem, 10-target-recognition, is considered subsequently. All of ten targets in TABLE I are utilized. Examples of ten targets are given in Fig. 14. Similarly, images taken at 17° depression are set forth to train the algorithms, while the ones collected at 15° depression are put forward to testing. For two targets with multi-variants,  $TG_1$  and  $TG_3$ , the standards are used for training, while the remaining are used for testing, from which the EOC difference on depression and configuration are simultaneously evaluated.

TABLE VI gives the recognition accuracies. We can see the classification via score-level fusion outperforms its competitors. The recognition rates for SUM and MAP are 0.9346 and 0.9339, compared to 0.9292 for MSRC, 0.9125 for SRC, and 0.8588 for SVM. SUM is 0.54%, 2.21%, and 7.58% better than the competitors. On the other hand, the proposed classification via kernel combination outperforms the corresponding competitors too. The recognition rates for CKLR1 and CKLR2 are 0.9584 and 0.9666, compared to 0.9488 for KLSF, 0.9442 for KLR, 0.9342 for KSRC, and 0.9232 for KSVM. CKLR2 is 1.78%, 2.24%, 3.24%, and 4.34% better than these competitors.

Moreover, the classification via kernel combination outperforms the one via score-level fusion. The recognition rates for CKLR1 and CKLR2 are 0.9584 and 0.9666, 2.38% and 2.45% better than SUM, 3.20% and 3.26% better than MAP.

TABLE VI. THE OVERALL RECOGNITION RATES IN 10-TARGET-RECOGNITION.

Classification in the spatial domain					Classification in the Hilbert space				
SVM	SRC	MSRC	SUM	MAP	KSVM	KSRC	KLR	KLSF	CKLR1 CKLR2
0.8588	0.9125	0.9292	0.9346	0.9339	0.9232	0.9342	0.9442	0.9488	0.9584 0.9666

To further observe the performance to individual target, the recognition rates with respect to each target class are drawn in Fig. 15, with the corresponding confusion matrices shown in Fig. 16.

From TABLE VI, Fig. 15, and Fig. 16, we can see that most of the targets could be correctly recognized except  $TG_1$  and  $TG_3$ , whose depression and configuration are significantly different between the images used for training and those used for testing. For  $TG_3$ , it is easy to be confused with  $TG_9$  due to the similar scattering phenomenology. By combining various kinds of information contained in the monogenic signal, SUM

Ground-truth	15°	SRC (0.9818)			KSRC (0.9903)			KLR (0.9866)			MSRC (0.9964)			SUM (0.9988)			MAP (0.9988)			CKLR1 (0.9963)			CKLR2 (0.9975)			
		TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	
		TG <sub>4</sub>	272	2	0	271	3	0	271	3	0	273	1	0	273	1	0	273	1	0	271	3	0	272	2	0
		TG <sub>5</sub>	10	261	3	2	269	3	4	266	4	1	271	1	0	274	0	0	274	0	0	274	0	0	274	0
	TG <sub>6</sub>	0	0	274	0	0	274	0	0	274	0	0	274	0	0	274	0	0	274	0	0	274	0	0	274	
	30°	SRC (0.9614)			KSRC (0.9722)			KLR (0.9686)			MSRC (0.9829)			SUM (0.9658)			MAP (0.9722)			CKLR1 (0.9712)			CKLR2 (0.9829)			
		TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	
		TG <sub>4</sub>	280	8	0	279	9	0	283	5	0	282	5	1	288	0	0	281	6	1	285	3	0	287	1	0
		TG <sub>5</sub>	13	398	9	4	413	3	5	412	3	1	413	6	1	385	34	2	413	5	28	391	1	10	402	8
	TG <sub>6</sub>	6	7	393	6	9	391	9	13	384	1	5	400	2	1	403	7	10	389	0	0	406	0	0	406	
45°	SRC (0.5366)			KSRC (0.5444)			KLR (0.5453)			MSRC (0.6394)			SUM (0.7186)			MAP (0.7221)			CKLR1 (0.7238)			CKLR2 (0.7456)				
	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>	TG <sub>4</sub>	TG <sub>5</sub>	TG <sub>6</sub>		
	TG <sub>4</sub>	107	195	1	115	188	0	114	189	0	262	25	16	200	95	8	199	96	8	213	89	1	224	50	29	
	TG <sub>5</sub>	33	383	7	32	386	5	30	387	6	112	195	116	11	340	72	12	333	78	30	338	55	50	246	127	
TG <sub>6</sub>	55	241	126	51	247	124	51	246	125	78	67	277	37	100	285	29	96	297	38	104	280	9	27	386		
		Prediction																								

