SAR Target Recognition via Sparse Representation of Monogenic Signal on Grassmann Manifolds

Ganggang Dong and Gangyao Kuang, *Member, IEEE*

Abstract—In this paper, classification via sparse representation of monogenic signal on Grassmann manifolds is presented for target recognition in SAR image. To capture the broad spectral information with maximal spatial localization of SAR image, a recently proposed vector-valued analytic signal, namely monogenic signal is exploited. Different from the conventional methods, where a single feature descriptor is generated using the monogenic signal in an Euclidean space, the multiple components of monogenic signal at various scale-spaces are viewed as points on a special type of Riemannian manifolds, Grassmann manifolds. The similarity between a pair of patterns (points) are measured by Grassmann distance metric. To exploit the nonlinear geometry structure further, we embed the sets of monogenic components into an implicit Reproducing Kernel Hilbert Space (RKHS), where the kernel-based sparse signal modeling can be learnt to reach the inference. Specifically, the sets of monogenic components resulting from the training samples are concatenated first to build a redundant dictionary. Then, the counterpart of the query is efficiently approximated by superposition of atoms of the dictionary. Notably, the representation coefficients of superposition are very parsimonious. The inference is drawn by evaluating which class of training patterns could recover the query as accurately as possible. The novelty of this paper comes from (i) the development of Grassmann manifolds formed by the multi-resolution monogenic signal; (ii) the definition of similarity between the sets of monogenic components on Grassmann manifolds for target recognition; (iii) the generalization of sparse signal modeling on Grassmann manifold; (iv) multiple comparative experiments for performance assessment.

Keywords—*The monogenic signal, Grassmann manifolds, target recognition, sparse representation, Hilbert space.*

I. INTRODUCTION

AUTOMATIC target recognition (ATR) has become a major interest for many defense applications, *e.g.*, surveillance, reconnaissance and tracking. Over the years, ATR from synthetic aperture radar (SAR) imaging sensors has been studied pervasively, yet it remains a difficult problem due to the extended operating conditions (EOC) [1], in which a single operational parameter, *e.g.*, configuration, articulated, depression, is significantly different between the images used for training and those for testing. At present, the most popularly used SAR ATR system comes from the one presented by Lincoln Laboratory [2]–[4]. It requires a sensor that could search large areas and provide fine resolution to detect and identify ground mobile targets in a variety of landscapes and deployments.

The Lincoln Laboratory ATR system comprises three separate stages: detection (or prescreening), discrimination, and classification. The first stage (prescreener), a two-parameter constant false alarm rate (CFAR) detector, passes all targets and eliminate only the obviously non-target-like naturally occurring clutters. The second stage (discriminator), a target-size matched filter, rejects all naturally occurring clutter that has been passed by the prescreener, and passes only man-made objects to the next stage. The third stage (classifier), a pattern matching algorithm, categorizes the remaining as a target of interest (of which there can be a number of class, *e.g.*, tank, armored personnel carrier, howitzer, *etc.*) or as an uninteresting man-made object. In this paper, our attention is mainly paid to the third stage, classification.

Recently, a multi-dimensional generalization of analytic signal, namely the monogenic signal, has been presented [5]. It has been built around the Riesz transform, a 2-D vector-valued extension of Hilbert transform [6]. Afterwards, the monogenic scale-space is presented by jointly considering the scale-space theory and the phase-based image processing in a unifying framework [7]. Furthermore, a new method to implement the monogenic scale-space on a rectangular domain is presented [8], where the Poisson scale space is extended to a vector valued scale space which is a harmonic vector field. Due to the nice mathematical properties of the monogenic scale space, it provides deep insight into the analysis of images. To analyze i(ntrinsic)1D and i2D local features of any curved 2D signal such as lines, edges, corners and junctions without the use of steering, a novel rotational invariant quadrature filter approach, the conformal monogenic signal has been presented in [9]. To achieve robust segmentation of ultrasound images, a new phase-based level set algorithms is presented [10], where a new speed term resulting from local phase and local orientation of monogenic signal is used to cope with attenuation artifact. To deal with the color image, the color monogenic signal has been presented in [11], where the formation of the monogenic signal has been extended to mappings with values in the vectorial part of the Clifford algebra $\mathbb{R}_{5,0}$. The presented theory is verified using two examples, color edge detection and color optical flow. To produce a parameter-free classifier, a new learning framework that leverages a key property of visual pattern discrimination, the local organizations of scales and directions is proposed in [12]. The localized comprehensive characterizations of scales and directions are enabled using steerable Riesz wavelet. Since the monogenic signal is itself a compact representation of features with little information loss and does not use steerable filters to create multi-orientation features, it is capable to capture the broad spectral informa-

tion with maximal spatial localization [13], [14], and hence applied for many pattern recognition fields. In [15], [16], the components derived from the monogenic signal at various scale spaces are encoded using local binary pattern techniques to generate a histogram for biometric recognition. Moreover, the monogenic signal is also introduced into target recognition in SAR images. In [17], an augmented monogenic feature vector is defined via uniformly down-sampling, normalization, and concatenation of the monogenic components. The resulting feature vector is fed into the classification system to draw the inference. In [18], the multiple components drawn from the monogenic signal are jointly considered into a unifying framework, where the similar constraint terms have been imposed on the component-specific classifications. In [19], a region covariance matrix descriptor is created using the monogenic signal at various scales. Since the region covariance matrix descriptor resides on a Riemannian manifold, rather than an Euclidean space, it is infeasible to be directly input to the conventional learning skills. To solve the problem, two approaches are presented in the preceding work. The first approach converts the proposed descriptor into the vector space using the logarithmic mapping, while the second one embeds it into the Hilbert space. Different from the preceding works, where the feature descriptor or learning skills are performed in the Euclidean vector space, the subspace-based learning problems are approached by reformulating it on a special type of Riemannian manifold namely Grassmann manifold [20], *i.e.*, a set of fixed-dimensional linear subspaces of a Euclidean space ($\mathbb{R}^d$). The similarity between two sets of subspaces can be then measured via distance metrics defined on Grassmann manifold, from which the decision can be made.

Recent advances in computer vision and machine learning suggest that a wide of problems can be dealt with more approximately by considering non-Eulidean geometry, *e.g.*, Riemannian manifolds [21]. In [22], the canonical correlations (or principal angle), which can be thought of as the angles between two fixed-dimensional subspaces, are utilized to compare with the set of images for object recognition. In [23], the problem in which data consist of linear subspaces instead of vectors are formulated on Grassmann manifolds, where feature extraction and classification are performed accordingly. In [24], the commonly used parametric models for videos and image sets are described using an unified framework of Grassmann and Stiefel manifolds. In [25], a discriminant analysis approach on Grassmannian manifold via a graph-embedding framework is presented for image set matching. In [26], the Grassmann manifolds is embedded into the space of symmetric matrices by an isometric mapping. Then, the commonly used learning technique, sparse coding and dictionary learning can be applied accordingly. Inspired by the works mentioned above, a novel classification approach via sparse representation on Grassmann manifold is proposed for target recognition in SAR images. Unlike the preceding works [17], [19], [27], the sets of components resulting from the monogenic signal at various scale spaces are viewed as points on Grassmann manifolds. According to the definition of Grassmann distance metrics in [20], the difference between points on Grassmann manifold (similarity of subspaces) can be measure by the distance

metric defined on Grassmann manifolds. Then, it is capable to predict the class membership of a query by comparing its monogenic signal representation with the training patterns item by item. To handle the nonlinear geometric structure in the dataset, this paper embeds the set of subspaces (points on Grassmann manifold) into an implicit Hilbert space (RKHS) by a positive definite kernel function on Grassmann space, *i.e.*, Grassmann kernel. Then, the kernel-based sparse signal modeling is learnt on Grassmann manifolds. Specifically, the sets of monogenic components resulting from the training samples are concatenated to form a redundant dictionary. The resulting dictionary is used to represent the counterpart of query as a sparse linear combination of their atoms. The optimal representation is obtained via optimizing a sparsity-constrained problem. The inference is reached according to the characteristics of optimal representation on reconstruction.

Contributions. As mentioned above, this paper introduces a classification method via sparse representation of monogenic signal on Grassmann manifolds. The sets of monogenic components at various scale spaces are regarded as points on Grassmann manifolds. The analytic comparisons of linear subspaces can be performed using Grassmann distance metrics. To the end, this paper embeds Grassmann manifold into an implicit RKHS, where the intrinsic nonlinear structure can be exploited. The decision can be made by learning kernel-based sparse signal modeling on Grassmann manifolds. Our contributions are therefore five-folds:

- We exploit the monogenic signal to capture the broad spectral information with maximal spatial localization of SAR image.
- We regard the sets of components of monogenic signal at different scale spaces as points on Grassmann manifold, rather than to produce a simple feature descriptor in an Euclidean space.
- We measure the similarity between the sets of monogenic components by Grassmann distance metric.
- We embed Grassmann manifold into an implicit Hilbert space to exploit the nonlinear geometric structure, followed by learning kernel-based sparse signal modeling to reach the inference.
- To demonstrate the advantage of proposed strategy, we perform extensive comparative experiments under the extended operating conditions.

