Classification via Sparse Representation of Steerable Wavelet Frames on Grassmann Manifold: Application to Target Recognition in SAR Image

Ganggang Dong, Gangyao Kuang, *Member, IEEE*, Na Wang, and Wei Wang

Abstract—Automatic target recognition has been studied widely over the years, yet it is still an open problem. The main obstacle consists in extended operating conditions, *e.g.*, depression angle change, configuration variation, articulation, occlusion. To deal with them, this paper proposes a new classification strategy. We develop a new representation model via the steerable wavelet frames. The proposed representation model is entirely viewed as an element on Grassmann manifolds. To achieve target classification, we embed Grassmann manifolds into an implicit Reproducing Kernel Hilbert Space (RKHS), where the kernel sparse learning can be applied. Specifically, the mappings of training sample in RKHS are concatenated to form an over-complete dictionary. It is then used to encode the counterpart of query as a linear combination of its atoms. By designed Grassmann kernel function, it is capable to obtain the sparse representation, from which the inference can be reached. The novelty of this paper comes from (i) the development of representation model by the set of directional components of Riesz transform; (ii) the quantitative measure of similarity for proposed representation model by Grassmann metric; (iii) the generation of global kernel function by Grassmann kernel. Extensive comparative studies are performed to demonstrate the advantage of proposed strategy.

Keywords—*SAR, Target Recognition, Riesz transform, Steerable Wavelet, Grassmann manifold, Sparse representation, Hilbert space.*

I. INTRODUCTION

AS an active sensor, synthetic aperture radar (**SAR**) could provide broad-area imaging at high resolution and acquire image during night as well as day. At present, it has been widely used, such as, monitoring the environment, recognizing and tracking ground vehicle, and mapping the Earth's resources. With the development of synthetic aperture technique, images collected from SAR sensor are too huge to be processed timely by photo analyst. This situation breeds an urgent need for automatic target recognition [1]. Though it has been studied over the decades, it is still an open problem. The main obstacle of target recognition lies in extended operating conditions, where a single operational parameter is significantly different between the images used for training and those for testing [2].

The specific challenge to be addressed is robust recognition of military target that can be collected at different pose and depression angle.

Early target recognition system relies on template matching [3]. The inference is reached by quantifying the similarity between the query and multiple templates generated by the training samples. It is ineffective under extended operating conditions, because slight changes in articulation or occlusion may result in significantly different scattering phenomenology. Hence it is difficult to quantify the similarity between the query and templates. Then another family of methods, correlation pattern recognition, was presented [4]. These methods utilize the global features (Fourier transform coefficients) that are optimized for standard operating conditions. The classification accuracy for extended operating conditions was reported to be limited. As widely reported, SAR image often results from both specular and diffuse reflections [5]. To take both types of scatterers into account, some parameterized statistical distribution models are used to represent SAR image amplitudes [6], [7]. Given SAR image of an unknown target and a specific statistical model conditioned on target class and pose, the task of target recognition and pose estimation can be achieved in terms of Bayesian hypothesis testing and minimum mean square error estimation. The log-likelihoods are obtained using the parameters estimated from the training samples. However, this family of methods usually suffer from the parameter estimation under limited samples. They easily fail when strong statistical relationship between the query and the observed does not exist. Again, some researchers have presented multiple discriminative features, and report their performance under training and testing scenarios [8], [9]. However, the relative performance of these features is difficult to be assessed based upon the empirical evaluation.

Recently, a general multi-dimensional analytic signal, the monogenic signal has been developed around the Riesz transform [10]. It possesses distinguishing property that the underlying feature extraction process is truly rotation-invariant. This also explains its success in many applications, *e.g.*, contour detection [11], local structure analysis [12], motion estimation [13], image registration [14], and pattern recognition [15], [16]. To consider the scale and directionality properties, some works have been done to transpose the monogenic signal to the wavelet domain. S. Olhede and G. Metrikas define the general vector-valued quaternion mother wavelet functions to combine the monogenic signal with the act of localization [17]. M. Unser *et al.* introduce the multi-resolution monogenic analysis

via Riesz-Laplace wavelet transform, where a complexified Riesz transform mapping a real-valued wavelet basis into a complex one are developed [18]. S. Held *et al.* present a new n-dimensional monogenic wavelet from the hyper-complex monogenic signal with the Riesz transform and isotropic wavelet frame [19]. M. Unser and D. Ville develop a general continuous-domain frameworks for steerable, reversible signal transformations [20]. R. Soulard *et al.* introduce a nontrivial extension of the monogenic framework to a non-marginal monogenic wavelet transform [21]. A. Depeursinge and A. Foncubierta-Rodriguez propose a rotation-covariant texture learning strategy using steerable Riesz wavelets [22]. These works prove that the steerable wavelet transform allows for translation, scaling, rotation, and hence provides great potential for model representation.

The majority of available techniques for pattern recognition assumes an underlying Euclidean space. Yet many advantages can be attained by considering these problems with a non-Euclidean geometry [23]. The relevant non-Euclidean space is the set of fixed dimensional subspaces, commonly called Grassmann manifold, a special type of Riemannian manifold with column orthogonalized. Over the years, many works are devoted to pattern recognition from exemplars that lie on certain manifolds. O. Tuzel *et al.* develop an approach for classifying points on a connected Riemannian manifold with the geometry of the non-Euclidean space [24]. P. Turaga *et al.* demonstrate that the commonly used parametric models for videos and image sets can be represented by the unified framework of Grassmann and Stiefel manifolds [25]. Y. Hu *et al.* propose a new approach for image set classification, where the sample of image set and its affine hull model are by jointly considered into a unified framework [26]. These works provide a good deal of enlightenment for classifying a set of signals.

To solve the obstacle of target recognition, this paper presents a new classification strategy. The steerable pyramid and tight Riesz wavelet frames are considered into Grassmann manifolds. Specifically, a steerable wavelet pyramid frames derived from the multi-orientation and multi-scale Riesz transform [27], is utilized to produce a new representation model[1]. The proposed representation model allows for a wide range of rotation, scaling and translation. Since the proposed representation model composes of a set of directional wavelet coefficients, it could not be dealt with by the conventional learning skill in Euclidean space. To handle the difficult training regimes, this paper considers the proposed representation model in non-Euclidean space. It is entirely regarded as an element on Grassmann manifolds. The measure of similarity between two proposed representation models is realized by accumulating Grassmann metric over the wavelet scale. The next problem is how to pursue the learning of classification. Inspired by the preceding works [29], [30], this paper embeds Grassmann manifold into an implicit Reproducing Kernel Hilbert Space. The proposed representation model is cast into the Hilbert space with a nonlinear mapping. Then, various

kernel analysis skills can be applied. This paper utilizes kernel sparse learning model due to the good performance. First, the mappings of the training sample in RKHS are concatenated to build a redundant dictionary to encode the counterpart of query as a linear combination of themselves. The presented regression model is solved with a designed Grassmann kernel. The decision is reached according to the characteristics of sparse representation on reconstruction, *i.e.*, evaluating which class of samples could produce the minimal reconstruction error.

Contributions. This paper proposes a classification strategy for target recognition in SAR image. The set of multi-orientation and multi-scale wavelet coefficients resulting from the high-order Riesz transform are used to generate a new representation model. The proposed representation model is entirely viewed as an element on Grassmann manifold. The measure of similarity for proposed representation model is implemented by aggregating Grassmann metric across the wavelet scale. To the end, we embed Grassmann manifold into an abstract Hilbert space, where kernel sparse learning model can be built to draw the inference. The main contribution of our proposed strategy can be summarized as follows.

- *The development of representation model via steerable wavelet frames.* The proposed representation model allows for invariance towards translations and rotations, and hence provides great potential for target recognition.
- *The definition of the measure of similarity for proposed representation model with Grassmann metric.*
- *The designed kernel function for sparse learning model.*
- *The utilization of kernel sparse representation on Grassmann manifolds.* The proposed representation model is cast into RKHS, where kernel analysis skills is applied.

The reminder of this paper is organized as follows. The next section provide the related works, the steerable wavelet frames, followed by the proposed strategy, kernel sparse representation on Grassmann manifolds. The effectiveness of proposed strategy has been evaluated in Section III, followed by the conclusion of this paper in Section IV.

II. THE PROPOSED STRATEGY

Radar targets, when observed from a given orientation, often produce a radar cross section (RCS) [31]. Different from the common optical sensors, SAR RCS often composes of both specular scatterers and and diffuse ones. The collection of scatterers contributing to a resolution cell in SAR image are usually dominated by a specular component resulting in glint at the location of the image. Thus, SAR images are typically non-literal, which results in the mutable scattering phenomenology even with small changes in pose, depression, and configuration. Due to the special image formation process, it is difficult to achieve satisfied performance using the conventional representation model. To deal with these issues, this paper develops a novel classification strategy. We represent the scattering phenomenology by the tight wavelet frame of Riesz transform. The directional coefficients of steerable wavelet frames are entirely viewed as an element on Grassmann manifold. It is

---

[1]Steerability refers to the property that the underlying wavelets can be rotated to any orientation by forming suitable linear combinations of a primary set of equiangular directional wavelet components [28].
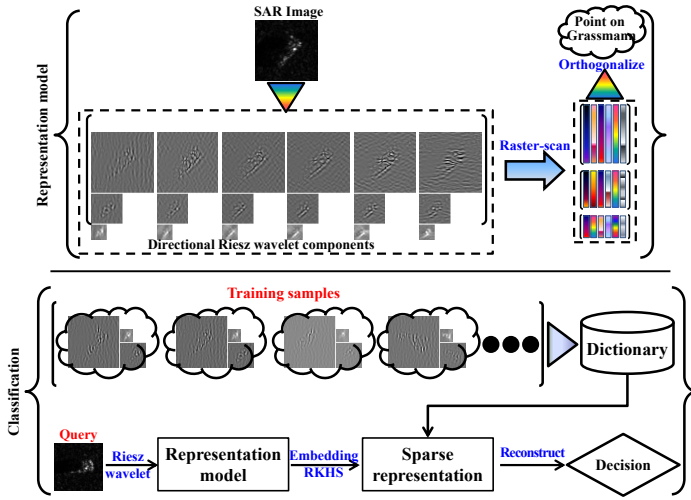
Fig. 1. The complete pipeline of proposed strategy. The proposed strategy composes of two separate phases: representation and classification. The first phase devotes to generating a representation model by performing the Nth-order and Jth-scale Riesz wavelet transform. The resulting model is entirely viewed on Grassmann manifolds. The second phase devotes to classification learning. The proposed representation model is cast into the Hilbert space. The mappings of training sample are concatenated to form a redundant dictionary to encode the counterpart of query as a linear combination of themselves. Grassmann kernel is used to generate the sparse representation, from which the inference can be reached.

then cast into the RKHS, where kernel analysis schemes can be applied. The pipeline for proposed strategy is shown in Fig. 1.

