Target Recognition in SAR Images via Classification on Riemannian Manifolds

Ganggang Dong and Gangyao Kuang, *Member, IEEE*

***Abstract*—In this letter, SAR target recognition via classification on Riemannian geometry is presented. To characterize SAR images, which have broad spectral information yet spatial localization, a 2-D analytic signal, the monogenic signal, is used. Then, the monogenic components are combined by computing a covariance matrix whose entries are the correlation of the components. Since the covariance matrix, a symmetric positive definite one, lies on the Riemannian manifold, it is unreasonable to be dealt with by the standard learning techniques. To address the problem, two classification schemes are proposed. The first maps the covariance matrix into the vector space, and feeds the resulting descriptor into a recently developed framework, sparse representation-based classification (SRC). The other embeds the Riemannian manifold into an implicit reproducing kernel Hilbert space (RKHS), followed by least square fitting technique to recover the test. The inference is reached by evaluating which class of samples could reconstruct the test as accurately as possible.**

***Index Terms*—Monogenic signal, Riemannian manifolds, sparse representation, target recognition, SAR.**

I. Introduction

Synthetic aperture radar (**SAR**) target recognition has become an essential research topic for many fields, *e.g.*, reconnaissance, surveillance, and tracking. Despite significant efforts by the signal processing and machine learning, recognizing targets in SAR images is still a challenging problem on account of its peculiar image formation (the specular reflections of coherent sources), which results in the mutable characteristics even with small changes in pose, depression and configuration.

The present approaches to SAR target recognition include template matching [1], feature-based method [2], model-based method [3], and the others [4], [5]. Template matching strategy classifies the query by comparing with the templates generated by the training samples. Its performance could be significantly affected by the variability of the template configuration, the test against target revetments, and the resolution of image. Feature-based method is enslaved to the accuracy of preprocessing steps, *e.g.*, noise reduction and image segmentation, which are still the open problems for SAR images. Model-based method refers to representing the pixel magnitude in SAR image by a parameterized statistical distribution model and predicting the class membership by evaluating which class of parameters could maximize the probability. It easily fails when strong statistical relationship does not exist between the present and the query, due to the difficulty of parameter estimation under limited samples.

The monogenic signal, an 2-D generalization of the analytic signal, has been successfully used in biometric recognition, optic flow estimation and harmonic analysis [6]. The main advantage consists in the invariance-equivariance property with respect to energy and orientation. With the monogenic representation, multiple components can be generated. Since the components are of high-dimension, it is unrealistic to be directly applied to classification system. The remaining problem is how to combine these components to make the computation tractable. An intuitive idea is to define an augmented feature vector via uniform down-sampling, normalization and concatenation of the components [7]. Then, different kinds of information could be unified in manner of concatenation. However, this strategy may be not capable to exploit the inter-correlation between the multiple components.

To exploit the inter-correlation of the monogenic components, a novel technique based on Riemannian geometry is presented. Unlike the preceding work, this paper combines the monogenic components by computing the covariance descriptor whose elements are the correlations of components themselves. The covariance descriptor, a symmetric positive definite (**SPD**) matrix, can be formulated as a connected Riemannian manifold, thus the standard learning skills in the vector space do not make sense. Since the derivatives at a point on the manifold lie in a vector space (the tangent space at the point), the manifold can be converted to the tangent space. This property provides a bridge linking the feature descriptor on the manifolds and the learning techniques in the vector space. To reach the inference, two classification schemes are presented. The first maps the covariance matrix into the tangent space with the matrix logarithm operator. The resulting feature vector is fed into a novel framework, SRC [8]. The second embeds the Riemannian manifold into an implicit RKHS, followed by a least square fitting method to recover the test. The decision is made by evaluating which class of samples could minimize the reconstruction error. Extensive experiments on MSTAR database demonstrate that the superior performance can be achieved using the proposed methods, in comparison to several popular algorithms.

II. Riemannian Geometry

This section presents a brief introduction to Riemannian geometry, focusing on the space of SPD matrices. A more detailed description can be found in Ref [9]. Point in the vector

The authors are with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, 410073, China (e-mail: dongganggang@.nudt.edu.cn)

Manuscript received May 20, 2014; revised June , 2014.
Digital Object Identifier 10.1109/LGRS.2014.2332076

space and the counterpart on the manifold are notated with the lowercase letter ($\mathbf{x} \in \mathbb{R}^m$) and the uppercase one ($\mathbf{X} \in \mathcal{M}$).