The reminder of this paper is organized as follows. Section II briefly reviews Grassmann manifolds theory. Section III devotes to the proposed classification via sparse representation on Grassmann manifolds, which is formed by the components of monogenic signal at different scale spaces. The effectiveness of the proposed methods has been verified in Section IV by several comparative experiments on MSTAR SAR database. Finally, Section V concludes this paper.

II. BACKGROUND

This section briefly reviews the theory of Grassmann manifolds. We address the interesting reader to Ref [20], [21] for further detail descriptions.

A. Stiefel and Grassmann manifold

The space of $d \times m$-dimensional matrices ($\mathbb{R}^{d \times m}, m < d$) with orthonormal columns is a special type of Riemannian manifold known as a Stiefel manifold $\mathcal{S}(m, d)$, where the notion of normality is formulated by an orthonormality condition,

$$\mathcal{S}(m, d) \triangleq \mathbf{Y} \in \mathbb{R}^{d \times m} : \mathbf{Y}^T \mathbf{Y} = \mathbf{I}_m \quad (1)$$

for m by m identity matrix $\mathbf{I}_m$. Grassmann manifold $\mathcal{G}(m, d)$ is a mathematical object with several similarities to Stiefel manifold. It can be defined as a quotient manifold of $\mathcal{S}(m, d)$ with the equivalence relation

$$\mathbf{Y}_1 \sim \mathbf{Y}_2 \ \ if \ and \ only \ if \ \ Span(\mathbf{Y}_1) = Span(\mathbf{Y}_2) \quad (2)$$

where $Span(\mathbf{Y})$ denotes the subspace spanned by columns of $\mathbf{Y} \in \mathcal{S}(m, d)$. Thus, a Grassmann manifold $\mathcal{G}(m, d)$ is the set of m-dimensional linear subspaces of $\mathbb{R}^d$.

$\mathcal{G}(m, d)$ is a $m(d - m)$-dimensional compact Riemanian manifold. An element of $\mathcal{G}(m, d)$ can be represented by an orthonormal matrix $\mathbf{Y}$ of size d by m such that $\mathbf{Y}^T \mathbf{Y} = \mathbf{I}_m$. Thus, we usually pay more attention to $Span(\mathbf{Y})$ rather than the specific values of $\mathbf{Y}$. Since the matrix representation of a point in $\mathcal{G}(m, d)$ is not unique, two matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are considered the same if and only if $Span(\mathbf{Y}_1) = Span(\mathbf{Y}_2)$.

Propositions 1. *For Grassmann manifold $\mathcal{G}(m, d)$, $\mathbf{Y}_1 \sim \mathbf{Y}_2$ if and only if $\mathbf{Y}_1 = \mathbf{Y}_2 R_m$ for some orthonormal matrix $R_m \in \mathcal{O}(m)$, where $\mathcal{O}(\cdot)$ denotes the orthogonal group.*

A canonical distance between two subspaces is the Riemannian distance, *i.e.*, the length of geodesic path connecting two corresponding points on Grassmann manifold. A more intuitive and computationally attractive way to define the geodesic distance derives from the principal angle. If denote by $\mathbf{Y}_1, \mathbf{Y}_2$ two orthonormal matrices with $d \times m$ in size, the principal angles $0 \leq \theta_1 \leq \cdots \leq \theta_m \leq \frac{\pi}{2}$ between two subspaces $Span(\mathbf{Y}_1)$ and $Span(\mathbf{Y}_2)$, are defined recursively by

$$\cos \theta_k = \max_{u_k \in Span(\mathbf{Y}_1)} \max_{v_k \in Span(\mathbf{Y}_2)} u_k^T v_k$$
$$subject \ to \ u_k^T u_k = 1, v_k^T v_k = 1, \quad (3)$$
$$u_k^T u_j = 0, v_k^T v_j = 0, \ for \ j = 1, ..., k - 1.$$

The first principal angle θ_1 corresponds to the smallest angle between a pair of unit vectors each from two subspaces, *i.e.*, the first canonical correlation. Similarly, the k-th principal angle and canonical correlation can be described recursively.

Riemannian manifold is a special type of differentiable manifold with a smooth inner product (Riemannian metric) endowed [28]. For two tangent vectors $\triangle_1, \triangle_2$ at point $\mathbf{Y}$, its Riemannian metric is defined as

$$\langle \triangle_1, \triangle_2 \rangle_{\mathbf{Y}} = tr(\triangle_1^T (\mathbf{I}_d - \frac{1}{2} \mathbf{Y} \mathbf{Y}^T) \triangle_2) = tr(\triangle_1^T \triangle_2). \quad (4)$$

The metric (4) induces a geodesic distance, the length of shortest curve connecting two points (m-dimensional subspaces), namely arc length. It is known that the principal angles are related to the geodesic distance [20],

$$\delta_{Arc}^2(\mathbf{Y}_1, \mathbf{Y}_2) = \sum_j \theta_j^2 = \|\Theta\|_2^2 \quad (5)$$

for $\Theta = [\theta_1, \theta_2, ..., \theta_m]$.

To obtain the principal angles, it is no need to solve the maximization problem (3). Instead, it can be derived from the singular value decomposition of the product of two matrices,

$$\mathbf{Y}_1^T \mathbf{Y}_2 = \mathbf{U S V} \quad (6)$$

for unitary matrices $\mathbf{U} = [u_1, ..., u_m], \mathbf{V} = [v_1, ..., v_m]$, and the diagonal matrix $\mathbf{S} = diag(\cos(\theta_1), ..., \cos(\theta_m))$.

B. Distances for subspaces

In this subsection, *distance* refers to any assignment of nonnegative values for a pair of points in space Ω. A valid metric is a distance that satisfies the additional axioms.

Propositions 2 (Metric). *A real-valued function $\delta : \Omega \times \Omega \mapsto \mathbb{R}$ is called a metric if*

- $\delta(x_1, x_2) \geq 0$;
- $\delta(x_1, x_2) = 0$ *if and only if $x_1 = x_2$;*
- $\delta(x_1, x_2) = \delta(x_2, x_1)$;
- $\delta(x_1, x_2) + d(x_2, x_3) \geq \delta(x_1, x_3)$;

for all $x_1, x_2, x_3 \in \Omega$.

According to the distance law, it is able to come the conclusion that the Grassmann distance should be invariant under different representations.

Propositions 3. *For any distance function $\delta(\cdot, \cdot) : \mathbb{R}^{d \times m} \times \mathbb{R}^{d \times m} \mapsto \mathbb{R}$, it is a Grassmann distance if $\delta(\mathbf{Y}_1, \mathbf{Y}_2) = \delta(\mathbf{Y}_1 R_1, \mathbf{Y}_2 R_2)$, for $\forall R_1, R_2 \in \mathcal{O}(m)$.*

In the following, several distances resulting from the principal angles for subspaces are provided [20], [23].

1) Projection distance. The projection distance is defined as the 2-norm of the sine of principal angle,

$$\delta_{Pj}^2(\mathbf{Y}_1, \mathbf{Y}_2) = \sum_{j=1}^m \sin^2(\theta_j) = m - \sum_{j=1}^m \cos^2(\theta_j)$$
$$= m - \|\mathbf{Y}_1^T \mathbf{Y}_2\|_F^2 = \frac{1}{2} \|\mathbf{Y}_1 \mathbf{Y}_1^T - \mathbf{Y}_2 \mathbf{Y}_2^T\|_F^2 \quad (7)$$

where $\| \cdot \|_F$ denotes the matrix Frobenius norm. It is a Grassmann distance and a metric too.

2) Binet-Cauchy distance. The Binet-Cauchy distance is defined as the product of canonical correlations

$$\delta_B^2(\mathbf{Y}_1, \mathbf{Y}_2) = 1 - \prod_{j=1}^m \cos^2(\theta_j) = 1 - \det(\mathbf{Y}_1^T \mathbf{Y}_2) \quad (8)$$

It is invariant under different representations, and furthermore is a metric.

3) Max correlation. The max correlation distance derives from the largest canonical correlation (or the smallest principal angle)

$$\delta_{Max}^2(\mathbf{Y}_1, \mathbf{Y}_2) = 1 - \cos^2(\theta_1) = \sin(\theta_1). \quad (9)$$

It is a Grassmann distance, yet not a metric.

4) Min correlation. The min correlation distance is conceptually the opposite of the max one, *i.e.*, the smallest canonical correlation (the largest principal angle),

$$\delta^2_{Min}(\mathbf{Y}_1, \mathbf{Y}_2) = 1 - \cos^2(\theta_m) = \sin(\theta_m). \quad (10)$$

It is a Grassmann distance and also a metric.