We first introduce the steerable wavelet frames, followed by the proposed representation model. The classification strategy, kernel sparse representation on Grassmann manifolds, is presented in the end.

A. Steerable Wavelet Frames

1) The Riesz Transform: The Riesz transform is an extension of Hilbert transform. For any signal $f(\mathbf{x}), \mathbf{x} = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n$, the Riesz transform is defined as $f_{\mathcal{R}_r}(\mathbf{x}) = \lim_{\epsilon \to \infty} c_n \int_{|y| < \epsilon} \frac{y_r}{|\mathbf{y}|^{n+1}} f(\mathbf{x} - \mathbf{y}) d\mathbf{y}$, where $c_n = \frac{\Gamma[(n+1)/2]}{\pi^{(n+1)/2}}$. The convolution kernel of Riesz transform is

$$\mathcal{R}\{f\}(\mathbf{x}) = \begin{pmatrix} \mathcal{R}_1\{f\}(\mathbf{x}) \\ \mathcal{R}_2\{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}_n\{f\}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} (h_1 * f)(\mathbf{x}) \\ (h_2 * f)(\mathbf{x}) \\ \vdots \\ (h_n * f)(\mathbf{x}) \end{pmatrix} \quad (1)$$

where $h_r = c_n \frac{x_r}{|\mathbf{x}|^{n+1}}$ is the rth sub-transform. The transfer function of Riesz transform can be expressed as

$$H(\mathbf{u}) = \left[-j\frac{u_1}{|\mathbf{u}|}, -j\frac{u_2}{|\mathbf{u}|}, ..., -j\frac{u_n}{|\mathbf{u}|} \right]^T \quad (2)$$

where $H_r = -i\frac{u_r}{|\mathbf{u}|}$ is the frequency response of the rth sub-transform. The Riesz transform keeps unit length in any direction, and hence has isotropic energy distribution. Moreover, it has many unique properties.

- Property 1: *translation-invariant*

$$\forall \mathbf{x}_0 \in \mathbb{R}^n, \mathcal{R}\{f(\cdot - \mathbf{x}_0)\}(\mathbf{x}) = \mathcal{R}\{f(\cdot)\}(\mathbf{x} - \mathbf{x}_0)$$

- Property 2: *scale-invariant*

$$\forall a \in \mathbb{R}^+, \mathcal{R}\left\{f\left(\frac{\cdot}{a}\right)\right\}(\mathbf{x}) = \mathcal{R}\{f(\cdot)\}\left(\frac{\mathbf{x}}{a}\right)$$

- Property 3: *steerability*

$$h_r(\mathbf{R_v}\mathbf{x}) = \langle \mathbf{v}, \mathbf{R}\delta(\mathbf{x}) \rangle = \sum_{l=1}^n v_l h_l(\mathbf{x})$$

where $\mathbf{R_v}$ denote a $n \times n$ rotation matrices whose r^{th} row is the unit vector $\mathbf{v} = [v_1, v_2, ..., v_n]$.

- Property 4: *inner-product preservation*

$$\|\mathcal{R}f\|_{L_2^n}^2 = \sum_{r=1}^n \|\mathcal{R}_r f\|_{L_2}^2 = \|f\|_{L_2}^2$$

2) Higher-Order Riesz Transform: To further exploit the angular selectivity property, M. Unser *et al.* develop the higher-order Riesz transform [20], [28],

$$\mathcal{R}^{(N)}\{f\}(\mathbf{x}) = \begin{pmatrix} \mathcal{R}^{(N,0,...,0)}\{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(N_1,N_2,...,N_n)}\{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(0,0,...,N)}\{f\}(\mathbf{x}) \end{pmatrix} \quad (3)$$

where $\mathcal{R}^N = \sqrt{\frac{|N|!}{N!}} \mathcal{R}_1^{N_1} \mathcal{R}_2^{N_2} \cdots \mathcal{R}_n^{N_n}$ is the multi-index, and $\frac{|N|!}{N!} = \binom{N}{N_1,...,N_2} = \frac{N!}{N_1!N_2!\cdots N_n!}$. The corresponding transfer function is $\mathcal{R}^N\{F\}(\mathbf{u}) = \sqrt{\frac{|N|!}{N!}} \frac{(-j\mathbf{u})^N}{\|\mathbf{u}\|^{|N|}} F(\mathbf{u})$.

3) Steerable Wavelet Frames: Thanks to shift- and scale-invariance, the Nth-order Riesz transform could convert any wavelet frame into another one [20]. For a signal with finite energy, we could produce the following wavelet decomposition: $\forall f \in L_2(\mathbb{R}^n), f(\mathbf{x}) = \sum_{r \in \mathbb{Z}} \sum_{l \in \mathbb{Z}^n} \{\langle f, \psi_{r,l} \rangle_{L_2} \psi_{r,l}(\mathbf{x})\}$. The wavelets at the rth scale are the dilated version of the ones at the finest scale: $\psi_{r,l}(\mathbf{x}) = 2^{-r(n/2)} \psi_{0,l}(\mathbf{x}/2^r)$. Then, a general wavelet frames can be obtained,

$$\psi_{r,l}^N = \mathcal{R}^N \psi_{r,l} = \sqrt{\frac{N!}{N_1!...N_n!}} \mathcal{R}_1^{N_1} \cdots \mathcal{R}_n^{N_n} \psi_{r,l} \quad (4)$$

The steerable wavelet frames inherits the fundamental properties of primary operator, and hence provides great potential for signal analysis.

B. A New Representation Model

Thanks to steerable Riesz wavelet, the property of scale invariance, translation-invariance, and rotation-invariance can be achieved. These peculiar properties motivates us to develop a new representation model, with which the extended operating conditions, for example, pose and depression change, structural modification, articulation, and occlusion can be handled.

For a 2-D signal (or image) f with $h \times w$ pixels in size, we first produce the Nth-order Riesz transform. A set of the 5th-order Riesz transform are pictorially shown in Fig. 2. Obviously, the higher-order Riesz transform has good angular selectivity. Since the higher-order Riesz transform could extract the Nth-order smooth derivatives, it commutes with translation and scaling operators. Another more remarkable property is that it commutes with spatial rotation.
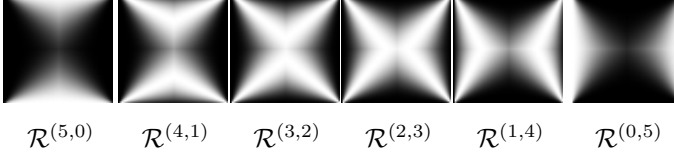

$$\mathcal{R}^{(5,0)} \quad \mathcal{R}^{(4,1)} \quad \mathcal{R}^{(3,2)} \quad \mathcal{R}^{(2,3)} \quad \mathcal{R}^{(1,4)} \quad \mathcal{R}^{(0,5)}$$

Fig. 2. The frequency response of the 5th-order Riesz transform.

Second, the higher-order Riesz transform could map any frame of $L_2(\mathbb{R}^n)$ into another one, it is therefore capable to produce the tight wavelet frames. The wavelet frames are a family of functions $\{\psi_{r,l}\}_{r \in \mathbb{Z}^n, l \in \mathbb{Z}}$ indexed by location and scale (or resolution) pair (r, l). The basis function at the rth-resolution is the rescaled and dilated version of the one at the finest resolution. Thus, the higher-order Riesz transform can be used to produce a family of self-reversible and tight pyramids, *i.e.*, the steerable wavelet frames. An instance of the $J(2)^{nd}$-scale and $N(5)^{th}$-order steerable wavelet frame is given in Fig. 3. It achieves a better decomposition of the image in terms of directional components, and hence are steerable in arbitrary directions.

To exploit the invariant properties, this paper develops a new representation model with the directional components of steerable wavelet frames. The proposed representation model composes of a set of directional pyramid wavelet coefficients. Then, the task to infer the membership of the query is converted to classify a set of wavelet coefficients to one of the training class. We expect that robust performance can be achieved by a set of pyramid coefficients rather than a single signal. The implementation of target classification on the proposed representation model refers to set-to-set match problem. Hence the conventional learning methods could not be applied any more. Recent studies prove that great advantages can be achieved by considering pattern classification with a non-Euclidean geometry [23]–[25], [30]. This inspires us to deal with the proposed representation model in a non-Euclidean space, *i.e.*, Grassmann manifolds.