Prediction

Fig. 13. The performance under EOC difference on depression. In each sub-table, the row shows the ground-truth of the query sample, while the column displays the class membership predicted by the algorithms to be studied. The diagonal entries demonstrate the recognition rate of each class, while the non-diagonal entries represent the misclassification rate. The entries in parentheses are the overall recognition rates.



Fig. 14. Examples of ten vehicle targets. Images of  $TG_1$ ,  $TG_2$ ,  $TG_3$ ,  $TG_4$ ,  $TG_5$ ,  $TG_6$ ,  $TG_7$ ,  $TG_8$ ,  $TG_9$ , and  $TG_{10}$  with an aspect view of  $40^\circ$  are listed from left to right.

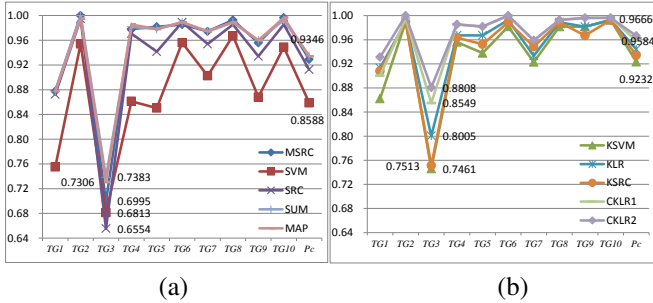


Fig. 15. The recognition rates with respect to the individual target class. Pc denotes the overall recognition rate; (a) and (b) plots the recognition rates obtained using the classifications in the spatial domain and the ones in the Hilbert space.

achieves the recognition rates of 0.7383, compared to 0.6995 for MSRC, 0.6554 for SRC, and 0.6813 for SVM. It is 3.88%, 8.29%, and 5.7% better than those competitors. Another score-level fusion scheme, MAP, also achieves the recognition rate of 0.7306, 3.11%, 7.52%, and 4.93% better than the competitors, MSRC, SRC, and SVM. The classification via kernel combination achieves the best performance. For  $TG_3$ , the recognition rate for CKLR2 is 0.8808, compared to 0.8005 for KLR, 0.7513 for KSRC, and 0.7461 for KSVM. It is 8.03%, 12.95%, and 13.47% better than the competitors. Furthermore, CKLR1 also generates the recognition rate of 0.8550, 5.4%, 10.4%, and 10.9% better than KLR, KSRC, and KSVM.

Another multi-configuration-target  $TG_1$  is also difficult to

be recognized. For example, 24.49% and 13.78% of the query samples are incorrectly recognized by SVM and KSVM. However, CKLR2 achieves the accuracy of 0.9311, compared to 0.9133 for KLR, 0.9082 for KSRC, and 0.8622 for KSVM, 1.8%, 2.3%, and 6.9% better than these competitors.

From the experimental results, we are able to come the following conclusions. For the challenging 10-target recognition problem, the classification via score-level fusion outperforms the baseline methods, especially under the nonliteral experimental setup, *e.g.*, different configurations and depressions. The classification via kernel combination always achieves the best performance. The results prove that the tolerance towards variation in configuration and depression could be improved by monogenic signal representation.