5) Procrustes distance. The Procrustes distance is defined as the vector 2-norm of $[\sin(\frac{\theta_1}{2}), ..., \sin(\frac{\theta_m}{2})]$,

$$\delta^2_{Pc1}(\mathbf{Y}_1, \mathbf{Y}_2) = 4 \sum_{j=1}^{m} \sin^2(\frac{\theta_j}{2}). \quad (11)$$

An alternative definition is the minimum Euclidean distance between different representations of two subspaces $Span(\mathbf{Y}_1)$ and $Span(\mathbf{Y}_2)$,

$$\delta^2_{Pc1}(\mathbf{Y}_1, \mathbf{Y}_2) = \|\mathbf{Y}_1\mathbf{U} - \mathbf{Y}_2\mathbf{V}\|^2_F$$
$$= \min_{R_1, R_2 \in \mathcal{O}(m)} \|\mathbf{Y}_1 R_1 - \mathbf{Y}_2 R_2\|^2_F \quad (12)$$

Another version of the Procrustes distance is to use the matrix 2-norm, rather than the matrix F-norm,

$$\delta^2_{Pc2}(\mathbf{Y}_1, \mathbf{Y}_2) = \|\mathbf{Y}_1\mathbf{U} - \mathbf{Y}_2\mathbf{V}\|^2_2 = 2\sin(\frac{\theta_m}{2}) \quad (13)$$

Obviously, the first and second Procrustes distances are derived by embedding the Grassmann manifold in the vector space $\mathbb{R}^{m \times d}$, then using 2- and Frobenius-norms in these spaces.

6) Fubini-Study distance. The Fubini-Study distance is derived via the Plucker embedding of $\mathcal{G}_{m,d}$ into the projective space, then using the Fubini-Study metric,

$$\delta^2_{FS}(\mathbf{Y}_1, \mathbf{Y}_2) = \arccos \| \det(\mathbf{Y}_1^T\mathbf{Y}_2)\|$$
$$= \arccos(\prod_{j=1}^{m} \cos(\theta_j)). \quad (14)$$

C. Kernel functions for subspaces

In this subsection, two kernel functions are provided, with which the metrics defined in Section II-B can be compatible.

Propositions 4. *A Grassmann kernel function is a positive definite kernel function on $\mathcal{G}(m, d)$ [23].*

1) Projection kernel. The projection distance can be understood by associating a point $Span(\mathbf{Y}) \in \mathcal{G}(m, d)$ with its projection matrix $\mathbf{YY}^T$ by an embedding, $\Psi_{Pj}(\cdot)$: $\mathcal{G}(m, d) \mapsto \mathbb{R}^{d \times d}$, *i.e.*, $Span(\mathbf{Y}) \mapsto \mathbf{YY}^T$. The map Ψ_{Pj} is an isometric embedding, and hence the Projection distance is simply an Euclidean distance in $\mathbb{R}^{d \times d}$. The corresponding inner product can be expressed as $\langle \mathbf{Y}_1, \mathbf{Y}_2 \rangle_{\mathcal{G}} = tr[(\mathbf{Y}_1\mathbf{Y}_1^T)(\mathbf{Y}_2\mathbf{Y}_2^T)] = \|\mathbf{Y}_1^T\mathbf{Y}_2\|^2_F$, and therefore the projection kernel

$$\kappa_{Pj}(\mathbf{Y}_1, \mathbf{Y}_2) = \|\mathbf{Y}_1^T\mathbf{Y}_2\|^2_F \quad (15)$$

is a Grassmann kernel.

2) Binet-Cauchy kernel. Another popularly used Grassmann kernel is the Binet-Cauchy kernel,

$$\kappa_{BC}(\mathbf{Y}_1, \mathbf{Y}_2) = (\det(\mathbf{Y}_1^T\mathbf{Y}_2))^2 \quad (16)$$

which induces the Binet-Cauchy distance (8).

III. SPARSE MODELING ON GRASSMANN MANOIFOLD

The monogenic signal, a 2-D extension of analytic signal, could seek the broad spectral information with maximal spatial localization, and hence is used to represent SAR image.

A. The monogenic signal

The analytic signal for real signal $f(x)$, is defined as

$$f_A(x) = f(x) + j\mathbf{H}\{f\}(x), x \in \mathbb{R} \quad (17)$$

where $\mathbf{H}\{\cdot\}$ denotes the Hilbert transform whose kernel is $h(x) = \frac{1}{\pi x}$. It is usually used to produce a complex-valued representation for 1-D signal. The Riesz transform is a natural isotropic extension of the Hilbert transform to multidimensional signal. For signal $f \in L^p(\mathbb{R}^n)$, Riesz transform is defined as

$$f_{R_j}(\mathbf{x}) \equiv \mathcal{R}_j(f)(\mathbf{x}) = h_j(\mathbf{x}) * f(\mathbf{x}) \quad (18)$$

where $h_j(\mathbf{x}) = c_n \frac{x_j}{\|\mathbf{x}\|^{n+1}}, \mathbf{x} = [x_1, ..., x_n] \in \mathbb{R}^n$ is the convolution kernel. For 2-D signal $f(\mathbf{x}), \mathbf{x} = [x_1, x_2]$, the Riesz transform can be expressed as

$$f_{R_2}(\mathbf{x}) = h_2(\mathbf{x}) * f(\mathbf{x}) = \left[h_2^1(\mathbf{x}) * f(\mathbf{x}), h_2^2(\mathbf{x}) * f(\mathbf{x}) \right]^T, \quad (19)$$

where $h_2^1(\mathbf{x})$ and $h_2^2(\mathbf{x})$ are the first and second-order Riesz kernels.

The linear combination of a 2-D signal with the Riesz transformed one yields a sophisticated 2-D analytic, namely the monogenic signal [5],

$$f_M(\mathbf{z}) = f(\mathbf{z}) - (i, j)f_{R_2}(\mathbf{z}) \quad (20)$$

where the units $\{i, j, 1\}$ forms an orthonormal basis of $\mathbb{R}^3$. From the viewpoint of Clifford algebra, the monogenic signal can be viewed as a special form of quaternion signal decomposition ($e = e_1 + ie_2 + je_3 + ke_4$) [29],

$$f_M(\mathbf{x}) = f(\mathbf{x}) - i(h_2^1 * f)(\mathbf{x}) - j(h_2^2 * f)(\mathbf{x}) \quad (21)$$

where the k-component is absent. Then, an intuitive idea is to decompose the original signal under Cartesian coordinate, *i.e.*, f, $f_1 = (h_2^1 * f)(\mathbf{x})$ and $f_2 = (h_2^2 * f)(\mathbf{x})$. Moreover, the original signal can be also orthogonally decomposed into three components under the polar coordinate, local amplitude, local phase, and local orientation [5], which are defined as

$$amplitude : A = \sqrt{f(\mathbf{z})^2 + |f_{R_2}(\mathbf{z})|^2}$$
$$phase : \phi = atan2(|f_{R_2}(\mathbf{z})|, f(\mathbf{z})) \in (-\pi, \pi] \quad (22)$$
$$orientation : \theta = atan(f_2(\mathbf{z})/f_1(\mathbf{z})) \in (-\frac{\pi}{2}, \frac{\pi}{2}]$$

To demonstrate phase and orientation components of monogenic signal, a simple diagram is given in Fig. 1.

The practical signal is of finite length, it is therefore needed to be infinitely extended by a bandpass filter to handle the broad Fourier spectra [8]. Then, the monogenic signal representation evolves into the process of computing multiple components of a bandpass version of the signal,

$$f_M(\mathbf{z}) = (h_{bp} * f)(\mathbf{z}) - (i, j)h_{bp} * f_R(\mathbf{z}) \quad (23)$$
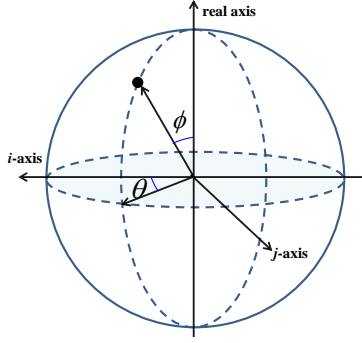
Fig. 1. The phase and orientation components of monogenic signal. The former represents the signal structural information, while the latter provides the geometric information.

where h_{bp} denotes the bandpass filter kernel.

To maintain the odd property of Riesz kernels with null DC component, the bandpass filter h_{bp} should be symmetric [30]. Considering the mathematical, computational, and empirical reasons, this paper employs log-Gabor filter, whose expression in the frequency domain is

$$G(\omega) = \exp\left\{ -\frac{[\log(\omega/\omega_0)]^2}{[\log(\sigma/\omega_0)]^2} \right\}, \qquad (24)$$

where ω_0 is the center frequency, and σ is the scaling factor of the bandwidth. The log-Gabor filter could capture the broad spectral information of the signal with compact spatial filter, as recommended in the preceding works [15]–[17], [19]. By tuning the scaling factor σ and center frequency ω_0, it is able to produce the multiresolution monogenic signal representation. Then, the monogenic signal is embedded into a log-Gabor scale space by convolving its components with the bandpass filter h_{bp} [7].

B. Grassmann Manifold Formed by The Monogenic Signal

Different from the preceding works, where a single feature descriptor is produced by the monogenic components in an Euclidean space, the sets of monogenic components at different scale spaces are viewed as points on Grassmann manifold $\mathcal{G}$.