Manifold is a topological space which is locally similar to Euclidean space. The space of $d \times m$-dimensional matrices is a special type of Riemann manifold, Stiefel manifold ($\mathcal{S}(m, d)$), if it satisfies $\mathcal{S}(m, d) \triangleq \mathbf{Y} \in \mathbb{R}^{d \times m} : \mathbf{Y}^T \mathbf{Y} = \mathbf{I}_m$, where $\mathbf{I}_m$ is $m \times m$-dimensional identity matrix. Grassmann manifold $\mathcal{G}(m, d)$ is a quotient manifold. It fulfills the following relation: $\mathbf{Y}_1 \sim \mathbf{Y}_2$, *if and only if* $Span(\mathbf{Y}_1) = Span(\mathbf{Y}_2)$, where $Span(\mathbf{Y})$ is the subspace spanned by the columns of $\mathbf{Y}$. The similarity between two elements on manifold is measured by geodesic distance. A direct approach to compute geodesic distance is the principal angles. Given two elements
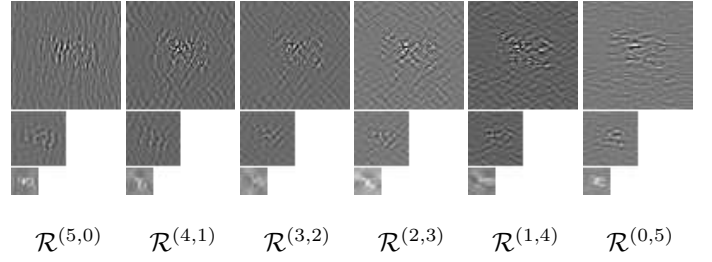

$$\mathcal{R}^{(5,0)} \quad \mathcal{R}^{(4,1)} \quad \mathcal{R}^{(3,2)} \quad \mathcal{R}^{(2,3)} \quad \mathcal{R}^{(1,4)} \quad \mathcal{R}^{(0,5)}$$

Fig. 3. Illustration of the 2th-scale and 5th-order steerable wavelet frames. In each sub-figure, the first image show the finest resolution component, while the second and third ones demonstrate the first- and second-scale ones.

$\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{d \times m}$, their principal angle is defined as

$$\cos \theta_k = \max_{u_k \in Span(\mathbf{Y}_1)} \max_{v_k \in Span(\mathbf{Y}_2)} u_k^T v_k$$
$$subject\ to \left(\begin{array}{c} u_k^T u_k = 1, v_k^T v_k = 1 \\ u_k^T u_j = 0, v_k^T v_j = 0, j = 1, ..., k-1 \end{array} \right) \quad (5)$$

The first principal angle θ_1 is defined as the smallest angle between a pair of unit vector, each of which result from $Span(\mathbf{Y}_1)$ and $Span(\mathbf{Y}_2)$. The kth principal angle can be obtained recursively. Actually, there is no need to solve (5) directly. Principal angle can be also obtained by singular value decomposition $\mathbf{Y}_1^T \mathbf{Y}_2 = \mathbf{U S V}$, where $\mathbf{U}, \mathbf{V}$ are unitary matrices, and $\mathbf{S} = diag(\cos(\theta_1), ..., \cos(\theta_m))$ is the diagonal matrix. For two elements on Grassmann manifolds, their similarity can be measured by the principal angles. The popularly used Grassmann metric includes:

- *Arc length*: $\delta_{Arc}^2(\mathbf{Y}_1, \mathbf{Y}_2) = \sum_r \theta_r^2 = \|\Theta\|_2^2$
- *Projection Metric*: $\delta_{Pj}^2(\mathbf{Y}_1, \mathbf{Y}_2) = \sum_{r=1}^{m} \sin^2(\theta_r)$
- *Binet-Cauchy Metric*: $\delta_{B}^2(\mathbf{Y}_1, \mathbf{Y}_2) = 1 - \prod_{r=1}^{m} \cos^2(\theta_r)$

Given 2-D signal $f \in \mathbb{R}^{h \times w}$, its Nth-order and J^{th}-scale steerable wavelet frames can be expressed

$$\left\{ \psi_l^{(N,0)}, \psi_l^{(N-1,1)}, \ldots, \psi_l^{(0,N)} \right\}, \psi_l^{(\cdot, \cdot)} \in \mathbb{R}^{h/2^l \times w/2^l}$$

$l = 0, ..., J - 1$. We reshape each wavelet component to be a d_l-dimensional ($d_l = (h/2^l)(w/2^l)$) single vector by raster-scanning the columns. The wavelet components at a certain scale are concatenated to form a matrix $\Psi_l \in \mathbb{R}^{d_l \times m}$, where $m = N + 1$. The orthogonal condition $\Psi_l^T \Psi_l = \mathbf{I}$ can be easily realized with the mathematical operations, *e.g.*, Schmidt orthogonalization. The resulting matrix fulfills, $\forall f_1, f_2$

$$\Psi_l(f_1) \sim \Psi_l(f_2), if\ and\ only\ if, Span(\Psi_l(f_1)) = Span(\Psi_l(f_2)).$$

The proposed representation model can be then entirely viewed as an element on Grassmann manifolds. The similarity between two representation models at a certain scale can be measured by Grassmann metric. The global similarity can be obtained by accumulating Grassmann metric across the wavelet scale. This measure of similarity can be then conjunction with a simple $\mathbf{k}^{th}$-nearest-neighbor classifier to achieve target classification.

Given n training samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ from L distinct classes, their representation models can be produced by building the Nth-order and J^{th}-scale steerable wavelet frames

$$\{\Psi_{N,J}(\mathbf{x}_1), \Psi_{N,J}(\mathbf{x}_2), ..., \Psi_{N,J}(\mathbf{x}_n)\}. \quad (6)$$
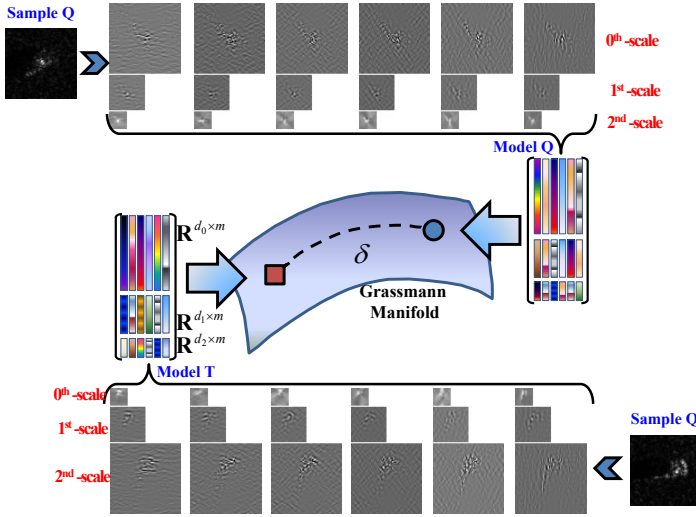
Fig. 4. The measure of similarity between a pair of proposed representation models. The set of Riesz wavelet components are entirely viewed as an element on Grassmann manifolds. The global similarity are defined by accumulating Grassmann metric over the wavelet scale. d_0, d_1 and d_2 denote the dimensionality of components at the 0th-, 1st-, and 2nd-scale.

These representation models are entirely viewed as the elements on Grassmann manifolds. For a query sample $\mathbf{y}$, its representation model $\Psi_{N,J}(\mathbf{y})$ can be obtained similarly. To measure the similarity between two proposed representation models, this paper defines the global similarity by aggregating Grassmann metric over the wavelet scale,

$$Sim(\Phi_{N,J}(\mathbf{y}), \Phi_{N,J}(\mathbf{x}_k)) = \sum_{r=0}^{J-1} Sim_l(\Phi_r(\mathbf{y}), \Phi_r(\mathbf{x}_k)) \quad (7)$$

where $Sim_l(\Phi_r(\mathbf{y}), \Phi_r(\mathbf{x}_k))$ is the similarity between two representation models at the rth-scale. It can be implemented using the Grassmann metric, *e.g.*, arc length, projection metric. Since the addition of any valid metrics is a valid one, the measurement (7) is a valid metric. To demonstrate the global similarity, a conceptual example is shown in Fig. 4, where two SAR images are compared with in terms of their representation models.

With the measurement defined in (7), the similarity between the probe and the gallery can be measured, from which the decision can be reached. The complete procedure for kNN on Grassmann manifold is listed in Algorithm 1.

C. Sparse Representation on Grassmann Manifold

The proposed representation model composes of a set of wavelet coefficients, whose dimensions are different, it is therefore difficult to be applied into the learning framework in Euclidean space. Though the developed similarity metric can be conjunction with a nearest neighbor classifier to implement target classification, there are some severe restrictions. Moreover, its recognition performance is limited. To achieve robustness to difficult training regimes, this paper embeds Grassmann manifold into an abstract RKHS. Then various kernel analysis

---

Algorithm 1 kNN on the Grassmann Manifolds ($NN^{\mathcal{G}}$)

Input: $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$: a set of training samples for L classes; $\mathbf{y}$: a query sample;
Output: identification of $\mathbf{y}$.

1: Produce the representation model of the query by the Nth-order and J^{th}-scale Riesz wavelet $\Psi_{N,J}(\mathbf{y})$;
2: **for** $k \leftarrow 1, n$ **do**
3:     Generate the representation model of the kth training sample $\Psi_{N,J}(\mathbf{x}_k)$;
4:     **for** $r \leftarrow 0, J-1$ **do**
5:         Quantify the similarity between the query and the kth training sample with Grassmann metric;
6:         Aggregate the similarity over the wavelet scale:

$$SimV_k+ = Sim(\Psi_r(\mathbf{y}), \Psi_r(\mathbf{x}_k))$$

7:     **end for**
8: **end for**
9: Reach the inference: $Label\{\arg\min_k\{SimV\}\}$.