Riemannian manifold is a differentiable manifold endowed with a smooth inner product (Riemannian metric) [9]. Point on the manifold has a neighborhood for which there exists a bijection mapping the neighborhood to $\mathbb{R}^m$. For differentiable manifold, it is capable to define the derivatives of the curves on the manifold. The derivatives at a point on the manifold $\left(\mathbf{X} \in \mathcal{M}\right)$ lie in a vector space $\mathcal{T}_{\mathbf{X}}$, *i.e.*, the tangent space at the point $\mathbf{X}$. The shortest curve connecting two points on the manifold $\left(\mathbf{X}, \mathbf{Y} \in \mathcal{M}\right)$ is called the geodesic, whose length gives the distance between the pair of points $\mathcal{D}_G(\mathbf{X}, \mathbf{Y})$.

The exponential map $\exp_{\mathbf{X}} : \mathcal{T}_{\mathbf{X}} \mapsto \mathcal{M}$ is a transform from the tangent vector to the point reached by the geodesic. Since the exponential map $\exp_{\mathbf{X}}$ is onto but only one-to-one in a neighborhood of $\mathbf{X}$, the inverse mapping $\log_{\mathbf{X}} : \mathcal{M} \mapsto \mathcal{T}_{\mathbf{X}}$ is uniquely defined only around a small neighborhood of point $\mathbf{X}$. The exponential map with respect to the Riemannian metric $\exp_{\mathbf{X}}(\mathbf{y}) = \mathbf{X}^{1/2}\exp(\mathbf{X}^{-1/2}\mathbf{y}\mathbf{X}^{-1/2})\mathbf{X}^{1/2}, \mathbf{y} \in \mathcal{T}_{\mathbf{X}}$ is a continuously differentiable mapping in both directions, thus the logarithm map is uniquely defined at all the points on the manifold, $\log_{\mathbf{X}}(\mathbf{Y}) = \mathbf{X}^{1/2}\log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2})\mathbf{X}^{1/2}$ for $\mathbf{Y} \in \mathcal{M}$. Clearly, $\exp_{\mathbf{X}}$ and $\log_{\mathbf{X}}$ are manifold-specific, as well as point-dependent.

The space of $d \times d$ dimensional SPD matrices can be formulated as a connected Riemannian manifold $\left(\mathbf{Sym}_d^+\right)$ [10], and hence are discussed with the language of Riemannian manifold hereafter. In [11], the most widely used similarity for SPD matrices, the affine invariant Riemannian metric (**AIRM**), is introduced. It shows that the AIRM induced Riemannian structure is invariant to inversion and similarity transforms. For two SPD matrices $\mathbf{X}, \mathbf{Y} \in \mathbf{Sym}_d^+$, the AIRM is defined as $\mathcal{D}(\mathbf{X}, \mathbf{Y}) = \|\log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2})\|_F$, where $\|\cdot\|_F$ is the frobenius norm.

Since the AIRM involves an intensive use of matrix inverse, square root, logarithm and exponent, it is computationally demanding. To overcome the computational limitations, Ref [12] presents a new simple family of metrics named Log-Euclidean,

$$\mathcal{D}(\mathbf{X}, \mathbf{Y}) = \|\log(\mathbf{X}) - \log(\mathbf{Y})\| \tag{1}$$

where $\log(\cdot)$ refers to the matrix logarithm operator. It is clear that Log-Euclidean metrics are also Euclidean distances in the vector space structure. Another family of Riemannian metric popularly used is based on the symmetrized Bregman matrix divergence [13]. This paper focuses on a speciality, the Jensen-Bregman LogDet Divergence (JBLD). For two SPD matrices $\mathbf{X}, \mathbf{Y} \in \mathbf{Sym}_d^+$, their JBLD is defined as

$$\mathcal{J}(\mathbf{X}, \mathbf{Y}) = \log\left(\det\left(\frac{\mathbf{X}+\mathbf{Y}}{2}\right)\right) - \frac{1}{2}\log(\det(\mathbf{X}\mathbf{Y})). \tag{2}$$

where $\det(\cdot)$ denotes the determinant. The JBLD only involves scalar logarithm operation, thus it is computationally efficient.