Given S-scale log-Gabor filter bank. For 2-D signal (or image) f with $w \times h$ pixels in size $f \in \mathbb{R}^{w \times h}$, the monogenic signal at different scales can be shown as $f_M^{(1)}, f_M^{(2)}, \cdots, f_M^{(S)}$, with the corresponding components under the polar coordinate expressed as,

$$\underbrace{A^{(1)}, \phi^{(1)}, \theta^{(1)}}_{f_M^{(1)}}, ..., \underbrace{A^{(k)}, \phi^{(k)}, \theta^{(k)}}_{f_M^{(k)}}, ..., \underbrace{A^{(S)}, \phi^{(S)}, \theta^{(S)}}_{f_M^{(S)}} \quad (25)$$

where $\{A^{(k)}, \phi^{(k)}, \theta^{(k)}\}_{k=1}^{S} \in \mathbb{R}^{w \times h}$ denote amplitude, phase, orientation components of monogenic signal at the k-th scale space. For each element in (25), it is reshaped as a single vector in $\mathbb{R}^{wh}$ by raster-scanning the columns. Then, the set of monogenic components (denote by $\mathbf{f}$) is in fact a wh-by-$3S$-dimensional matrix $f \in \mathbb{R}^{wh \times 3S}$,

$$\mathbf{f} = [A^{(1)(:)}, \phi^{(1)}(:), \theta^{(1)}(:), ..., A^{(S)(:)}, \phi^{(S)}(:), \theta^{(S)}(:)] \quad (26)$$

Though the column vectors of $\mathbf{f}$ are not orthonormal to each other, it is easy to produce a set of orthonormal basis for the range of $\mathbf{f}$ by some mathematical manipulations, *e.g.*, Schmidt orthogonalization law, and hence to satisfy the orthonormality condition (1), *i.e.*, $\mathbf{f}^T\mathbf{f} = \mathbf{I}$. Considering the equivalence relation (2), $\mathbf{f}_1 \sim \mathbf{f}_2$ if and only if $Span(\mathbf{f}_1) \sim Span(\mathbf{f}_2)$, it is reasonable to regard the set of monogenic components $\mathbf{f}$ to be a point on Grassmann manifold $\mathcal{G}(m, d)$ where $m = 3S$ and $d = wh$.

For a pair of the sets of monogenic components resulting from two images, $\mathbf{f}_1$ and $\mathbf{f}_2$, the distance between $Span(\mathbf{f}_1)$ and $Span(\mathbf{f}_2)$ can be quantitatively measured by the principal angles $\Theta = [\theta_1, ..., \theta_m]$ using the inner product of vectors according to (3), as pictorially shown in Fig. 2. By comparing the query sample with the training ones on Grassmann manifold item by item, it is able to predict the class membership of the query by a simple nearest neighbor classifier, as briefly summarized in Algorithm 1.
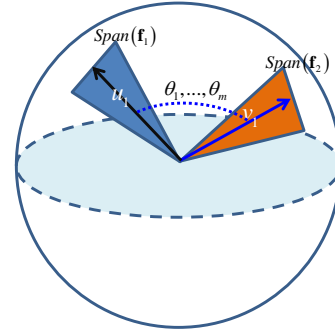

Fig. 2. Principal angles and Geodesic distance. $Span(\mathbf{f}_1)$ and $Span(\mathbf{f}_2)$ denote two subspaces formed by multiple components of monogenic signal. The similarity between two subspaces is quantified by the principal angles, as shown in (5).

Algorithm 1 K-Nearest-Neighbor-Classification on Grassmann Manifold (kNN$^{\mathcal{G}}$)

Input: $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$: a set of training samples for L target classes; $\mathbf{y}$: a query sample;
Output: identification of $\mathbf{y}$.
1: Produce the multiple monogenic representation of query sample, $\mathbf{f}_\mathbf{y} \in \mathbb{R}^{d \times m}$ by (25); Generate the corresponding orthonormal basis $Span(\mathbf{f}_\mathbf{y})$; initial $k = 1$;
2: **for** $k \leftarrow 1, N$ **do**
3:　　Represent the kth training sample by the components of monogenic signal at different scale-space, $\mathbf{f}_k \in \mathbb{R}^{d \times m}$;
4:　　Produce the orthonormal basis for $\mathbf{f}_k$;
5:　　Quantify the similarity between points, $Span(\mathbf{f}_\mathbf{y})$ and $Span(\mathbf{f}_k)$ via Grassmann distance $\delta(\mathbf{f}_\mathbf{y}, \mathbf{f}_k)$;
6:　　Record the similarity $dist(k) = \delta(\mathbf{f}_\mathbf{y}, \mathbf{f}_k)$;
7: **end for**
8: identity($\mathbf{y}$)=identity($\arg\min_k\{dist(k)\}$).

Given N training samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ from L distinct classes, we are capable to produce N training patterns on

$\mathcal{G}(m,d)$ by the components of monogenic signal,

$$Span(\mathbf{f}_1), Span(\mathbf{f}_2), ..., Span(\mathbf{f}_N)$$

where $\{\mathbf{f}_j\}_{j=1}^N$ denotes the monogenic representation of the k-th training sample $\mathbf{x}_k$. For a query samples $\mathbf{y}$, we compare its monogenic representation $Span(\mathbf{f_y})$ with the training patterns item by item (as described in Section II-B). According to the idea of nearest neighbor classification, the identity of the query can be assigned as the one nearest to it. Provided a proper distance metric on Grassmann manifolds, the problem of classifying the set of monogenic components are converted to point to point classification on $\mathcal{G}(m,d)$. To demonstrate the Grassmann distance for classification, a conceptual example are drawn in Fig. 3.
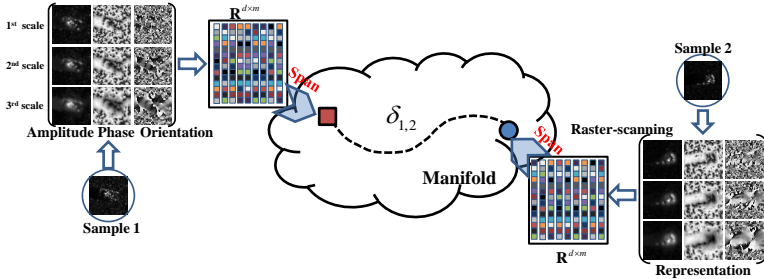

Fig. 3. Illustration of the distance between a pair of samples on Grassmann manifolds. The sets of monogenic components are viewed as points on Grassmann manifold $\mathcal{M}$. The problem of classifying a query sample evolves into searching the nearest neighbor points on Grassmann manifold.

C. Sparse Signal Modeling on Grassmann Manifold

To exploit the nonlinear geometric structure further, we embed Grassmann manifolds into an implicit Reproducing Kernel Hilbert Space (defined in Law (5)), where sparse signal modeling can be learnt to reach the inference.

Propositions 5. *For a nonempty set Ω and a Hilbert space of functions $(g : \Omega \mapsto \mathbb{R})$ $\mathcal{H}$, $\mathcal{H}$ is a Reproducing Kernel Hilbert Space endowed with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, if there exists a function $\kappa : \Omega \times \Omega$ with the following properties:*

- $\langle g, \kappa(x, \cdot) \rangle = g(x)$ *for all* $g \in \mathcal{H}$, *in particular* $\langle \kappa(x, \cdot), \kappa(y, \cdot) \rangle = \kappa(x, y)$;
- κ *spans* $\mathcal{H}$, *that is,* $\mathcal{H} = \overline{Span\{\kappa(x, \cdot) \| x \in \Omega\}}$, *where* $\overline{\mathbf{X}}$ *denotes the completion of the set* $\mathbf{X}$.

Given N training patterns on Grassmann manifolds, $\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_N$, we project them into the RKHS by an embedding function $\phi : \mathcal{G}(m,d) \mapsto \mathcal{H}$. The images of training patterns in RKHS are then concatenated to form a redundant dictionary,

$$\mathcal{D} = [\phi(\mathcal{P}_1), \phi(\mathcal{P}_2), ..., \phi(\mathcal{P}_N)].$$

For any query point on Grassmann manifolds $\mathcal{P}_\mathbf{y}$, the counterpart image in RKHS is expressed as a linear combination of atoms of $\mathcal{D}$,

$$\phi(\mathcal{P}_\mathbf{y}) = \phi(\mathcal{P}_1)\alpha_1 + \phi(\mathcal{P}_2)\alpha_2 + \cdots + \phi(\mathcal{P}_N)\alpha_N = \mathcal{D}\alpha \quad (27)$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_N] \in \mathbb{R}^N$ is the weight coefficients. The system (27) is actually a set of linear regression models, where the training patterns $\phi(\mathcal{P}_1), \phi(\mathcal{P}_2), ..., \phi(\mathcal{P}_N)$ play the role of regressors, and $\phi(\mathcal{P}_\mathbf{y})$ is the corresponding response. Then, the next question is how to produce the unique representation for the linear regression models.

The preceding works claim that the sparest representation is naturally discriminative and hence present a creative framework, namely sparse representation-based classification [31]. The comprehensive review for sparse signal representation in machine learning and computer vision can be addressed to Ref [32]–[34]. Moreover, sparse signal modeling has also been introduced into Riemannian manifolds. The representative applications include image set matching, action recognition, and video-based recognition [26], [35]–[42].