---

schemes can be applied. Inspired by the preceding works [32]–[34], we design kernel sparse signal modeling.

Given an embedded function $\phi(\cdot) : \mathcal{G} \mapsto \mathcal{H}$, by which Grassmann manifolds can be embedded into the Hilbert space, the proposed representation model at a certain scale Ψ_r can be cast into the Hilbert space $\phi(\Psi_r)$. For a pair of representation models, $\Psi_r(\mathbf{x}_k)$ and $\Psi_r(\mathbf{x}_l)$, we could define a Grassmann kernel function,

$$\kappa(\Psi_r(\mathbf{x}_k), \Psi_r(\mathbf{x}_l)) = \langle \phi(\Psi_r(\mathbf{x}_k)), \phi(\Psi_r(\mathbf{x}_l)) \rangle \quad (8)$$

Jihun Hamm and Daniel D. Lee prove that the projection kernel and Binet-Cauchy kernel are valid Grassmann kernel [29]. They are widely used in the preceding works.

- *Projection Kernel*: $\kappa_{Pj}(\mathbf{Y}_1, \mathbf{Y}_2) = \|\mathbf{Y}_1^T\mathbf{Y}_2\|_F^2$
- *Binet-Cauchy Kernel*: $\kappa_{BC}(\mathbf{Y}_1, \mathbf{Y}_2) = det(\mathbf{Y}_1^T\mathbf{Y}_2)^2$

Since the addition of any valid kernels yields a valid one, we could form a global kernel function by assembling the ones over the wavelet scale,

$$\kappa(\cdot, \cdot) = \sum_{r=0}^{J-1} \beta_r \kappa_r(\cdot, \cdot) \quad (9)$$

where $\kappa_r(\cdot, \cdot)$ is the kernel function at the rth-scale. β_r is the weight value for the rth-scale kernel function. This paper sets these weights to be unit. The details on the selection of kernel weight will be discussed in Section III. The experiments in Section III corroborate that the discrimination power can be enhanced by assembling the kernel function over the scale.

Given a set of training samples, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, we first generate their representation models by the Nth-order and J^{th}-scale steerable wavelet. The resulting representation models are cast into the Hilbert space by $\phi(\cdot)$,

$$\{\phi(\Psi_{N,J}(\mathbf{x}_1)), \phi(\Psi_{N,J}(\mathbf{x}_2)), ..., \phi(\Psi_{N,J}(\mathbf{x}_n))\}. \quad (10)$$

These mappings in the Hilbert space are concatenated to form a redundant dictionary $\mathcal{D}$. The resulting dictionary is then used

to encode the counterpart of query $\phi(\Psi_{N,J}(\mathbf{y}))$ as a linear combination of themselves,

$$\phi(\Psi_{N,J}(\mathbf{y})) = \phi(\Psi_{N,J}(\mathbf{x}_1))\alpha_1 + \phi(\Psi_{N,J}(\mathbf{x}_2))\alpha_2 \\ + \cdots + \phi(\Psi_{N,J}(\mathbf{x}_n))\alpha_n = \mathcal{D}\alpha \quad (11)$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n] \in \mathbb{R}^n$ is the representation vector. In (11), the recognition problem is cast as one of classifying among multiple linear regression, where the mappings of training sample, $\{\phi(\Psi_{N,J}(\mathbf{x}_r))\}_{r=1}^n$ play the role of regressors (or predictor), and the one of query $\phi(\Psi_{N,J}(\mathbf{y}))$ is the corresponding response. The next problem is how to seek the optimal regression coefficients $\hat{\alpha}$. For notation convenience, the mapping of query and training, $\phi(\Psi_{N,J}(\mathbf{y}))$ and $\phi(\Psi_{N,J}(\mathbf{x}_k))$, will be abbreviated as $\phi_\mathbf{y}$ and $\phi_{\mathbf{x}_k}$.

The membership of the query is unknown, hence it is represented with the entire training set, *i.e.*, the training sample of all classes. Since a query sample can be sufficiently represented using only the training samples from the same class [35], the entries of the coefficient vector α should be zero except those associated with the unknown class. This representation is naturally sparse if the number of classes and training samples is reasonably large. The popularly used approach to produce the sparse solution is to restrict the feasible set via sparsity constraint, for example, ℓ_1-norm minimization,

$$\min_\alpha \|\alpha\|_1 \text{ subject to } \|\phi_\mathbf{y} - \mathcal{D}\alpha\|_2^2 < \varepsilon \quad (12)$$

where ε is the allowed error tolerance. The objective function expresses the sparsity level, while the constraint term gives the reconstruction error (*i.e.*, the fidelity term). However, it is difficult to be solved directly, due to the absence of the embedded function $\phi(\cdot)$. To circumvent the problem, we first unfold the fidelity term (it is re-notated by $g(\alpha)$),

$$g(\alpha) = \|\phi_\mathbf{y} - \mathcal{D}\alpha\|_2^2 \\ = (\phi_\mathbf{y} - \mathcal{D}\alpha)^T(\phi_\mathbf{y} - \mathcal{D}\alpha) \quad (13) \\ = \phi_\mathbf{y}^T\phi_\mathbf{y} - 2\phi_\mathbf{y}^T\mathcal{D}\alpha + \alpha^T\mathcal{D}^T\mathcal{D}\alpha$$

Then, some items of the fidelity can be replaced using the kernel trick. By introducing the proposed Grassmann kernel (8), the fidelity term can be re-written as:

$$\phi_\mathbf{y}^T\phi_\mathbf{y} = \langle \phi_\mathbf{y}, \phi_\mathbf{y} \rangle = \kappa(\Psi_{N,J}(\mathbf{y}), \Psi_{N,J}(\mathbf{y}))$$

$$\phi_\mathbf{y}^T\mathcal{D} = \begin{bmatrix} \langle\phi_\mathbf{y}, \phi_{\mathbf{x}_1}\rangle \\ \langle\phi_\mathbf{y}, \phi_{\mathbf{x}_2}\rangle \\ \vdots \\ \langle\phi_\mathbf{y}, \phi_{\mathbf{x}_n}\rangle \end{bmatrix} = \begin{bmatrix} \kappa(\Psi_{N,J}(\mathbf{y}), \Psi_{N,J}(\mathbf{x}_1)) \\ \kappa(\Psi_{N,J}(\mathbf{y}), \Psi_{N,J}(\mathbf{x}_2)) \\ \vdots \\ \kappa(\Psi_{N,J}(\mathbf{y}), \Psi_{N,J}(\mathbf{x}_n)) \end{bmatrix}$$

and (14), where $\kappa(\Psi_{N,J}(\mathbf{x}_k), \Psi_{N,J}(\mathbf{x}_l))$ can be realized by assembling Grassmann kernel functions over the wavelet scale

$$\kappa(\Psi_{N,J}(\mathbf{x}_k), \Psi_{N,J}(\mathbf{x}_l)) = \sum_{r=0}^{J-1}\{\beta_r\kappa_r(\Psi_r(\mathbf{x}_k), \Psi_r(\mathbf{x}_l))\}$$

Given a Grassmann kernel, (12) can be then tractable accordingly. From (12) to (14), we found that the embedded function $\phi(\cdot)$ does not participate in the numerical implementation of sparse representation. It only plays a role of bridge that

links Grassmann manifold and Hilbert space. By integrating the fidelity into (12), the problem can be then solved using the common optimization skills, such as the feature-sign search algorithm [36]. Finally, the inference can be reached according to the characteristics of sparse representation on reconstruction, *i.e.*, evaluating which class of samples could result in the minimum reconstruction error,

$$\min_{k=1,...,L}\{\|\phi_\mathbf{y} - \mathcal{D}\delta_k(\hat{\alpha})\|_2^2\} \quad (15)$$

where $\delta(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^n$ preserves the entries associated with the kth training class and sets the remaining to be zero. The complete procedure of propose strategy, sparse representation on Grassmann manifold, is shown in Algorithm 2 ($\text{KSR}^\mathcal{G}$).

Algorithm 2 Sparse Representation on Grassmann Manifolds

Input: $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$: a set of training samples for L classes;
   $\mathbf{y}$: query sample;
   κ: Grassmann kernel;
   λ: the regularization parameter;
Output: identification of $\mathbf{y}$.
 1: Produce the representation model of query by the Nth-order and J^{th}-scale Riesz transform $\Psi_{N,J}(\mathbf{y})$;
 2: Generate the representation model of the training samples $\Psi_{N,J}(\mathbf{x}_1), \Psi_{N,J}(\mathbf{x}_2), \ldots, \Psi_{N,J}(\mathbf{x}_n)$;
 3: **for** $r \leftarrow 1, n$ **do**
 4:    Compute the similarity between the query and the rth training sample in terms of their representation models in RKHS, $\kappa(\Psi_{N,J}(\mathbf{y}), \Psi_{N,J}(\mathbf{x}_r)) \rightarrow [\phi_\mathbf{y}\mathcal{D}]_r$;
 5:    **for** $l \leftarrow 1, n$ **do**
 6:       Quantify the similarity between the rth and l^{th} training sample in terms of their representation models in RKHS, $\kappa(\Psi_{N,J}(\mathbf{x}_r), \Psi_{N,J}(\mathbf{x}_l)) \rightarrow [\mathcal{D}^T\mathcal{D}]_{r,l}$;
 7:    **end for**
 8: **end for**
 9: Seek the optimal regression coefficients $\hat{\alpha}$;
10: Make the decision: $\min_k\{\|\phi_\mathbf{y} - \mathcal{D}\delta_k(\hat{\alpha})\|_2^2\}$.