III. The Proposed Methods

A. The Monogenic Signal

The monogenic signal, a 2-D generalization of the analytic signal, has been widely used in biometric recognition, optic flow estimation and signal processing [6]. It has been built around Riesz transform. For any 2-D signal $f(\mathbf{z}), \mathbf{z}=(x, y)^T$ and the frequency spectra $F(\mathbf{u}), \mathbf{u} = (u, v)^T$, the Riesz transformed one in the frequency domain is $F_R(\mathbf{u}) = \frac{j\mathbf{u}}{|\mathbf{u}|}F(\mathbf{u})$. The corresponding signal in the spatial domain is $f_R(\mathbf{z}) = -\frac{\mathbf{z}}{2\pi|\mathbf{z}|^3} * f(\mathbf{z}) = [f_R^u, f_R^v]^T$, where $*$ denotes the convolution; f_R^u, f_R^v are the components with respect to u and v. The multiplication in the Fourier domain corresponds to the convolution in the spatial domain, thus the Riesz transform is numerically realized as the multiplication in the Fourier domain. The combination of signal with the Riesz transformed one yields a sophisticated 2-D analytic signal named the monogenic signal,

$$f_M(\mathbf{z}) = f(\mathbf{z}) - (i, j)f_R(\mathbf{z}) \tag{3}$$

where i, j are the imagery units. As a special form of quaternion decomposition, (3) can be retrieved by three components, local amplitude, local phase, and local orientation [6], [14]

$$\begin{aligned} amplitude &: A = \sqrt{f(\mathbf{z})^2 + |f_R(\mathbf{z})|^2} \\ phase &: \phi = atan2(|f_R(\mathbf{z})|, f(\mathbf{z})) \\ orientation &: \theta = atan2(f_R^u(\mathbf{z}), f_R^v(\mathbf{z})) \end{aligned} \tag{4}$$

In monogenic representation, the local amplitude represents local energetic information, and is invariant with respect to local structure, while the local phase depicts local structural information, and is invariant with respect to local energy but changes if the local structure varies. Thus the monogenic signal fulfills the invariance-equivariance property with respect to energy and orientation.

Since the practical signal is of finite length, it is necessary to perform infinite extension with the band-pass filter. Following the preceding works [7], [14], the log-Gabor filters are used due to its capability to capture the broad spectral information with compact support. Suppose S-scale log-Gabor filters are performed, then the monogenic signal at various scales can be generated, $\left[f_M^1, f_M^2, \cdots, f_M^S\right]$. The resulting components can be shown as $\mathbf{f} = [A_1, \phi_1, \theta_1, A_2, \phi_2, \theta_2, \cdots, A_S, \phi_S, \theta_S]$, where $\{A_k, \phi_k, \theta_k\}_{k=1}^S$ are local amplitude, local phase and local orientation of monogenic signal at the kth scale. The feature tube $\mathbf{f}$ is unrealistic to be directly applied to classifier, due to the high data dimension. To solve the problem, Ref [7] defines an augmented feature vector via uniform downsampling, normalization, and concatenation of the monogenic components. However, this method may be not capable to exploit the inter-correlation of multiple components.

B. A Novel Combination Technique

Covariance descriptor, an effective tool to combine different features, has been successfully used in many fields [10], [13], due to the advantage to exploit the inter-correlation of various features. Thus, various kinds of information can be unified. Inspired by the preceding works, this paper combines the monogenic components by computing the covariance descriptor. Given a 2-D signal of $w \times h$ pixels in size, the monogenic feature tube, $\{\mathbf{f}_k \in \mathbb{R}^d\}_{k=1}^{wh}$, can be obtained by (4). Then the $d \times d$-dimensional covariance matrix of $\mathbf{f}$ can be computed by

$$\mathbb{C}_v(p, q) = \frac{1}{wh-1}\sum_{k=1}^{wh}(\mathbf{f}_k(p) - \bar{\mathbf{f}}(p))(\mathbf{f}_k(q) - \bar{\mathbf{f}}(q)) \tag{5}$$

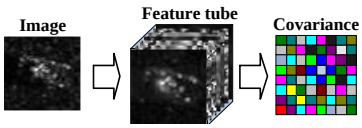

Fig. 1. The generation of the covariance descriptor for monogenic signal.

where $\{\bar{\mathbf{f}}(p)\}_{p=1}^{d}$ is the mean of the pth component. For the covariance matrix $\mathbb{C}_v$, the diagonal elements give the variance of each component, while the non-diagonal ones present the correlations between a pair of components. Thus, it is capable to combine the information in the monogenic components. Moreover, it is of low-dimension because there are only $m = d(d+1)/2$ unique values for symmetry. The generation of the descriptor is pictorially shown in Fig. 1.