Inspired by the preceding works, this paper proposes to seek the parsimonious representation by sparsity constraint (ℓ_1-nrom minimization),

$$\min_\alpha \|\alpha\|_1 \; subject \; to \; \|\phi(\mathcal{P}_\mathbf{y}) - \mathcal{D}\alpha\|_2^2 < \varepsilon \quad (28)$$

where $\| \cdot \|_1$ is the ℓ_1-norm. It sums the absolute value of all elements. ε is the allowed error tolerance. By some mathematical manipulations, the constraint in (28) changes to a penalty as following,

$$\min_\alpha \|\phi(\mathcal{P}_\mathbf{y}) - \mathcal{D}\alpha\|_2^2 + \lambda\|\alpha\|_1 \quad (29)$$

where λ is a constant to make a tradeoff between the sparsity level and the fidelity (accuracy of reconstruction). Since ℓ_1-norm is not differentiable at the origin, (29) is not differentiable, although it is convex. To solve the problem (29), we resort to the kernel tricks [43], [44], as popularly recommended in the conventional works on machine learning [45]–[47].

To simplify the objective function, this paper unfolds the reconstruction term of (28),

$$\begin{aligned} g(\alpha) &= \|\phi(\mathcal{P}_\mathbf{y}) - \mathcal{D}\alpha\|_2^2 \\ &= (\phi(\mathcal{P}_\mathbf{y}) - \mathcal{D}\alpha)^T(\phi(\mathcal{P}_\mathbf{y}) - \mathcal{D}\alpha) \quad (30) \\ &= \phi(\mathcal{P}_y)^T\phi(\mathcal{P}_y) - 2\phi(\mathcal{P}_y)^T\mathcal{D}\alpha + \alpha^T\mathcal{D}^T\mathcal{D}\alpha \end{aligned}$$

Thanks to the Grassmann kernel, (30) can be further simplified as

$$\phi(\mathcal{P}_y)^T\phi(\mathcal{P}_y) = \langle \phi(\mathcal{P}_y), \phi(\mathcal{P}_y) \rangle = \kappa(\mathcal{P}_y, \mathcal{P}_y)$$

,

$$\phi(\mathcal{P}_y)^T\mathcal{D} = \begin{bmatrix} \langle \phi(\mathcal{P}_y), \phi(\mathcal{P}_1) \rangle \\ \langle \phi(\mathcal{P}_y), \phi(\mathcal{P}_2) \rangle \\ \vdots \\ \langle \phi(\mathcal{P}_y), \phi(\mathcal{P}_N) \rangle \end{bmatrix} = \begin{bmatrix} \kappa(\mathcal{P}_y, \mathcal{P}_1) \\ \kappa(\mathcal{P}_y, \mathcal{P}_2) \\ \vdots \\ \kappa(\mathcal{P}_y, \mathcal{P}_N) \rangle \end{bmatrix} \quad (31)$$

and (32).

Obviously, $\phi(\mathcal{P}_y)^T\mathcal{D} \in \mathbb{R}^N$ is a similarity vector whose entry denotes the similarity between the query and the training patterns. $\mathcal{D}^T\mathcal{D} \in \mathbb{R}^{N \times N}$ is a kernel Gram matrix whose element gives the distance between a pair of training samples. Then, the optimization problem (29) is tractable since it only

$$\mathcal{D}^T\mathcal{D} = \begin{bmatrix} \langle \phi(\mathcal{P}_1), \phi(\mathcal{P}_1) \rangle & \langle \phi(\mathcal{P}_1), \phi(\mathcal{P}_2) \rangle & \cdots & \langle \phi(\mathcal{P}_1), \phi(\mathcal{P}_N) \rangle \\ \langle \phi(\mathcal{P}_2), \phi(\mathcal{P}_1) \rangle & \langle \phi(\mathcal{P}_2), \phi(\mathcal{P}_2) \rangle & \cdots & \langle \phi(\mathcal{P}_2), \phi(\mathcal{P}_N) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \phi(\mathcal{P}_N), \phi(\mathcal{P}_1) \rangle & \langle \phi(\mathcal{P}_N), \phi(\mathcal{P}_2) \rangle & \cdots & \langle \phi(\mathcal{P}_N), \phi(\mathcal{P}_N) \rangle \end{bmatrix} = \begin{bmatrix} \kappa(\mathcal{P}_1,\mathcal{P}_1) & \kappa(\mathcal{P}_1,\mathcal{P}_2) & \cdots & \kappa(\mathcal{P}_1,\mathcal{P}_N) \\ \kappa(\mathcal{P}_2,\mathcal{P}_1) & \kappa(\mathcal{P}_2,\mathcal{P}_2) & \cdots & \kappa(\mathcal{P}_2,\mathcal{P}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathcal{P}_N,\mathcal{P}_1) & \kappa(\mathcal{P}_N,\mathcal{P}_2) & \cdots & \kappa(\mathcal{P}_N,\mathcal{P}_N) \end{bmatrix} \tag{32}$$

refers to the kernel operation $\kappa(\cdot,\cdot)$ (or inner product) instead of the mapping ϕ.

Provided Grassmann kernel function, it is easy to attain the optimal representation $\hat{\alpha}$ via convex optimization toolkits, because the objective function (29) is differential and convex. Have computed the optimal representation $\hat{\alpha}$ of the regression models (27), the inference can be reached by evaluating which class of training patterns could recover the query one as accurately as possible,

$$\min_k \|\phi(\mathcal{P}_y) - \mathcal{D}\psi_k(\hat{\alpha})\|_2^2 \tag{33}$$

where $\psi_k(\cdot)$ maintains the elements associated with the k-th class and sets the remaining to be zeros. The proposed method is pictorially demonstrated in Fig. 4, the corresponding procedure is given in Algorithm 2.
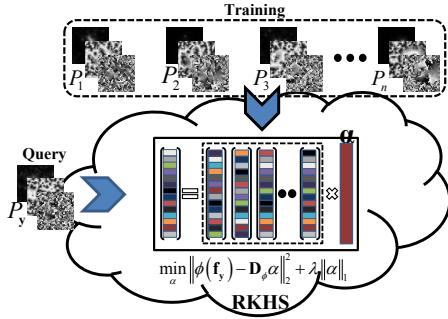

Fig. 4. Pictorial demonstration of the proposed method. Grassmann manifold is embedded into an implicit Hilbert space, in which the training patterns (points on Grassmann manifold) are used to build a redundant dictionary to encode the query one as a sparse linear combination of them. The inference can be reached by evaluating which class of training pattern could recover the query one as accurately as possible.

D. Computational Complexity

The computational cost of Algorithm 1 consists in the monogenic signal representation, orthogonalization, and the distance metric quantization. The quantization of distance between two points on Grassmannian manifolds refers to singular value decomposition, which has the computational complexity of $\mathcal{O}(m^3)$. Then we could approximate the whole computational complexity of Algorithm 1 as $\mathcal{O}(dm + d^2m + m^3)$. Similarly, the computational burden of Algorithm 2 composes of the monogenic signal representation, orthogonalization, the quantization of similarity between query and training samples, ℓ_1-norm minimization optimization. Therefore, the whole procedure of Algorithm 2 demands for $\mathcal{O}(dm + d^2m + m^3 + N^2L)$

Algorithm 2 Sparse Representation on Grassmann Manifold

Input: $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$: a set of training samples for L target classes;
　　$\mathbf{y}$: a query sample;
　　λ: a regularization parameter;
Output: identification of $\mathbf{y}$.
1: Select a Grassmann kernel function;
2: Compute the monogenic representation of training samples, $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N$ and the query one, $\mathbf{f_y}$ by (25);
3: Derive the orthonormal basis for $\{\mathbf{f}_j\}_{j=1}^N$ and $\mathbf{f_y}$;
4: **for** $i \leftarrow 1, N$ **do**
5: 　Compute the similarity of query and the i-th training sample, $\kappa(\mathcal{P}_y, \mathcal{P}_i) \rightarrow [\phi(\mathcal{P}_y)^T\mathcal{D}]_i$;
6: 　**for** $j \leftarrow 1, N$ **do**
7: 　　Quantify the similarity of the i-th and j-th training samples, $\kappa(\mathcal{P}_i, \mathcal{P}_j) \rightarrow [\mathcal{D}^T\mathcal{D}]_{i,j}$;
8: 　**end for**
9: **end for**
10: Solve the ℓ_1-norm minimization problem (28) to obtain the sparsest representation coefficients $\hat{\alpha}$;
11: Compute the residual associated with each target class;
12: Reach the inference by seeking the minimum residual.

---

operations for sparsity L. Though the proposed method is computationally complex than the conventional ones, it can be easily overcome with the developed hardware device.

IV. EXPERIMENTS AND DISCUSSIONS

To verify the proposed method, extensive comparative experiments are performed on MSTAR dataset, a gallery collected using a 10 GHz SAR sensor with $1ft \times 1ft$ resolution in range and azimuth. The images are collected at various depression angles over $0 \sim 359°$ 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 images are of around 128×128 pixels in size, and cropped to 64×64 pixels, with which the center patch is preserved. To make the computation tractable, the original image is downsampled by a factor of ρ. To balance the computational cost and accuracy, 3-scale log-Gabor filter bank is formed to produce the multi-resolution monogenic signal representation. The parameters of log-Gabor filters are set as following: $\frac{\sigma}{\omega_0} = 0.28$. The center frequency is tuned as a function of minimum wavelength, $\omega_0 = \frac{1}{\lambda_m \mu^{s-1}}$ for $\lambda_m = 12$ and $\mu = 3$.