## V. CONCLUSION

In this paper, we introduce a novel classification framework on the monogenic scale-space for target recognition in SAR image. The proposed method applies multiple linear regression models to the features derived from the components of the monogenic signal at different scales, and reaches the inference by evaluating which class of samples could recover the query as accurately as possible. It is robust towards EOC difference on depression and configuration, as well as noise corruption. The feasibility of the proposed framework has been successfully tested on MSTAR database, which is a standard testbed for SAR image interpretation. From the experimental results on MSTAR database, we draw the following conclusions: (i) The monogenic signal could effectively capture the characteristics of SAR image, *e.g.*, broad spectral information and simultaneous spatial localization, with compact support, and hence produce the excellent performance for target recognition. (ii) The classification via score-level fusion could combine the benefits of individual classification while tackling their drawbacks. Thus, the significant improvement in accuracy can be obtained. (iii) The classification via kernel combination could deal with the dataset whose target classes are not linearly

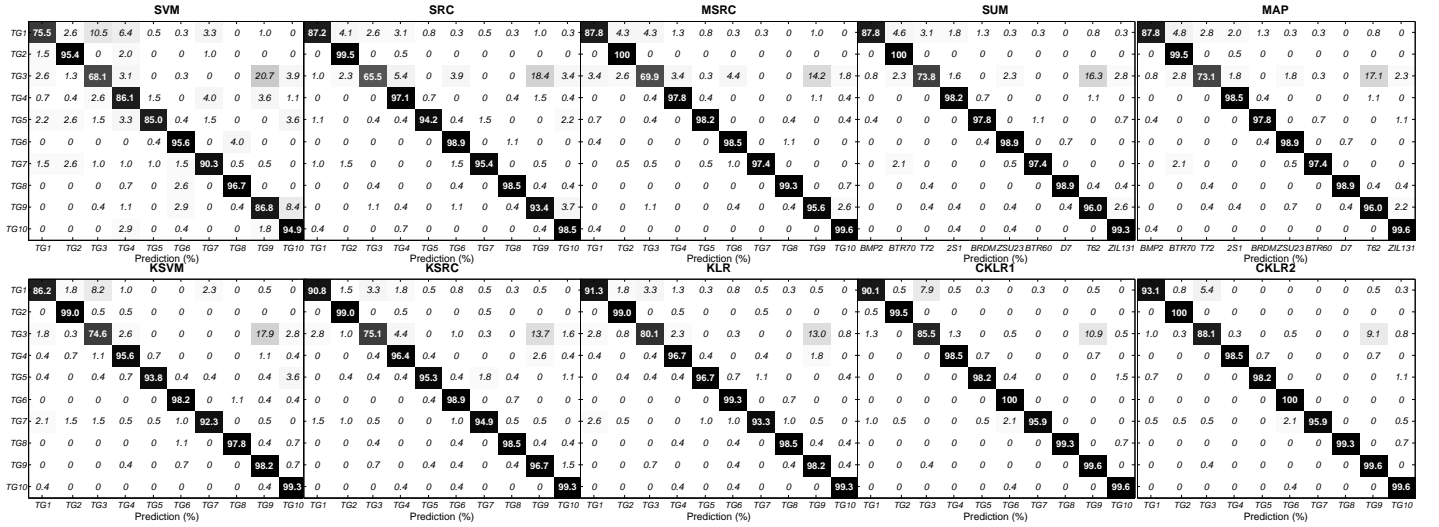


Fig. 16. The confusion matrices of ten-target-recognition. TG<sub>1</sub>, TG<sub>2</sub>, TG<sub>3</sub>, TG<sub>4</sub>, TG<sub>5</sub>, TG<sub>6</sub>, TG<sub>7</sub>, TG<sub>8</sub>, TG<sub>9</sub>, TG<sub>10</sub> denote BMP2, BTR70, T72, 2S1, BRDM\_2, ZSU23/4, BTR60, D7, T62, and ZIL131. In each sub-figure, the diagonal entries provides the recognition rates of each targets.

separable, and hence improve the recognition accuracy. (iv) The proposed classification framework is valid in the case of random noise corruption. The stronger the noise, the better the proposed method performs compared to the competitors. (v) The proposed method could improve the recognition accuracy, especially under the extended operating conditions, *e.g.*, structural modifications, variations in depression angle.