III. EXPERIMENT AND DISCUSSION

The proposed strategy are verified with MSTAR database, a gallery collected using a 10 GHz SAR sensor with 1×1-ft resolution in range and azimuth. To the best of our knowledge, it is the only publicly released database available for SAR target recognition. We first discuss the effect of some related factors on target recognition performance, followed by several fundamental verifications. Multiple comparative studies are performed finally. Images are collected at various depression angles over $0 \sim 359°$ range of aspect view [2]. Images available are of around 128×128 pixels. To standardize the input, we extract the center 64×64 pixels patches. The cropped image is used to build the steerable wavelet frames.

---

[2]The depression angle refers to the angle between the line of sight from the radar to target and the horizontal plane at the radar.

$$\mathcal{D}^T\mathcal{D} = \begin{bmatrix} \langle\phi(\Psi_{N,J}(\mathbf{x}_1)),\phi(\Psi_{N,J}(\mathbf{x}_1))\rangle & \langle\phi(\Psi_{N,J}(\mathbf{x}_1)),\phi(\Psi_{N,J}(\mathbf{x}_2))\rangle & \cdots & \langle\phi(\Psi_{N,J}(\mathbf{x}_1)),\phi(\Psi_{N,J}(\mathbf{x}_n))\rangle \\ \langle\phi(\Psi_{N,J}(\mathbf{x}_2)),\phi(\Psi_{N,J}(\mathbf{x}_1))\rangle & \langle\phi(\Psi_{N,J}(\mathbf{x}_2)),\phi(\Psi_{N,J}(\mathbf{x}_2))\rangle & \cdots & \langle\phi(\Psi_{N,J}(\mathbf{x}_2)),\phi(\Psi_{N,J}(\mathbf{x}_n))\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\phi(\Psi_{N,J}(\mathbf{x}_n)),\phi(\Psi_{N,J}(\mathbf{x}_1))\rangle & \langle\phi(\Psi_{N,J}(\mathbf{x}_n)),\phi(\Psi_{N,J}(\mathbf{x}_2))\rangle & \cdots & \langle\phi(\Psi_{N,J}(\mathbf{x}_n)),\phi(\Psi_{N,J}(\mathbf{x}_n))\rangle \end{bmatrix}$$
$$= \begin{bmatrix} \kappa(\Psi_{N,J}(\mathbf{x}_1),\Psi_{N,J}(\mathbf{x}_1)) & \kappa(\Psi_{N,J}(\mathbf{x}_1),\Psi_{N,J}(\mathbf{x}_2)) & \cdots & \kappa(\Psi_{N,J}(\mathbf{x}_1),\Psi_{N,J}(\mathbf{x}_n)) \\ \kappa(\Psi_{N,J}(\mathbf{x}_2),\Psi_{N,J}(\mathbf{x}_1)) & \kappa(\Psi_{N,J}(\mathbf{x}_2),\Psi_{N,J}(\mathbf{x}_2)) & \cdots & \kappa(\Psi_{N,J}(\mathbf{x}_2),\Psi_{N,J}(\mathbf{x}_n)) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\Psi_{N,J}(\mathbf{x}_n),\Psi_{N,J}(\mathbf{x}_1)) & \kappa(\Psi_{N,J}(\mathbf{x}_n),\Psi_{N,J}(\mathbf{x}_2)) & \cdots & \kappa(\Psi_{N,J}(\mathbf{x}_n),\Psi_{N,J}(\mathbf{x}_n)) \end{bmatrix} \tag{14}$$

A. Effect of the Related Factors on Performance

We first perform two sets of experiments for different order of Riesz transform and different scale of wavelet. Four military vehicles, BMP2, BTR60, T72, and T62 are used to perform 4-class target recognition. The number of aspect views available for different targets are tabulated in TABLE I. The number of aspect views available for training are in bold, while the ones for testing are underlined. The items in bracket mark the configuration by the series number. The detail explanation on dataset can be found in Section III-C. Three metrics, arc length, projection metric, and Binet-Cauchy metric, and two Grassmann kernels, projection kernel and Binet-Cauchy kernel are utilized to demonstrate the performance.

TABLE I.   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*) <u>196</u> (*SN_9566*) <u>196</u> (*SN_c21*)	196 (*SN_132*) <u>195</u> (*SN_812*) <u>191</u> (*SN_s7*)	<u>195</u>	<u>273</u>	<u>1246</u>

1) Scale of Wavelet Frames: The recognition performance across different scale of wavelet frame is shown in TABLE II, where the $N(5)^{th}$-order Riesz transform is used to generate the representation model. Experiments are performed by changing the scale of wavelet from $J(0)^{th}$ to $J(2)^{nd}$. The $J(0)^{th}$-scale wavelet refers to the original $N(5)^{th}$-order Riesz transform coefficients, while the first- and second-scale wavelet are the rescaled and dilated ones, corresponding to the wavelet coefficients of 32×32, 16×16 in size. The bar diagram is shown in Fig. 5. We can see that the performance has been improved with the multi-scale wavelet frames. The best performance, 0.8957, is obtained when the $J(2)^{nd}$-scale wavelet are performed. The experimental results demonstrate that each wavelet coefficients have some discrimination power. The performance can be improved by aggregating the discrimination information across the scale.

2) Order of Riesz Transform: The recognition accuracy across the order of Riesz transform is shown in TABLE III, where $J(2)^{nd}$-scale pyramid wavelet are built. Experiments are conducted by adjusting the order of Riesz transform from $N(2)^{nd}$ to $N(6)^{th}$. The bar diagram is drawn in Fig. 6.

From TABLE III (a), we could see that the recognition rates obtained using Projection metric and Binet-Cauchy metric are

TABLE II.   THE RECOGNITION ACCURACY ACROSS THE SCALE OF WAVELET FRAMES.

(a) The accuracies obtained using Algorithm 1

Metric	$J=0$	$J=1$	$J=2$
Arc length	0.8283	0.8555	0.8844
Projection	0.8307	0.8604	0.8836
Binet-Cauchy	0.8202	0.8644	0.8957

(b) The accuracies obtained using Algorithm 2

Kernel	$J=0$	$J=1$	$J=2$
Projection	0.8684	0.8788	0.9278
Binet-Cauchy	0.8620	0.8941	0.9230
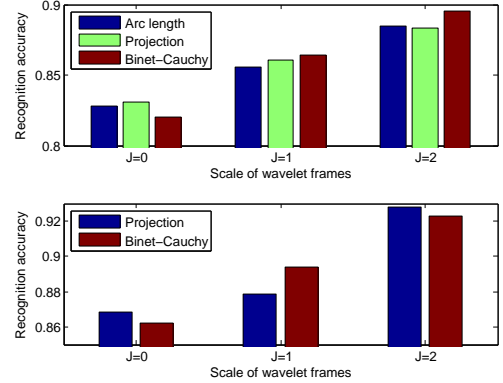
Fig. 5. The pictorial illustration of recognition performance across the scale of wavelet frames.

irregularly varied. The recognition accuracy obtained by the $N(5)^{th}$-order Riesz transform is slightly better than the others. From TABLE III (b), we could see that Projection kernel outperforms Binet-Cauchy one. The recognition accuracy obtained using Binet-Cauchy kernel fluctuates more sharply than the one obtained using Projection kernel. This is because the definition of Binet-Cauchy kernel refers to determinant operation, while Projection kernel only involves addition operation. The $N(5)^{th}$-order Riesz transform consistently achieves the best performance.

Based on the above experiments, $J(2)^{nd}$-scale and $N(5)^{th}$-order Riesz transform is used to produce the proposed representation model. This settings will be utilized in the following comparative studies.

TABLE III.   THE RECOGNITION ACCURACY ACROSS THE ORDER OF RIESZ TRANSFORM.

(a) The rates obtained using Algorithm 1

Metric	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
Arc length	0.8684	0.8700	0.8700	0.8844	0.8748
Projection	0.8708	0.8700	0.8756	0.8836	0.8764
Binet-Cauchy	0.8652	0.8692	0.8820	0.8957	0.8716

(b) The rates obtained using Algorithm 2

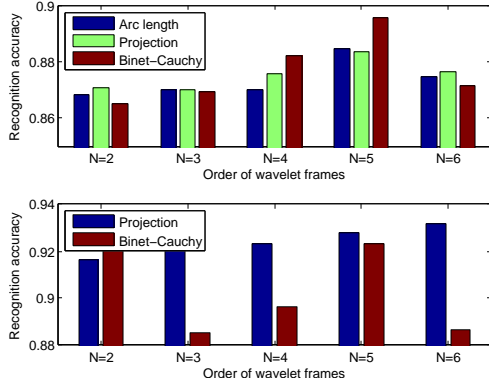| Kernel	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
Projection	0.9165	0.9213	0.9230	0.9278	0.9318
Binet-Cauchy	0.9238	0.8852	0.8965	0.9230	0.8868

Fig. 6.   The pictorial illustration of recognition performance across the order of wavelet frames.

B. Fundamental Verifications

The subsequent experiments devote to several fundamental verifications, such as the study of kernel fusion, the validation of kernel function, and the comparison with our preceding work. The experimental setting is same to the precious one.