A. Performance of Grassmann Distance Metrics

For a point in Grassmann manifold $\mathbf{Y} \in \mathcal{G}(m, d)$, we usually pay more attention to the linear subspaces spanned by the

column vectors of $\mathbf{Y}$, rather than the specific values of $\mathbf{Y}$. Thus, it is capable to classify a query point by measuring the distance between the query point and the training ones, as described in Section III-B.

To verify the performance of various Grassmann distance metrics (shown in Section II-B), a four-class target recognition experiment is performed. Four vehicle targets, BMP2, T72, BTR60, and T62 are employed, among which BMP2 and T72 have several variants marked by serial number. The number of aspect view images available for each target is listed in TABLE I, where images used for training are in bold, while the ones available for testing are underlined. The original image (a finite 2-D signal) is orthogonally decomposed into three components, local amplitude, local phase, and local orientation (under polar coordinate) to form linear subspaces. The choice of images available for training and those for testing will be explained in Section IV-C1.

TABLE I. THE NUMBER OF ASPECT VIEWS FOR DIFFERENT TARGETS

Depr.	BMP2	T72	BTR60	T62	Total
17°	**233** (*SN_9563*) 232 (*SN_9566*) 233 (*SN_c21*)	**232** (*SN_132*) 231 (*SN_812*) 228 (*SN_s7*)	**256**	**299**	**1020**
15°	195 (*SN_9563*) 196 (*SN_9566*) 196 (*SN_c21*)	196 (*SN_132*) 195 (*SN_812*) 191 (*SN_s7*)	195	273	1246

TABLE II. THE RECOGNITION RATES OBTAINED BY VARIOUS DISTANCE METRICS; (A) $\rho = \frac{1}{64}$, (B) $\rho = \frac{1}{16}$.

Dist.	δ_{Arc}	δ_{Pj}	δ_{BC}	δ_{Max}	δ_{Min}	δ_{Pc1}	δ_{Pc2}	δ_{FS}
	$\rho = \frac{1}{64} \mapsto \mathcal{G}(64,9)$							
NN	0.8668	0.8844	0.8547	0.5947	0.7255	0.8708	0.7255	0.8547
kNN	0.8941	0.8925	0.8700	0.6132	0.7456	0.8973	0.7456	0.8700
	$\rho = \frac{1}{16} \mapsto \mathcal{G}(256,9)$							
NN	0.9021	0.8941	0.8957	0.6100	0.7921	0.9053	0.7921	0.8957
kNN	0.9133	0.9077	0.9101	0.6172	0.8226	0.9109	0.8226	0.9101

TABLE II gives the recognition performance of various Grassmann distance metrics obtained by Algorithm 1. According to the experimental results, we could come the conclusion as follows:

- The recognition performance of min correlation is equivalent to the ones obtained by the second version of Procrustes distance. This is because the second version of Procrustes distance has the immediate relationship with the min correlation distance, as described by

$$\delta_{Min}^2(\mathbf{Y}_1, \mathbf{Y}_2) = \sin^2 \theta_m = 1 - (1 - 2\sin^2(\frac{\theta_m}{2}))^2$$

where $\delta_{Pc2}^2(\mathbf{Y}_1, \mathbf{Y}_2) = 4\sin^2(\frac{\theta_m}{2})$. Therefore, the Fubini-Study distance behaves similarly to the Binet-Cauchy distance, because the former is monotonically related to the latter.

- The performances obtained by $\delta_{Max}, \delta_{Min}$, and δ_{Pc2} are much worse than the remaining metrics. With the downsampled factor of $\frac{1}{64}$, the recognition rates for $\delta_{Max}, \delta_{Min}$, and δ_{Pc2} are 0.5947, 0.7255, 0.7255 by

kNN, compared to 0.8668 for δ_{Arc}, 0.8844 for δ_{Pj}, 0.8547 for δ_{BC}, 0.8708 for δ_{Pc1}, and 0.8547 for δ_{FS}. This is because three distances mentioned above only refers to a single principal angle (θ_1 or θ_m), and hence are susceptible to noise.

- Though both δ_{Max} and δ_{Min} utilize a single principal angle, their performances are drastically different. The best recognition rate for δ_{Max} is 0.6172, compared to 0.8226 for δ_{Min}. This is because δ_{Max} will be close to zeros for most of the data when the subspaces are concentrated and have nonzero intersections.

- By using all principal angles, the arc length δ_{Arc}, the Projection distance δ_{Pj}, and the first Procrustes distance δ_{Pc1} consistently achieve the best performance in comparison to the others'. The results can be interpreted as they are better suited for discriminating the subspaces of MSTAR database.

B. Cartesian Coordinate and Polar Coordinate

In Section III-A, we provide two approaches to generate the monogenic components, *i.e.*, the decomposition under Cartesian coordinate and polar coordinate. The former decomposition is redundant, while the latter one is orthogonal [5], [15]. To assess which approach is more effective to target recognition in SAR images, several experiments is performed. The experimental setup is similar to the one in Section IV-A. The experimental results are shown in TABLE III.

TABLE III. THE RECOGNITION RATES OBTAINED BY VARIOUS DISTANCE METRICS UNDER CARTESIAN COORDINATE AND POLAR COORDINATE ($\rho = \frac{1}{16}$).

Dist.	δ_{Arc}	δ_{Pj}	δ_{BC}	δ_{Max}	δ_{Min}	δ_{Pc1}	δ_{Pc2}	δ_{FS}
	NN							
Cart.	0.8852	0.8852	0.8772	0.7055	0.8363	0.8868	0.8363	0.8772
Polar	0.9021	0.8941	0.8957	0.6100	0.7921	0.9053	0.7921	0.8957
	k-NN							
Cart.	0.8949	0.8981	0.8828	0.7014	0.8475	0.8989	0.8475	0.8828
Polar	0.9133	0.9077	0.9101	0.6172	0.8226	0.9109	0.8226	0.9101

From TABLE III, it can be seen that the performance under Cartesian coordinate is significantly different to the one under polar coordinate. For $\delta_{Max}, \delta_{Min}$ and δ_{Pc2}, the recognition rates under Cartesian coordinate are much better than the ones under polar coordinate, while the classification accuracies under Cartesian coordinate are better than the one under polar coordinate for $\delta_{Arc}, \delta_{Pj}, \delta_{BC}, \delta_{PC1}$, and δ_{FS}. Moreover, the arc length, the Projection distance, the first version of Procrustes distance, and the Fubini-Study distance perform much better than the remaining metrics, as conforming to the preceding experiments. The best accuracy, 0.9133, is obtained by the arc length (geodesic distance) under the polar coordinate.

C. Quantitative Comparison with the Conventional Methods

In the previous subsections, we evaluate the effect of related factors on performance, including the distance metric

and signal decomposition. The subsequent experiments contribute to the quantitative comparison with the state-of-the-art algorithms. The downsampling ratio is set as $\rho = \frac{1}{16}$, corresponding to Grassmann manifold $\mathcal{G}(256, 9)$. The methods to be studied, including the proposed method and the reference algorithms are summarized in TABLE IV.

1) EOC Difference on Configuration: At present, more attentions are paid to target recognition under extended operating conditions, *i.e.*, the algorithm is trained under a standard experimental setup, while tested under another nonliteral one. In this subsection, we first evaluate target recognition under EOC difference on configuration.

Under the realistic battle-fields, there may be many physical target configurations that can be categorized as a single target type of military significance. In this context, the configuration refers to physical difference and small structural modifications. TABLE V lists some exemplars for target configuration.

Following the previous experiments, four targets, BMP2, T72, BTR60, and T62 are employed, among which BMP2 and T62 have three configuration variants denoted by series number, as shown in TABLE I. The standards, SN_9563 for BMP2 and SN_132 for T72 collected at $17°$ depression are used to train the algorithms, while the remaining (SN_9566, SN_c21 for BMP2, SN_812, SN_s7 for T72) taken at $15°$ depression are used for testing, *i.e.*, the configurations used for testing are not contained in the training set. The experimental results are tabulated in TABLE VI.