The excellent performance shown by classification on the monogenic scale-space is the direct result of coupling the monogenic signal with sparse signal modeling. Thus, an intriguing question for our future work is whether the monogenic signal and sparse representation techniques could be used to the other stages for ATR, *i.e.*, target detection and discrimination. We believe that the full potential of the monogenic signal and sparse representation in target detection and discrimination is yet to be uncovered. Secondly, the features used in Section III-B are defined via uniformly downsampling, normalization, and concatenation of the components of the monogenic signal at different scales. Then,  $\ell_1$ -norm minimization problem becomes computationally tractable. In the future, we will explore some other techniques to combine the multiple components, and improve the performance simultaneously. Thirdly, the kernel function used in Section III-C is a general function, Gaussian RBF parameterized by the width  $\gamma$ . To improve the performance further, more attention will be paid to the task-specific kernels.

#### ACKNOWLEDGMENT

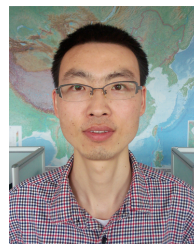
This work was supported by the National Natural Science Foundation of China under Grant 61201338 and 61401477. The authors would like to thank the handling Associate Editor and the anonymous reviewers for their constructive comments and suggestions for this paper.

#### REFERENCES

- [1] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions - a tutorial," in *Algorithms for Synthetic Aperture Radar Imagery III*. SPIE, Jun. 1996, pp. 228–242.
- [2] T. Ross, S. Worrell, V. Velten, J. Mossing, and M. Bryant, "Standard SAR ATR evaluation experiments using the MSTAR public release data set," in *Algorithms for Synthetic Aperture Radar Imagery V*. SPIE, Apr. 1998, pp. 566–573.
- [3] D. Dudgeon and R. Lacoss, "An overview of automatic target recognition," *Lincoln Lab J.*, vol. 6, no. 1, pp. 3–10, 1993.
- [4] 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.
- [5] J. Zhou, Z. Shi, X. Cheng, and Q. Fu, "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] G. Owirka, S. Verbout, and L. Novak, "Template-based SAR ATR performance using different image enhancement techniques," in *Algorithms for Synthetic Aperture Radar Imagery VI*. Orlando, Florida: SPIE, Apr. 1999, pp. 302–319.
- [7] R. Singh and B. Kumar, "Performance of the extended maximum average correlation height filter and the polynomial distance classifier correlation filter for multiclass SAR detection and classification," in *Algorithms for SAR Imagery IX, Proc. SPIE*, vol. 4727, Aug. 2002.
- [8] D. Casasent and A. Nehemiah, "Confuser rejection performance of EMACH filters for MSTAR ATR," in *Optical Pattern Recognition XVII, Proc. SPIE*, vol. 6245, 2006.
- [9] K. Copsey and A. Webb, "Bayesian gamma mixture model approach to radar target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1201–1217, Oct. 2003.
- [10] M. DeVore and J. O'Sullivan, "Quantitative statistical assessment of conditional models for synthetic aperture radar," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 113–125, Feb. 2004.
- [11] J. O'Sullivan, M. DeVore, V. Kedia, and M. Miller, "SAR ATR performance using a conditionally Gaussian model," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 1, pp. 91–108, Jan. 2001.
- [12] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.