1) The Study of Kernel Fusion: In this paper, the global kernel function is generated by accumulating the Grassmann kernel over the wavelet scale, as defined in (9). It is natural to ask how to generate the kernel weight. An intuitive idea is to set the kernel weight equally. In addition, several studies recommend to generate a learned kernel weights [34], [37]–[40]. Inspired by these works, this paper introduces two rule for kernel fusion. The first rule produces the kernel weight according to the classification accuracy obtained by each single kernel function, as developed in [38]. It can be expressed as

$$\beta_r = \frac{R_r - \delta}{\sum_{j=0}^{J-1} \{R_j - \delta\}} \quad (16)$$

where R_r denotes the recognition rate obtained using the rth scale kernel $\kappa_r(\cdot, \cdot)$, and δ is a certain threshold that should be less than the minimum of accuracy obtained using single kernel function.

The second rules generates the kernel weight via the kernel alignment strategy [39]. The kernel alignment is a measure of similarity between two kernel Gram matrix [41],

$$A(\mathcal{K}_m, \mathcal{K}_n) = \frac{\langle \mathcal{K}_m, \mathcal{K}_n \rangle_F}{\sqrt{\langle \mathcal{K}_m, \mathcal{K}_m \rangle_F \cdot \langle \mathcal{K}_n, \mathcal{K}_n \rangle_F}}$$

where $\langle \mathcal{K}_m, \mathcal{K}_n \rangle_F = \sum_i \sum_j \kappa_m(\mathbf{x}_i, \mathbf{x}_j) \kappa_n(\mathbf{x}_i, \mathbf{x}_j)$. The kernel weight is then set as

$$\beta_r = \frac{A(\mathcal{K}_r, \mathcal{K}_d)}{\sum_{j=0}^{J-1} A(\mathcal{K}_j, \mathcal{K}_d)} \quad (17)$$

where $\mathcal{K}_d$ is the ideal kernel matrix whose elements are defined as

$$\mathcal{K}_d(i, j) = \begin{cases} 1 & \text{if } \mathbf{x}_i, \mathbf{x}_j \text{ from the same class} \\ 0 & \text{otherwise.} \end{cases}$$

To study the effect of kernel weight on recognition performance, a set of experiments are performed, where the equal weight is compared with the learned one. The experimental results are given in TABLE IV, where Learn$_1$ and Learn$_2$ denote the kernel weight obtained by (16) and (17).

TABLE IV.   THE RECOGNITION ACCURACY OBTAINED USING VARIOUS KERNEL WEIGHT SCHEMES.

	Equal	Learn$_1$	Learn$_2$
Projection Kernel	0.9318	0.9352	0.9334
Binet-Cauchy Kernel	0.9230	0.9246	0.9238

From TABLE IV, we would come the conclusion as follows. The performance of two learned kernel weight schemes are better than equal weight. The recognition rate obtained by Learn$_1$ is better than the one obtained by kernel alignment scheme. The performance improvement obtained by projection kernel is more significant than the one obtained by Binet-Cauchy kernel.

2) The Validation of Kernel Function: As proved by Jihun Hamm and Daninel D. Lee [29], the projection kernel and Binet-Cauchy kernel are valid Grassmann kernel function. They are associated with the projection metric and Binet-Cauchy metric, respectively. Though several kernel functions are widely used in the preceding works [40], [42], they are valid only in the vector space, and hence unreasonable to be utilized on Grassmann manifolds. To verify this proposition, this paper performs a set of experiments. Gaussian RBF kernel is conjunction with Grassmann metric to form a psedo-kernel function. It is then compared with the proposed method. Gaussian RBF is defined as

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{ -\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \right\}$$

where γ is the width parameter. The measure of similarity $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$ is realized by Grassmann metric. The experimental results are given in TABLE V.

TABLE V.   THE VERIFICATION OF KERNEL FUNCTION.

Gaussian RBF			Grassmann Kernel	
Projection metric	BC metric	Arc length	Projection kernel	BC Kernel
0.8720	0.8307	0.8323	0.9318	0.9230

Though Grassmann metric is used to generate the measure of similarity, the recognition accuracy of Gaussian RBF is still much lower than the one obtained using Grassmann kernel. This is because Gaussian RBF is not a valid Grassmann kernel, and hence unreasonable to be applied on Grassmann manifolds. The performance of Binet-Cauchy metric and arc length are worsen than projection metric.

3) The comparison of Representation Model: This paper develops a new representation model via the steerable wavelet frames. Though the proposed method is an improvement version of our preceding works [43], sparse representation of monogenic signal on Grassmann manifold, their core ideas are significantly different, as summarized in the following items.

- **The formation of Grassmann manifolds.** This paper builds Grassmann manifolds with steerable wavelet coefficients, while our preceding work constructs Grassmann manifolds by the multi-resolution monogenic signal.
- **The measure of similarity.** This paper generates the measure of similarity by accumulating Grassmann metric over the wavelet scale.
- **The definition of kernel function.** This paper defines the global kernel function by assembling Grassmann kernel over the wavelet scale.

To test the performance of proposed representation model, a set of experiments are performed, where the steerable wavelet representation are compared with the multi-resolution monogenic signal. The experimental results are shown in TABLE VI.

TABLE VI. THE COMPARISON OF MONOGENIC SIGNAL REPRESENTATION WITH STEERABLE WAVELET REPRESENTATION.

Model	Monogenic signal [43]		Steerable wavelet	
Classifier	SVM	SRC	SVM	SRC
Accuracy	0.9085	0.9270	0.9229	0.9318

From TABLE VI, we can see that whether SVM of SRC are utilized, the performance of steerable wavelet representation model are consistently better than the one of multi-resolution monogenic signal representation.

4) The comparison of learning model: This paper develops a new representation model via the steerable wavelet frames. The proposed representation model is then cast into the Hilbert space. To achieve target classification, the proposed representation model is fed into kernel sparse learning. Another popularly used kernel analysis skill is support vector machine learning. To demonstrate the advantage of sparse learning model, a set of experiment are performed, where the sparse learning is compared with SVM. To make the comparison fair, the proposed representation model is set as the input of SVM. In addition, the Grassmann kernel is utilized to search the support vector. The experimental results are given in TABLE VII.

TABLE VII. THE COMPARISON OF SPARSE LEARNING AND SUPPORT VECTOR MACHINE LEARNING.

Classifier	SVM		SRC	
Kernel	Projection	Binet-Cauchy	Projection	Binet-Cauchy
Accuracy	0.9229	0.9213	0.9318	0.9230

As can be seen from TABLE VII, the performance of sparse learning is better than SVM. The improvement obtained using the projection kernel is more significant than the Binet-Cauchy kernel.

C. Performance Evaluation

Target recognition in SAR image is a non-trivial problem due to the peculiar working mechanism. Unlike the optical sensor, SAR imaging of the same target taken at different aspect views will result in significantly different phenomenology. To deal with these problems, the preceding works develop several algorithms by the monogenic signal [16], [44]–[46]. Actually, these works have been done around the $J(0)^{th}$-scale and $N(1)^{st}$-order Riesz transform. Therefore, these algorithms are employed as the baseline for performance comparison. In addition, the prototype of sparse signal modeling [35] and support vector machine learning are included. It should be noted that Grassmann kernel is utilized in the framework of SVM, *i.e.*, the input of SVM is the proposed representation model. The methods to be studied are summarized in TABLE VIII. To model the real battle-field condition, we pursue four kinds of experiments, configuration variation, nose corruption, articulation and occlusion, small sample size.

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

Abbre.	Full name (description)	Ref.
SRC	Sparse representation-based classification	[35]
MSRC	Sparse representation of monogenic signal	[44]
SRC$^{\mathcal{R}}$	Sparse representation on Riemannian manifolds	[45]
TJSR	Joint Sparse Representation of monogenic signal	[46]
SVM$^{\mathcal{G}}$	Support vector machine learning with Grassmann kernel	—
KSRC	Kernel sparse representation-based classifier	[32]
KLMC	Kernel linear coding of monogenic signal	[45]
KSR$^{\mathcal{M}}$	Monogenic signal representation on Grassmann manifolds	[43]
KSR$^{\mathcal{G}}$	Steerable wavelet representation on Grassmann manifolds	Alg.2

1) Configuration Variations: Under the realistic battlefield, there may be many different physical target configurations that could be categorized in a single class. In this context, the configuration simply refers to physical change and structural modifications. Examples of such variability are listed in TABLE IX. This sub-section deals with target recognition on different configurations. We train the classification algorithms by the standard, and test them by the variants.

TABLE IX. EXEMPLARS FOR CONFIGURATION VARIATION

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

Four military vehicles, BMP2, T72, BTR60, and T62 are selected. BMP2 and BTR60 are armored personnel carrier, while T72 and T62 are main-battle tank. BMP2 and T72 have several configuration variants with small structural modifications. The details on configuration can be found in TABLE I. Their are three variants for BMP2 and T72, SN_9563, SN_9566, SN_c21 (BMP2), SN_132, SN_812, and SN_s7 (T72). Only SN_9563 and SN_132 taken at a 17° depression angle are used to train the algorithms, while the remaining, SN_9566, SN_c21 and SN_132, SN_s7 collected at a 15° depression angle are used for testing. Both the configuration and depression angle are different between the images used for training and those for testing.

The performance comparison is shown in TABLE X, where both the overall recognition rate and confusion matrix are

TABLE X.   THE PERFORMANCE FOR TARGET RECOGNITION ON CONFIGURATION VARIATION.