C. Sparse Representation of Log-Covariance

The space of covariance matrices with a give dimension in size is not closed with respect to multiplication, and hence does not form a vector space. Most of the usual learning techniques make sense only in an Euclidean space. Thus it would be unreasonable to directly apply the covariance matrix to the standard learning techniques. The conclusion is corroborated by the experiments conducted in Section IV-A. Thanks to the logarithm map [12], the space of SPD matrices can be mapped into the vector space of symmetric matrices. Given a covariance matrix $\mathbb{C}_v \in \mathbf{Sym}_d^+$ and its eigenvalue decomposition $\mathbb{C}_v = \mathbf{U}\Delta\mathbf{U}^T$, where $\mathbf{U}$ composes the eigenvectors and Δ is the diagonal matrix of eigenvalues, the log-covariance can be realized by $\log(\mathbb{C}_v) = \mathbf{U}\log(\Delta)\mathbf{U}^T$. Since the log-covariance forms a vector space with the metric defined by (1), the existing learning skills can be applied. To generate the minimal representation, the upper triangular entries of log-covariance matrix are utilized to form a feature vector,

$$vec(\mathbf{y}) = \left[y_{1,1}, \sqrt{2}y_{1,2}, \sqrt{2}y_{1,3}, \cdots, y_{2,2}, \cdots, y_{d,d}\right]^T \quad (6)$$

for $\mathbf{y} = \log(\mathbb{C}_v)$. Multiplying the non-diagonal elements by $\sqrt{2}$ is to maintain the frobenius norm. To predict the class membership, the resulting vector is fed into SRC [8].

Suppose there are n training samples $\mathbf{x}_1, ..., \mathbf{x}_n$ from K distinct classes. The covariance matrices of the training samples are $\{\mathbb{C}_v^1; ...; \mathbb{C}_v^n\}$. Then, a dictionary can be formed by concatenating the feature vectors (6) of the training samples, $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_n] \in \mathbb{R}^{m\times n}$, where $\mathbf{d}_i = vec(log(\mathbb{C}_v^i))$. It is assumed that features of samples that belong to a single class approximately lie in a subspace [8]. For any test $\mathbf{x}$, its feature vector $\mathbf{d}_\mathbf{x}$ can be generated from the covariance matrix $\mathbb{C}_v^\mathbf{x}$ using (6) and (5). In theory, the feature vector $\mathbf{d}_\mathbf{x}$ can be well recovered by the counterparts of the samples that belong to the same class. Since the class membership is unknown initially, $\mathbf{d}_\mathbf{x}$ is represented over the whole dictionary $\mathbf{D}$,

$$\mathbf{d}_\mathbf{x} = \mathbf{d}_1\alpha_1 + \mathbf{d}_2\alpha_2 + \cdots + \mathbf{d}_n\alpha_n = \mathbf{D}\alpha, \quad (7)$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]^T \in \mathbb{R}^n$ is the representation. Considering the underdetermined system ($m < n$), the solution of (7) is not unique. The popular strategy is to seek the most parsimonious representation via ℓ_1-norm minimization [8]

$$\min_\alpha \|\alpha\|_1 \ s.t. \ \|\mathbf{d}_\mathbf{x} - \mathbf{D}\alpha\|_2 \le \varepsilon, \quad (8)$$

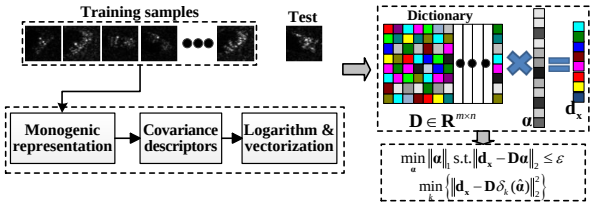

Fig. 2. Classification via sparse representation of log-covariance. The dictionary, composing of the vectorized log-covariance matrices of the training samples, is used to represent the counterpart of the test as a linear combination of them. Via sparsity constraint, the representation can be uniquely generated.

where ε is the error tolerance. By some optimization toolkit, the optimal representation $\hat{\alpha}$ can be obtained. The decision is made by evaluating which class of samples could minimize the reconstruction error,

$$\min_{1\le k\le K} \{\|\mathbf{d}_\mathbf{x} - \mathbf{D}\delta_k(\hat{\alpha})\|_2^2\}, \quad (9)$$

where $\delta_k(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^n$ preserves the entries associated with the kth class, and sets the others to be zeros. The block diagram of the proposed method is displayed in Fig.2.