From TABLE VI, it is able to come the conclusions as follows.

i) The performance obtained by the algorithms whose input derives from the monogenic signal is much better than the methods whose input results from the original intensity values. The recognition rates for MSRC, SRC$^{\mathcal{R}}$, TJSR, and kNN$^{\mathcal{G}}$ are 0.8748, 0.9117, 0.9117, and 0.9133, compared to 0.8443 for SVM, 0.8708 for SRC. Moreover, the accuracies for KLMC, KSR$^{\mathcal{G}_1}$, and KSR$^{\mathcal{G}_1}$ are 0.9238, 0.9270, and 0.9390, compared to 0.9165 for KLRC, 0.8796 for KSVM, and 0.9045 for KSRC. The experimental result demonstrate that the monogenic signal representation could effectively characterize SAR image in comparison to the raw intensity information.

ii) The recognition rate for MSRC is 0.8748, compared to 0.9117 for SRC$^{\mathcal{R}}$, and 0.9133 for kNN$^{\mathcal{G}}$, 3.69 percent and 3.85 percent lower than the classification strategy on Riemannian manifolds. This is due mainly to a novel feature representation on Grassmannian manifolds for target recognition.

iii) The classification on Grassmann manifolds performs much better than the one on Riemannian manifolds. The accuracy for KSR$^{\mathcal{G}_1}$ and KSR$^{\mathcal{G}_2}$ are 0.9270 and 0.9390, compared to 0.9238 for KLMC. Though kNN$^{\mathcal{G}}$ achieves the performance (0.9133) similarly to SRC$^{\mathcal{R}}$ (0.9117), it should be noted that SRC$^{\mathcal{R}}$ utilizes sparse signal modeling technique to produce the inference, while the proposed method only employs a simple nearest neighbor classifier.

iV) The performance of classification in RKHS consistently exceeds the one of classification in the spatial domain. The recognition rate for KSR$^{\mathcal{G}_2}$ is 0.9390, compared to 0.9133 for kNN$^{\mathcal{G}}$, while the accuracy for KLMC is 0.9238, compared to 0.9117 for SRC$^{\mathcal{R}}$. The similar results can be obtained by comparing SVM, SRC with KSVM KSRC. This is because the original dataset is not linearly separable, and it can be circumvented by embedding it into an implicit Hilbert space. Then, the linear separability between different target classes can be enhanced.

2) Robustness towards Noise Corruption: For this experiment, we evaluate the recognition performance under random noise corruption. Three distance metrics, the arc length distance (geodesic distance), the Projection distance, and the first version of Procrustes distance are used to measure the difference between a pair of points. The corresponding classification methods are noted by kNN$^{\mathcal{G}_1}$, kNN$^{\mathcal{G}_2}$, and kNN$^{\mathcal{G}_3}$, respectively. The experimental setup is similar to the one in Section IV-C1.

To simulate the noise, we corrupt a percentage of randomly chosen pixels from the image available for testing, replacing their pixel values (intensity) with independent and identically distributed samples drawn from a uniform distribution $\mathcal{U}[0, u_{max}]$, where u_{max} is the largest possible pixel value. The sites whose pixel values are replaced are randomly chosen from the whole image. To demonstrate noise corruption, a series of exemplar images are given in Fig. 5, where 5%, 10%, 15%, 20%, 25%, and 30% of pixels of the original image are randomly corrupted.
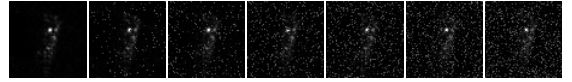

Fig. 5. Illustration of noise corruption. The set of images demonstrate that 0%, 5%, 10%, 15%, 20%, 25%, and 30% of pixels are corrupted.

TABLE VII provides the recognition accuracies obtained by various algorithms across the level of noise corruption, where the robust version of SRC, which solves the extended ℓ_1-norm minimization problem [31], is utilized to make the comparison. Obviously, the result is self-evident. It is able to see that the proposed method dramatically outperforms the competitors. From 0 percent upto 30 percent corruption, the accuracy for kNN$^{\mathcal{G}_1}$ slightly declines from 0.8949 to 0.8732. Similarly, the recognition rate for kNN$^{\mathcal{G}_2}$ drops from 0.8981 to 0.8724, and the accuracy for kNN$^{\mathcal{G}_3}$ drops from 0.8989 to 0.8756. Three Grassmann distance metrics have witnessed only a slight drop of 2.17%, 2.25%, and 2.33% when the level of corruption increased from 0% to 30%. On the contrary, the reference methods generate a drastic drop for recognition accuracy. The recognition rate for MSRC declines from 0.8748 to 0.5393, while the accuracy for SRC drops from 0.8708 to 0.4655. At 25 percent corruption, none of the competitors achieves higher than 0.7030 recognition rate, while the proposed algorithm achieve 0.8740, 0.8796, 0.8780 accuracies. Even at 30 percent corruption, the recognition rate for classification on Grassmann manifolds is still 0.8732, 0.8724, 0.8756, nearly 35% better than the nearest competitor. The experimental results demonstrate that the proposed classification technique is applicable to noise corruption environment, and hence could be applied to realistic battle-fields.

3) EOC Difference on Depression: To the end, we evaluate target recognition under EOC difference on depression. Three

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

Abbre.	Full name (description)	*Ref.*	Input	Hilbert (Y/N)	Riemannian (Y/N)	Grassmann (Y/N)
k-NN	k-Nearest-Neighbor classifier	[48]	Intensity values	N	N	N
SVM	Linear support vector machine learning	[49]	Intensity values	N	N	N
SRC	Sparse representation-based classification	[31]	Intensity values	N	N	N
MSRC	Sparse representation of monogenic signal	[17]	Augmented monogenic feature vector	N	N	N
SRC$^{\mathcal{R}}$	Sparse representation on Riemannian manifold	[19]	Monogenic covariance descriptor	N	Y	N
TJSR	Joint Sparse Representation of monogenic signal	[18]	Monogenic components	N	N	N
kNN$^{\mathcal{G}}$	kNN on Grassmann manifold	*Alg.* 1	The set of monogenic components	N	N	Y
KSVM	Kernel support vector machine	[50]	Intensity values	Y (**RBF**)	N	N
KSRC	Kernel sparse representation-based classifier	[45], [51]	Intensity values	Y (**RBF**)	N	N
KLRC	Kernel linear representation-based classifier	[52]	Intensity values	Y (**RBF**)	N	N
KLMC	Kernel linear coding of monogenic signal	[19]	Monogenic covariance descriptor	Y (**RBF**)	Y	N
KSR$^{\mathcal{G}_1}$	Sparse representation on Grassmann manifold	*Alg.* 2	The set of monogenic components	Y (κ_{Pj})	N	Y
KSR$^{\mathcal{G}_2}$	Sparse representation on Grassmann manifold	*Alg.* 2	The set of monogenic components	Y (κ_{BC})	N	Y

TABLE V. EXEMPLARS FOR CONFIGURATION

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

TABLE VI. THE RECOGNITION RATES OBTAINED BY VARIOUS METHODS UNDER EOC DIFFERENCE OF CONFIGURATION.

Spatial	SVM	SRC	MSRC	SRC$^{\mathcal{R}}$	TJSR	kNN$^{\mathcal{G}}$
	0.8443	0.8708	0.8748	0.9117	0.9117	0.9133
Hilbert	KLRC	KSVM	KSRC	KLMC	KSR$^{\mathcal{G}_1}$	KSR$^{\mathcal{G}_2}$
	0.9165	0.8796	0.9045	0.9238	0.9270	0.9390

TABLE VII. THE RECOGNITION RATES OBTAINED BY VARIOUS METHODS UNDER RANDOM NOISE CORRUPTION.

	SVM	SRC	MSRC	SRC$^{\mathcal{R}}$	TJSR	kNN$^{\mathcal{G}_1}$	kNN$^{\mathcal{G}_2}$	kNN$^{\mathcal{G}_3}$
0%	0.8443	0.8708	0.8748	0.9117	0.9117	0.8949	0.8981	0.8989
5%	0.7833	0.8587	0.8515	0.7873	0.8427	0.8796	0.8852	0.8836
10%	0.6404	0.7945	0.8090	0.7408	0.8041	0.8989	0.8989	0.8997
15%	0.4935	0.6653	0.7504	0.7022	0.7480	0.8788	0.8820	0.8812
20%	0.4606	0.5618	0.6661	0.6846	0.7415	0.8836	0.8812	0.8860
25%	0.4510	0.5016	0.5939	0.6100	0.7030	0.8740	0.8796	0.8780
30%	0.4454	0.4655	0.5393	0.5923	0.6821	0.8732	0.8724	0.8756
Drop	40.53%	33.55%	34.75%	31.94%	22.96%	2.17%	2.25%	2.33%

TABLE VIII. THE NUMBER OF ASPECT VIEW IMAGES AVAILABLE FOR 3 TARGETS AT VARIOUS DEPRESSIONS.

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

vehicle targets, 2S1, BRDM_2, ZSU_23/4 are employed, two of which, BRDM_2 and ZSU_23/4 have the articulated variants. The articulated refers to a physical change, *e.g.*, open hatch, rotated gun turret. An example of the articulated variant is shown in Fig. 6.
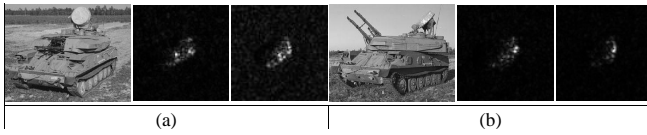

Fig. 6. Illustration of the articulated; (a) and (b) draw SAR images of ZSU_23/4 with turret straight and articulated taken at 30° and 45° depressions.

As mentioned above, images in MSTAR database are collected at various depression angles. To evaluate the recognition performance under different depressions, images taken at an operating condition of 17° depression are used to train the algorithms, while the ones captured at an operating condition of 30° and 45° are used for testing. The number of aspect view images available for three targets at different depressions are shown in TABLE VIII, where the entries in parentheses denote the articulated variants. The experimental results are tabulated in TABLE IX.