(a) Confusion matrix obtained using SRC [35]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9617	0.0255	0.0102	0.0026	
T72	0.0466	0.6477	0.0078	0.2979	0.8708
BTR60	0.0154	0.0	0.9692	0.0154	
T62	0.0	0.0147	0.0	0.9853	

(b) Confusion matrix obtained using MSRC [44]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9286	0.0230	0.0255	0.0230	
T72	0.0415	0.6736	0.0259	0.2591	0.8748
BTR60	0.0	0.0	0.9897	0.0103	
T62	0.0	0.0	0.0	1.0	

(c) Confusion matrix obtained using TJSR [46]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9107	0.0510	0.0230	0.0153	
T72	0.0181	0.8349	0.0	0.1425	0.9117
BTR60	0.0256	0.0103	0.9385	0.0256	
T62	0.0	0.0037	0.0	0.9963	

(d) Confusion matrix obtained using $\text{SVM}^{\mathcal{G}}$ (Projection)

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9311	0.0255	0.0306	0.0128	
T72	0.0337	0.8394	0.0	0.1269	0.9229
BTR60	0.0154	0.0	0.9744	0.0103	
T62	0.0037	0.0037	0.0	0.9927	

(e) Confusion matrix obtained using KSRC [32]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9464	0.0383	0.0102	0.0051	
T72	0.0363	0.7746	0.0052	0.1839	0.9045
BTR60	0.0256	0.0	0.9641	0.0103	
T62	0.0	0.0073	0.0073	0.9853	

(f) Confusion matrix obtained using CKLR [16]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9719	0.0281	0.0	0.0	
T72	0.0259	0.8394	0.0026	0.1321	0.9325
BTR60	0.0256	0.0	0.9641	0.0103	
T62	0.0	0.0037	0.0	0.9963	

(g) Confusion matrix obtained using $\text{KSR}^{\mathcal{M}}$ [43]

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9311	0.0357	0.0281	0.0051	
T72	0.0570	0.8420	0.0026	0.0984	0.9270
BTR60	0.0	0.0	0.9897	0.0103	
T62	0.0	0.0	0.0037	0.9963	

(h) Confusion matrix obtained using $\text{KSR}^{\mathcal{G}}$ (Algorithm 2)

Target	BMP2	T72	BTR60	T62	Overall
BMP2	0.9566	0.0408	0.0	0.0026	
T72	0.0596	0.8264	0.0	0.1140	0.9318
BTR60	0.0	0.0	0.9949	0.0051	
T62	0.0	0.0	0.0	1.0	

given. The performance of BTR60 and T62 are satisfied. The lowest recognition rate of BTR60 is 0.9385, while the lowest one of T62 is 0.9853. On the other side, the recognition accuracy for BMP2 and T72 are drastically varied. The recognition accuracy of BMP2 fluctuates from 0.8903 to 0.9719, while the recognition rate for T72 changes from 0.6477 to 0.8420. This is because both the configuration and depression angle are different between the images for training and those for testing. The performance of proposed method are similar to the one of CKLR, much better than the remaining. The results demonstrate that the proposed strategy could deal with the configuration variation even under the quasi-battle circumstance.

2) Random Noise Corruption: Pixels in SAR image usually exhibit speckle due to the coherent combinations of returns from the the scatterers. The relative motion of scatters is on the order of a wavelength of the incident signal. Thus, it is necessary to test the performance under noise corruption. Again, four kinds of targets, BMP2, BTR60, T72, and T62, are employed. To simulate noise, we corrupt a percentage of randomly chosen pixels from each of the images available for testing, replacing their intensity values with the independent and identically distributed samples drawn from a uniform distribution $\mathcal{U}[0, u_M]$, where u_M is the largest possible pixel value. The corrupted pixels are randomly chosen, *i.e.*, their locations are unknown for the algorithms.
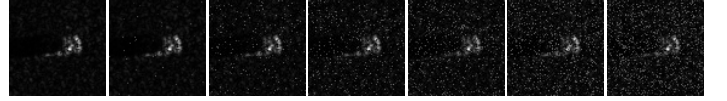

Fig. 7. Illustration of random noise corruption. The set of images demonstrate that 0%, 5%, 10%, 15%, 20%, 25%, and 30% of pixels are corrupted.

TABLE XI.   THE PERFORMANCE FOR PROPOSED REPRESENTATION MODEL UNDER RANDOM NOISE CORRUPTION.

	SRC	MSRC	$\text{SRC}^{\mathcal{R}}$	TJSR	CKLR	$\text{KSR}^{\mathcal{G}}$
0%	0.8764	0.8868	0.9117	0.9117	0.9325	0.9318
5%	0.8587	0.8515	0.7873	0.8427	0.8619	0.9238
10%	0.7945	0.8090	0.7408	0.8041	0.8515	0.9205
15%	0.6653	0.7504	0.7022	0.7480	0.8314	0.8941
20%	0.5618	0.6661	0.6846	0.7415	0.8001	0.8531
25%	0.5016	0.5939	0.6100	0.7030	0.7817	0.8339
30%	0.4655	0.5393	0.5923	0.6821	0.7504	0.8098
Drop	41.09%	34.75%	31.94%	22.96%	18.21%	12.12%

The recognition rate is listed in TABLE XI. For sparse signal representation, the robust version, the extended ℓ_1-minimization is used to seek the regression coefficients [35]. We vary the percentage of corrupted pixels from 0 percent to 30 percent, by which the scattering phenomenology of target can be manually changed. Some examples of noise corruption are demonstrated in Fig. 7. The proposed strategy consistently outperforms the reference methods. From 5 percent upto 30 percent noise corruption, the proposed strategy is 6.19%, 6.90, 6.27%, 5.30%, 5.22%, 5.94% better than the nearest competitors. The drop of recognition accuracy for proposed representation model is 10.12%, significantly lower than the competitors. Even at 30 percent noise corruption, the performance of proposed representation model is 34.43 percent, 27.05 percent, 21.75 percent, 12.77 percent, and 5.94 percent better than SRC, MSRC, $\text{SRC}^{\mathcal{R}}$, TJSR, and CKLR. The results demonstrate that the proposed strategy is robust towards noise corruption.

3) Articulation and Occlusion: The invariance to articulation and occlusion, *e.g.*, turret rotation, open hatch, is desirable in the real batter-fields. This sub-section considers target articulation and occlusion. Three kinds of military vehicles, 2S1, BRDM2, and ZSU23/4 are employed, among which

BRDM2 and ZSU23/4 have articulated variants. Example of the standard and articulated variants are demonstrated in Fig. 8.

Fig. 8.   Illustration of articulation, (a) turret straight, (b) turret articulated.

To evaluate the performance under articulation and occlusion, the standard taken at a 17° depression angle are used to train the algorithm, while the variants collected at 30° and 45° depression angle are used for testing. The number of aspect views available for three targets taken at 17°, 30°, and 45° depression angle are tabulated in TABLE XII. The items in bracket give the articulated variants.

TABLE XII.   IMAGES AVAILABLE FOR STANDARDS AND VARIANTS.

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

The recognition performance are given in TABLE XIII, where the overall recognition rate is in parentheses. The drop of recognition accuracy is pictorially shown in Fig. 9. The recognition accuracy degrades drastically when the depression angle varies from 30° to 45°. At the first scenario, a moderate change of 13° from 17° to 30° depression angle, the recognition performance is satisfied. The performance of proposed strategy is slightly better than the baseline. At the second scenario, a drastic change of 28° from 17° to 45° depression, the result is significantly different. Though the recognition accuracy of all methods are drastically decreased, the proposed strategy still achieve the best performance. It is 5.32%, 6.97%, 14.11%, 13.76%, and 24.04% better than the competitors. This set of experiments proves that the proposed strategy could handle articulation and occlusion.

4) Small Sample Size: Another important issue for target recognition is the robustness to limited training samples. This sub-section evaluates small sample size problem. BMP2, T72, BTR60, and T62 are employed. The details on aspect view can be found in TABLE I. Different from the preceding experiment, the size of training set are gradually decreased from 1020 to 307, corresponding to 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3 times of the original size. Samples available for testing are kept unchanged, *i.e.*, the total number of testing sample is 1246. The experimental results are shown in TABLE XIV.

With the decrease of training set, the recognition performance is deteriorated too. However, the proposed strategy consistently achieve the satisfied performance. Even with the 307 training samples, the performance of proposed representation model (SVM$^{\mathcal{G}}$ and KSR$^{\mathcal{G}}$) is much better than the others. The experimental results demonstrate that the proposed strategy

TABLE XIII.   THE PERFORMANCE FOR TARGET RECOGNITION ON DEPRESSION VARIATION.