D. Least square fitting in RKHS

The preceding method transforms the covariance matrix into the tangent space with the logarithm operator. This mapping may result in some information loss, because the logarithm map is uniquely defined only around a small neighborhood of the point on the manifold [10], [11]. To circumvent the problem, our idea here is to embed the Riemannian manifold into an implicit RKHS ($\mathcal{H}$), in which the test can be linearly represented by the training samples.

As verified in the preceding works [15], the logDet divergence metric

$$\kappa(\mathbb{C}_v^1, \mathbb{C}_v^2) = \exp\{-\beta\mathcal{J}_{ld}(\mathbb{C}_v^1, \mathbb{C}_v^2)\} \quad (10)$$

forms a Riemannian kernel, where β is a smooth parameter. Given an implicit embedding function $\phi(\cdot) : \mathbf{Sym}_d^+ \mapsto \mathcal{H}$, which casts the covariance matrix into the RKHS. We build a dictionary by catting the covariance matrices of the training samples in $\mathcal{H}$, $\widetilde{\mathbf{D}} = [\phi(\mathbb{C}_v^1), \phi(\mathbb{C}_v^2), ..., \phi(\mathbb{C}_v^n)]$, where $\mathbb{C}_v^i$ results from the ith training sample. For a new test $\mathbf{x}$, its covariance matrix in $\mathcal{H}$ can be represented as a linear combination of atoms in $\widetilde{\mathbf{D}}$,

$$\phi(\mathbb{C}_v^\mathbf{x}) = \phi(\mathbb{C}_v^1)\alpha_1 + \phi(\mathbb{C}_v^2)\alpha_2 + \cdots + \phi(\mathbb{C}_v^n)\alpha_n = \widetilde{\mathbf{D}}\alpha \quad (11)$$

To produce the unique solution, linear coding method

$$\min_\alpha \|\alpha\|_2^2, s.t., \|\phi(\mathbb{C}_v^\mathbf{x}) - \widetilde{\mathbf{D}}\alpha\|_2^2 \le \varepsilon, \quad (12)$$

is employed [5], [16]. Different from (8), ℓ_2-norm minimization is used to limit the feasible set of the representation. This is because that the data separability between different classes has been enhanced with the embedding function $\phi(\cdot)$. Then, a simple linear regression model could achieve the satisfactory performance yet with much lower computational cost in comparison to sparse technique (8). To find the solver of (12), we rewrite it as the unconstrained one,

$$\min_\alpha \left\{\|\phi(\mathbb{C}_v^\mathbf{x}) - \widetilde{\mathbf{D}}\alpha\|_2^2 + \lambda\|\alpha\|_2^2\right\} \quad (13)$$

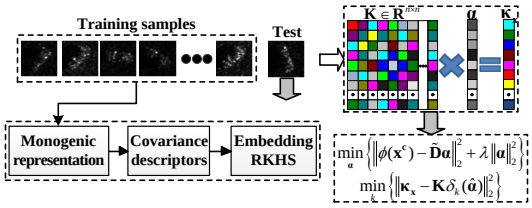

Fig. 3. Classification via least square fitting in $\mathcal{H}$. The dictionary, consisting of the covariance matrices of the training samples, is used to encode the counterpart of the test as a linear combination of them. By least square fitting, the representation can be generated, from which the decision can be drawn.

TABLE I
THREE MODES OF COVARIANCE MATRICES TO BE STUDIED.