- [13] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4321, Nov. 2006.
- [15] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 609–616.
- [16] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 801–808.
- [17] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.
- [18] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [19] V. Berisha, N. Shah, D. Waagen, H. Schmitt, S. Bellofiore, A. Spanias, and D. Cochran, "Sparse manifold learning with applications to SAR image classification," in *Proc. IEEE Conf. Acoustic, Sonar and Signal Processing (ICASSP)*, 2007, pp. 1089–1092.
- [20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality analysis by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [21] J. Thiagarajan, N. Karthikeyan, K. Peter, P. Knee, A. Spanias, and V. Berisha, "Sparse representation for automatic target classification in SAR images," in *Int'l Sym. Communcitaion, Control and Signal Processing*, 2010, pp. 1–4.
- [22] H. Zhang, N. Nasrabadi, Y. Zhang, and T. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2495, Jul. 2011.
- [23] G. Dong, N. Wang, and G. Kuang, "Sparse representation of monogenic signal: with application to target recognition in SAR images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 952–956, Aug. 2014.
- [24] G. Dong and G. Kuang, "Target recognition in SAR images via classification on Riemannian manifolds," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 199–203, Jan. 2015.
- [25] —, "Target recognition via information aggregation through Dempster-Shafer's evidence theory," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1247–1251, Jun. 2015.
- [26] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, 2001.
- [27] —, "The monogenic scale-space: A unifying approach to phase-based image processing in scale space," *J. Math. Imag. Vis.*, vol. 21, no. 1, pp. 5–26, 2004.
- [28] Stein and Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton, New Jersey: Princeton University Press, 1971.
- [29] M. Unser, D. Sage, and D. Ville, "Multiresolution monogenic signal analysis using the Riesz-Laplace wavelet transform," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2042–2058, Nov. 2009.
- [30] R. Souillard, P. Carr, and C. Fernandez-Maloigne, "Vector extension of monogenic wavelets for geometric representation of color images," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1070–1083, Mar. 2013.
- [31] M. Yang, L. Zhang, S. Shiu, and D. Zhang, "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.
- [32] M. Alessandrini, H. Liebgott, D. Friboulet, and O. Bernard, "Myocardial motion estimation from medical images using the monogenic signal," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1084–1095, Mar. 2013.
- [33] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
- [34] G. Camps-Valls, L. Gomez-Chova, J. Muoz-Mari, J. Vila-Francis, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [35] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constraint linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.
- [36] G. Dong, N. Wang, G. Kuang, and Y. Zhang, "Kernel linear representation: application to target recognition in synthetic aperture radar images," *SPIE J. App. Remote Sens.*, vol. 8, no. 1, Aug. 2014.
- [37] L. Zhang, W. Zhou, P. Chang, J. Liu, Z. Yan, T. Wang, and F. Li, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.
- [38] S. Gao, I. W. Tsang, and L. Chia, "Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 423–434, Feb. 2013.
- [39] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yang, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1043, Jun. 2010.
- [40] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.
- [41] S. Lyu, "Mercer kernels for object recognition with local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 223–229.
- [42] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [43] M. Felsberg, R. Duits, and L. Florack, "The monogenic scale space on a rectangular domain and its features," *Int'l J. Comput. Vis.*, vol. 64, no. 2, pp. 187–201, 2005.
- [44] W. Hwang, H. Wang, H. Kim, S. Kee, and J. Kim, "Face recognition system using multiple face model of hybrid fourier feature under uncontrolled illumination variation," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1152–1165, Apr. 2011.
- [45] V. Sierra-Vzquez and I. Serrano-Pedraza, "Application of Riesz transforms to the isotropic AM-PM decomposition of geometrical-optical illusion images," *J. Opt. Soc. Am. A*, vol. 27, no. 4, pp. 781–796, 2010.
- [46] G. H. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Kluwer, Netherlands: Dordrecht, 1995.
- [47] M. Felsberg and G. Sommer, "The multidimensional isotropic generalization of quadrature filters in geometric algebra," in *Proc. Int. Workshop Algebraic Frames Perception-Action Cycle*. Lecture Notes in Computer Science, Sep. 2000, pp. 175–185.
- [48] O. Fleischmann, "2D signal analysis by generalized Hilbert transforms," Ph.D. dissertation, Christian Albrechts Univ., Keil, German, Aug. 2008. [Online]. Available: <http://www.ks.informatik.uni-kiel.de/>
- [49] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 342–347, Feb. 2008.
- [50] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [51] Q. Zhao and J. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 643–654, Apr. 2001.
- [52] N. Cristianini and J. Shawe-Taylor, *Kernel Methods for Pattern Analysis*. New York: Cambridge University Press, 2006.



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

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



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