The first experimental scenario (shown in the first row of TABLE IX) handles a medium EOC difference on depression. It trains the algorithms at an operating condition of a 17° depression, and tests them at an operating condition of a 30° depression, *i.e.*, a change of 13° from 17° to 30° exists. As can be seen, all of the methods performs perfectly. None of the methods achieves lower than 0.9 recognition rate. KSRC

even produces the recognition rate of 0.9569. The recognition rate for the proposed method is 0.9300, 0.9354, and 0.9381, slightly lower than the previous competitors, *e.g.*, 0.9515 for SRC$^{\mathcal{R}}$, 0.9524 for TJSR.

The second experimental scenario (demonstrated in the second row of TABLE IX) deals with a major EOC difference on depression. The previous methods are further tested at an operating condition of a 45° depression, *i.e.*, a drastic change of 28° from 17° to 45° exists. Obviously, the performance sharply deteriorates. The recognition rate for KSVM even achieves a drop of 41.82 percent from 0.9443 to 0.5261. Similarly, the recognition rate for SRC and KSRC also obtain a drop of 41.25 percent and 40.15 percent. On the contrary, the proposed classification on Grassmann manifold produces a slight drop of 24.01 percent, 21.58 percent, and 20.99 for recognition accuracy. Moreover, it is significantly better than the former competitors. The recognition rate for KSR$^{\mathcal{G}_2}$ is 0.7282, compared to 0.6787 for KLMC, 0.5444 for KSRC, 0.5261 for KSVM, 4.95 percent, 18.38 percent, and 20.21 percent better than the reference methods. The results fully prove that the proposed classification on Grassmann manifolds is robust toward depression variations.

TABLE IX.   THE RECOGNITION RATES OBTAINED BY VARIOUS
METHODS UNDER EOC DIFFERENCE ON DEPRESSION.

	SRC	MSRC	SRC$^{\mathcal{R}}$	TJSR	kNN$^{\mathcal{G}}$	KSVM	KSRC	KLMC	KSR$^{\mathcal{G}_1}$	KSR$^{\mathcal{G}_2}$
$30°$	0.9488	0.9497	0.9515	0.9524	0.9300	0.9443	0.9569	0.9443	0.9354	0.9381
$45°$	0.5366	0.6263	0.6359	0.7073	0.6899	0.5261	0.5444	0.6787	0.7195	0.7282

V. CONCLUSION

In this paper, a novel classification technique via sparse representation on Grassmann manifolds is presented for target recognition in SAR images. The proposed method exploits the monogenic signal to characterize the broad spectral information yet maximal spatial localization of SAR image. Different from the preceding works, where a simple feature descriptor is produced in an Euclidean space, the set of components resulting from the multiscale monogenic signal are viewed as points on Grassmann manifold. Though it is incapable to obtain the specific value of the point, we could quantitatively compare a pair of points on Grassmann manifold via Grassmann distance metrics, *e.g.*, the arc length distance, the Projection distance, the Procrustes distance, *etc*. For any query pattern, it is capable to produce the inference by a simple nearest neighbor classifier. On the other hand, this paper embeds Grassmann manifold into an implicit Hilbert space, where the query pattern can be expressed as a linear combination of a dictionary, whose elements compose of the training patterns. Following the though of sparse signal modeling, we produce the parsimonious representation by limiting the feasible set with a sparsity constraint. The optimal representation can be obtained by solve ℓ_1-norm minimization optimization problem. The inference is reached by evaluating which class of pattern could recover the query one as accurately as possible. The feasibility of the proposed technique has been tested by some comparative experiments on MSTAR SAR database, a standard testbed for SAR image interpretation. From the experimental results, it is able to come the conclusions as follows:

- The monogenic signal representation could capture the discriminative characteristics of SAR image much more effectively in comparison to the original intensity value.
- The classification on manifolds could exploit the nonlinear geometric structure in the dataset, and hence outperform the classification in the Euclidean space.
- The performance of classification on Grassmann manifold always exceeds the one of classification on the conventional Riemannian manifold.
- By embedding Grassmann manifold into a Hilbert space with a nonlinear mapping, the proposed method could handle the dataset whose classes are not linearly separable.
- The proposed method is much more robust under extended operating conditions, *e.g.*, different configurations, different depressions, and noise corruption, in comparison to the competitors.

The great performance produced by classification on Grassmann manifold is actually the direct result of coupling the monogenic signal representation and sparse signal modeling. Our dictionary in sparse signal modeling composes of the training patterns themselves. Thus, an interesting question for our future work is whether the dictionary learning skills in machine learning could be applied to the proposed classification framework.

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] D. Dudgeon and R. Lacoss, "An overview of automatic target recognition," *Lincoln Lab J.*, vol. 6, no. 1, pp. 3–10, 1993.

[3] 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.

[4] D. Kreithen, S. Halversen, and G. Owirka, "Discriminating targets from clutter," *Lincoln Lab J.*, vol. 6, no. 1, pp. 25–52, 1993.

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

[6] Stein and Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton, New Jersey: Princeton University Press, 1971.

[7] M. Felsberg and G. Sommer, "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.

[8] 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.

[9] L. Wietzke, O. Fleischmann, and G. Sommer, "2D image analysis by generalized Hilbert transforms in conformal space," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. Part II, LNCS 5303, 2008, pp. 638–649.

[10] A. Belaid, D. Boukerroui, Y. Maingourd, and J.-F. Lerallut, "Phase-based level set segmentation of ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 138–147, Jan. 2011.

[11] G. Demarcq, L. Mascarilla, M. Berthier, and P. Courtellemont, "The color monogenic signal: Application to color edge detection and color optical flow," *J. Math. Imag. Vis.*, vol. 40, pp. 269–284, 2011.

[12] A. Depeursinge and A. Foncubierta-Rodriguez, "Rotation-covariant texture learning using steerable Riesz wavelets," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 898–908, Feb. 2014.

[13] M. Unser and D. Ville, "Wavelet steerability and the higher-order Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 636–649, Mar. 2010.

[14] R. Soulard, 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.

[15] 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.

[16] X. Huang, G. Zhao, W. Zheng, and M. Pietikinen, "Spatiotemporal local monogenic binary patterns for facial expression recognition," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 476–480, May 2012.

[17] 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.

[18] G. Dong, G. Kuang, L. Zhao, and L. Jun, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Minor revision 2015.

[19] 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.

[20] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithm with orthogonality constraints," *SIAM Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, Feb. 1999.

[21] J. Jost, *Riemannian Geometry and Geometric Analysis*, 3rd ed. NY: Springer, 2002.

[22] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1005–1018, Jun. 2007.

[23] J. Hamm and D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proc. Int'l Conf. Mach. Learn.*, 2008, pp. 376–383.

[24] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2273–2286, Nov. 2011.

[25] M. Harandi, C. Sanderson, S. Shirazi, and B. Lovell, "Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2705–2712.

[26] M. Harandi, C. Sanderson, C. Shen, and B. Lovell, "Dictionary learning and sparse coding on Grassmann manifolds: An extrinsic solution," in *Proc. IEEE Int'l Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 3120–3127.

[27] G. Dong, G. Kuang, L. Zhao, L. Jun, and L. Min, "Joint sparse representation of monogenic components: with application to automatic target recognition in SAR imagery," in *Proc. IEEE Int'l Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 549–552.

[28] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. NY: Academic Press, 2002.

[29] S. Olhede and G. Metikas, "The monogenic wavelet transform," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3426–3441, Sep. 2009.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

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

[35] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[36] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos, "Tensor sparse coding for region covariances," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. Part IV, Sep. 2010, pp. 722–735.

[37] M. Zhao, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

[38] M. Harandi, C. Sanderson, R. Hartley, and B. Lovell, "Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. Part II, LNCS 7573, 2012, pp. 216–229.

[39] Y. Hu, A. Mian, and R. Owens, "Face recognition using sparse approximated nearest points between image sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1992–2004, Oct. 2012.

[40] S. Chen, C. Sanderson, M. Harandi, and B. Lovell, "Improved image set classification via joint sparse approximated nearest subspaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 452–459.

[41] K. Guo, P. Ishwar, and J. Konrad, "Action recognition from video using feature covariance matrices," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2479–2494, Jun. 2013.

[42] Y. Zhang, Z. Jiang, and L. Davis, "Discriminative tensor sparse coding for image classification," in *Proc. British Machine Vision Conf. (BMVC)*, 2013, pp. 1–11.

[43] B. Schlkopf and A. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press, 2001.

[44] S. Lyu, "Mercer kernels for object recognition with local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 223–229.

[45] 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.

[46] 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.

[47] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

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

[49] 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.

[50] N. Cristianini and J. Shawe-Taylor, *Kernel Methods for Pattern Analysis*. New York: Cambridge University Press, 2006.

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

[52] 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.
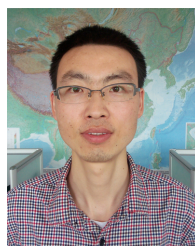
Ganggang Dong received the B. Eng. degree in UAV application engineering from Academy of Artillery, Hefei, China, in 2004 and the M.A.Eng. degree in information and communcation engineering from National University of Defense Technology, Changsha, China, in 2012. He is currently working toward the Ph.D. degree in 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 Central South University of Technology, Changsha, China, in 1998 and 1991, respectively, and the Ph.D. degree from National University of Defense Technology, Changsha, in 1995. He is currently a Professor and Director of the Remote Sensing Information Processing Laboratory in 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.