Mode1	$[A_1, \phi_1, \theta_1, A_2, \phi_2, \theta_2, \cdots, A_S, \phi_S, \theta_S]$	d=3S
Mode2	$[\mathbf{x}, A_1, \phi_1, \theta_1, A_2, \phi_2, \theta_2, \cdots, A_S, \phi_S, \theta_S]$	d=3S+1
Mode3	$[\mathbf{r}, \mathbf{c}, \mathbf{x}, A_1, \phi_1, \theta_1, A_2, \phi_2, \theta_2, \cdots, A_S, \phi_S, \theta_S]$	d=3S+3

for tradeoff parameter λ. The objective function of (13) (denote by $g(\alpha)$) can be further expanded as

$$\begin{aligned} g(\alpha) &= \phi(\mathbb{C}_v^{\mathbf{x}})^T\phi(\mathbb{C}_v^{\mathbf{x}}) - 2\alpha^T\widetilde{\mathbf{D}}^T\phi(\mathbb{C}_{\mathbf{x}}) + \alpha^T\widetilde{\mathbf{D}}^T\widetilde{\mathbf{D}}\alpha + \lambda\|\alpha\|_2^2 \\ &= \kappa(\mathbb{C}_v^{\mathbf{x}}, \mathbb{C}_v^{\mathbf{x}}) - 2\alpha^T\mathcal{K}_{\mathbf{x}} + \alpha^T\mathbb{K}\alpha + \lambda\alpha^T\alpha \end{aligned} \tag{14}$$

where $\mathbb{K} = \widetilde{\mathbf{D}}^T\widetilde{\mathbf{D}} \in \mathbb{R}^{n\times n}, \{\mathbb{K}_{(i,j)} = \kappa(\mathbb{C}_v^i, \mathbb{C}_v^j)\}_{i,j=1}^n$; $\mathcal{K}_{\mathbf{x}} \in \mathbb{R}^n, \{\mathcal{K}_{\mathbf{x}}(i) = \kappa(\mathbb{C}_v^{\mathbf{x}}, \mathbb{C}_v^i)\}_{i=1}^n$.

By setting the first order derivatives with respect to α to be zero, $\frac{\partial g(\alpha)}{\partial \alpha} = 2\mathbb{K}\alpha - 2\mathcal{K}_{\mathbf{x}} + 2\lambda\alpha = 0$, the solution can be obtained, $\hat{\alpha} = (\mathbb{K} + \lambda\mathbf{I})^{-1}\mathcal{K}_{\mathbf{x}}$, where $\mathbf{I}$ is the identity matrix whose size is same to $\mathbb{K}$. Obviously, the solution of (11) is in close form, thus it is computationally attractive than sparse coding (8). Similar to (9), the inference is reached by evaluating which class of the samples could recover the test as accurately as possible,

$$\min_{1\le k\le K} \|\mathcal{K}_{\mathbf{x}} - \mathbb{K}\delta_k(\hat{\alpha})\|_2. \tag{15}$$

The flowchart of the proposed method is shown in Fig.3.

IV. EXPERIMENTS AND DISCUSSIONS

To verify the proposed method, several experiments are performed on MSTAR database, a gallery collected using a 10 GHz SAR sensor in spotlight mode for several vehicle targets. For each target, images are captured at various depressions over a full $0 \sim 359^\circ$ range of aspect view. The chip images are of around 128×128-pixel in size, and cropped to 80×80-pixel. For the covariance descriptor, three feature modes are considered, as shown in TABLE I, where $\mathbf{x}$ is the original image; $\mathbf{r}$ and $\mathbf{c}$ are the row and column indexes.

A. Preliminary verification

This subsection devotes to preliminary evaluation. Three targets, BMP2, BTR70, and T72 are used, among which BMP2 and T72 consist of multiple variants. The standards, SN_9563 (BMP2), SN_132 (T72), and SN_c71 (BTR70) taken at 17° depression are used for training, while the remaining, SN_9566, SN_c21, SN_812 and SN_s7, captured at 15° are used for testing. The number of aspect views available for training and testing are listed in TABLE II.

TABLE II
THE NUMBER OF ASPECT VIEWS AVAILABLE FOR TRAINING AND TESTING.

Depr.	BMP2	BTR70	T72	SUM
17°	**232** [SN_9563]	**233**	**232** [SN_132]	**698**
15°	196 [SN_9566] 196 [SN_c21]	233	195 [SN_812] 191 [SN_s7]	974

TABLE III
THE RECOGNITION RATES OF THE METHODS TO BE STUDIED.