(a) Confusion matrix obtained using SRC [35]

Target	30°-Depression (**0.9488**)			45°-Depression (**0.5366**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9757	0.0243	0.0	0.7789	0.2013	0.0198
BRDM2	0.0595	0.9190	0.0214	0.4232	0.5650	0.0118
ZSU23	0.0172	0.0222	0.9606	0.6706	0.0237	0.3057

(b) Confusion matrix obtained using MSRC [44]

Target	30°-Depression (**0.9497**)			45°-Depression (**0.6394**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9618	0.0	0.0382	0.8647	0.0825	0.0528
BRDM2	0.0429	0.9262	0.0310	0.2648	0.4610	0.2742
ZSU23	0.0123	0.0222	0.9655	0.1848	0.1588	0.6564

(c) Confusion matrix obtained using SRC$^{\mathcal{R}}$ [45]

Target	30°-Depression (**0.9515**)			45°-Depression (**0.6359**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9132	0.0868	0.0	0.4620	0.5182	0.0198
BRDM2	0.0119	0.9810	0.0071	0.0591	0.9267	0.0142
ZSU23	0.0148	0.0369	0.9483	0.1872	0.3436	0.4692

(d) Confusion matrix obtained using TJSR [46]

Target	30°-Depression (**0.9524**)			45°-Depression (**0.7073**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9861	0.0069	0.0069	0.5644	0.3366	0.0990
BRDM2	0.0214	0.9000	0.0786	0.0757	0.6738	0.2506
ZSU23	0.0025	0.0148	0.9828	0.0095	0.1469	0.8436

(e) Confusion matrix obtained using CKLR [16]

Target	30°-Depression (**0.9712**)			45°-Depression (**0.7238**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9896	0.0104	0.0	0.7030	0.2937	0.0033
BRDM2	0.0667	0.9310	0.0024	0.0709	0.7991	0.1300
ZSU23	0.0	0.0	1.0	0.0900	0.2464	0.6635

(e) Confusion matrix obtained using KSR$^{\mathcal{M}}$ [43]

Target	30°-Depression (**0.9381**)			45°-Depression (**0.7282**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9722	0.0	0.0278	0.6865	0.2442	0.0693
BRDM2	0.0	0.8690	0.1310	0.0515	0.7096	0.2389
ZSU23	0.0049	0.0099	0.9852	0.0142	0.2180	0.7678

(f) Confusion matrix obtained using KSR$^{\mathcal{G}}$(Algorithm 2)

Target	30°-Depression (**0.9820**)			45°-Depression (**0.7770**)		
	2S1	BRDM2	ZSU23	2S1	BRDM2	ZSU23
2S1	0.9861	0.0139	0.0	0.7030	0.2871	0.0099
BRDM2	0.0214	0.9667	0.0119	0.0662	0.7943	0.1395
ZSU23	0.0049	0.0	0.9951	0.1185	0.0687	0.8128

TABLE XIV.   RECOGNITION ACCURACY ACROSS THE SIZE OF TRAINING SET.

	SRC	MSRC	TJSR	SVM$^{\mathcal{G}}$	KSR$^{\mathcal{M}}$	CKLR	KSR$^{\mathcal{G}}$
1020	0.8708	0.8748	0.9117	0.9229	0.9270	0.9325	0.9318
918	0.8748	0.8820	0.9077	0.9189	0.9350	0.9213	0.9270
816	0.8764	0.8828	0.8973	0.9141	0.9262	0.9205	0.9173
713	0.8724	0.8764	0.8900	0.9093	0.9165	0.9125	0.9165
612	0.8716	0.8740	0.8981	0.9133	0.9262	0.9117	0.9222
511	0.8604	0.8620	0.8836	0.8916	0.8989	0.8981	0.9141
408	0.8387	0.8499	0.8700	0.8892	0.8933	0.8812	0.8957
307	0.7978	0.8266	0.8266	0.8684	0.8531	0.8627	0.8700

could deal with the difficult training regime of small sample size.

IV. CONCLUSION

To handle extended operating conditions, this paper recommends a novel representation model via the steerable pyramids and tight wavelet frames. The proposed representation model
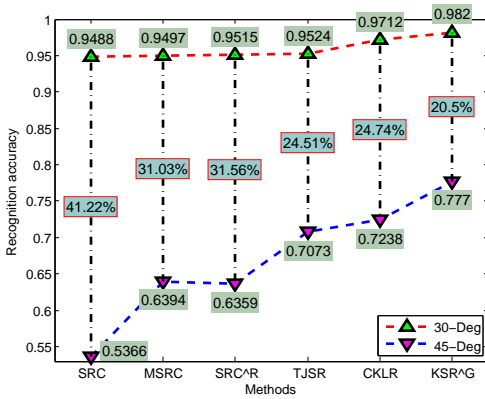
Fig. 9. The drop of recognition performance when the depression angle is varied from 30° to 45°. The percents provide the drop of recognition accuracy.

commutes with translation, scaling, and spatial rotations. To deal with difficult training regimes, the proposed representation model is cast into the abstract Hilbert space. The main contributions of this work include: i) we represent SAR image via the Nth-order and J^{th}-scale Riesz wavelet transform; ii) we define the measure of (dis)similarity between two proposed representation models via Grassmann metric; iii) we embed the Grassmann manifold into an abstract Hilbert space.

To test the feasibility of proposed strategy, multiple comparative studies have been done on MSTAR SAR database. The experimental results demonstrate: (i) the proposed representation model could characterize the scattering phenomenology of SAR image, (ii) the proposed strategy could deal with configuration variation and depression change because the representation model allows for translation, scaling, and spatial rotation. (iii) the proposed strategy demonstrates great potential for real-world applications, *e.g.*, noise corruption, articulation, occlusion, and small size of training set.

The performance achieved using the proposed strategy is the result of coupling steerable wavelet frames and kernel analysis of sparse coding. The redundant dictionary is formed by concatenating the mappings of training sample directly. An intriguing question for our future work is how to generate a more compact and discriminative dictionary using the developed learning skills. This topic has also been exploited in the recent studies [30], [33]. On the other hand, the recent development on kernel fusion, *i.e.*, multiple kernel learning in the framework of sparse representation, can be also exploited.
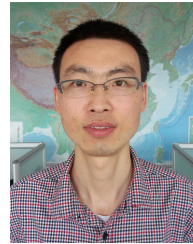
ACKNOWLEDGMENT

REFERENCES

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

[2] 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.

[3] L. Novak, G. Owirka, and W. Brower, "Performance of 10- and 20-target MSE classifiers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, pp. 1279–1289, Oct. 2000.

[4] 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*, vol. 4727. SPIE, 2002, pp. 265–279.

[5] L. Potter and R. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[6] 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.

[7] 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.

[8] C. Olson and D. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 103–113, Jan. 1997.

[9] U. Srinivas, V. Monga, and R. Raj, "SAR automatic target recognition using discriminative graphical models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 591–606, Jan. 2014.

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

[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] D. Zang and G. Sommer, "Signal modeling for two dimensional image structures," *J. Vis. Commun. Image Represent.*, vol. 18, no. 1, pp. 81–99, 2007.

[13] 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.

[14] M. Mellor and M. Brady, "Phase mutual information as a similarity measure for registration," *Med. Image Anal.*, vol. 9, no. 4, pp. 187–201, 2005.

[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] G. Dong and G. Kuang, "Classification on the monogenic scale space: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2527–2539, Aug. 2015.

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

[18] 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.

[19] S. Held, M. Storath, P. Massopust, and B. Forster, "Steerable wavelet frames based on the Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 653–667, Mar. 2010.

[20] M. Unser and D. Ville, "Wavelet steerability and the higher-order Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 636–652, Mar. 2010.

[21] 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.

[22] 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.

[23] P.-A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation," *Acta Applicandae Mathematica*, vol. 80, pp. 199–220, 2004.

[24] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, Oct. 2008.

[25] 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.

[26] 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.

[27] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 3136–3144, Sep. 1991.

[28] M. Unser, N. Chenouard, and D. Ville, "Steerable pyramids and tight wavelet frames in L_2 (R^d)," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2705–2721, Oct. 2011.

[29] J. Hamm and D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," 2008, pp. 376–383.

[30] 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.

[31] 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.

[32] 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.

[33] H. V. Nguyen, V. Patel, N. Nasrabadi, and R. Chellappa, "Design of nonlinear kernel dictionaries for object recognition," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5123–5135, Dec. 2013.

[34] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "Multiple kernel sparse representations for supervised and unsupervised learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2905–2915, Jul. 2014.

[35] 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.

[36] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Neural Inf. Process. Syst.*, 2007, pp. 801–808.

[37] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3013–3024, Jul. 2014.

[38] H. Tanabe, T. B. Ho, C. H. Nguyen, and S. Kawasaki, "Simple but effective methods for combining kernels in computational biology," in *IEEE International Conference on Research, Innovation and Vision for the Future*, Jul. 2008, pp. 71–78.

[39] S. Qiu and T. Lane, "A framework for multiple kernel support vector regression and its applications to siRNA efficacy prediction," no. 2, pp. 190–199, Jun. 2009.

[40] M. Gonen and A. Ethem, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, July 2011.

[41] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. S. Kan-dola, "On kernel-target alignment," in *Proc. Neural Inf. Process. Syst.*, 2001, pp. 367–373.

[42] G. Camps-Valls, L. Gomez-Chova, J. Muoz-Mari, J. Vila-Francs, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[43] G. Dong and G. Kuang, "SAR target recognition via sparse representation of monogenic signal on Grassmann manfolds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1308–1319, Mar. 2016.

[44] 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.

[45] 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.

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

Ganggang Dong received the B.S degree in UAV application engineering from the Academy of Artillery, Hefei, China, in 2004 and the M.S. degree and Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2012 and 2016. His research interests include the applications of compressed sensing and sparse representations, SAR image interpretation, and manifolds learning. He has authored 12 scientific papers in peer-reviewed journals and conferences, including IEEE TIP, IEEE JSTARS, IEEE GRSL, IEEE SPL, SPIE JARS, and IGARSS.

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.

Na Wang received the B.S. degree and M.A. Eng. degree in automated command and control, information and communciation engineering from the Air Force Engineering University, Xian, China, in 2005, 2008, and the Ph. D. degree in information and communciation engineering from the National University of Defense Technology, Changsha, China, in 2012.

Her research interests include the polarmetric SAR image interpretation, automatic target detection.
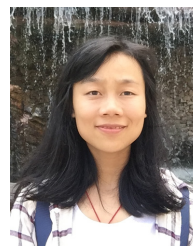
Wei Wang received the B.S degree in Lanzhou University, Lanzhou, China, in 2007 and the M.S. degree and Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2010 and 2013.

She is currently a lecturer in the College of Opto-electronics Science and Engineering, National University of Defense Technology. Her research interests include computer vision, pattern recognition, and shape matching.