	$\text{SRC}^{\mathbf{E}}$	$\text{kNN}^{\mathbf{R}}$	$\text{SVM}^{\mathbf{R}}$	$\text{SRC}^{\mathbf{R}}$	$\text{KSVM}^{\mathbf{R}}$	KLSF
Mode1	0.7721	0.8789	0.9281	0.9446	0.9414	0.9559
Mode2	0.7700	0.8819	0.9322	0.9466	0.9435	0.9569
Mode3	0.7618	0.8768	0.9333	0.9538	0.9456	0.9610

TABLE IV
THE NUMBER OF ASPECT VIEWS AVAILABLE FOR 10 VEHICLE TARGETS.

Depr.	BMP2	BTR70	T72	BTR60	2S1	BRDM	D7	T62	ZIL131	ZSU
17°	232	233	233	256	299	298	299	299	299	299
15°	392	196	386	195	274	274	274	273	274	274

TABLE III lists the recognition rates of the studied methods. The reference methods include SVM and nearest neighbor classifier (kNN). Two proposed schemes are notated as $\text{SRC}^{\mathbf{R}}$ (sparse representation of log-covariance) and KLSF (least square fitting in RKHS). $\text{SRC}^{\mathbf{E}}$ denotes the classification via sparse representation of covariance (not transformed into the vector space). $\text{SVM}^{\mathbf{R}}$, $\text{KSVM}^{\mathbf{R}}$(with Gaussian RBF kernel), $\text{kNN}^{\mathbf{R}}$, and $\text{SRC}^{\mathbf{R}}$ are based on the covariance descriptor (6).

From TABLE III, some conclusions can be reached. First, the accuracies of $\text{SRC}^{\mathbf{E}}$ with mode1, mode2, and mode3 are 0.7721, 0.7700, and 0.7618, vastly inferior to all of the other methods. This results verify that the covariance matrix lies on the Riemannian manifold, and it is unreasonable to be directly applied to the standard learning technique. The shortage can be covered by mapping the point into the vector space by logarithm operator, because point on the manifold has a neighborhood for which there is a bijection mapping the neighborhood to the vector space. Second, the recognition rates of $\text{SRC}^{\mathbf{R}}$ with mode1, mode2, and mode3 are 0.9446, 0.9466, and 0.9538, 1.71% and 6.91% better in average than $\text{SVM}^{\mathbf{R}}$ and $\text{kNN}^{\mathbf{R}}$. The results demonstrate that sparse representation technique could predict the class membership more accurately than $\text{SVM}^{\mathbf{R}}$ and $\text{kNN}^{\mathbf{R}}$. Third, the recognition rates of KLSF on mode1, mode2, and mode3 are 0.9559, 0.9569, and 0.9610, 0.96% and 1.44% better in average than $\text{SRC}^{\mathbf{R}}$ and $\text{KSVM}^{\mathbf{R}}$. The results prove that it may result in some information loss by mapping the point on the manifold into the vector space by logarithm operator, because the mapping makes sense only around a neighborhood of the point.

B. Comparison with the other methods

A more challenging problem, 10-target recognition, is considered subsequently. The details on aspect views for training and testing can be found in TABLE IV. To observe the individual accuracies, the confusion matrices are shown in Fig. 4. kNN, SVM, KSVM, and SRC are based on intensity with the dimension reduced by uniform down-sampling, while KLSF, $\text{SVM}^{\mathbf{R}}$ and $\text{SRC}^{\mathbf{R}}$ are performed on the **mode3** covariance descriptor. MSRC denotes our preceding work [7]. From Fig 4, several generic conclusions can be drawn.

First, the accuracies of BMP2 and T72 are much lower than the others. Their accuracies are varied from 0.7551 to 0.8826, and from 0.5362 to 0.8212. This is because the configuration as well as the depression are significantly different between the images for training and these for testing. Second, the monogenic signal based methods significantly outperform the intensity based ones. The recognition rates of SVM$^{\mathbf{R}}$, MSRC and SRC$^{\mathbf{R}}$ are 0.8994, 0.9292, and 0.9385, compared to 0.8741 for kNN, 0.8588 for SVM, and 0.9125 for SRC. The results prove that the monogenic signal could characterize SAR image effectively. Third, MSRC performs much better than SRC, and worse than SRC$^{\mathbf{R}}$. This is because the information in monogenic signal is combined by simple concatenation. Thus, it is incapable to exploit the inter-correlation of the components, while SRC$^{\mathbf{R}}$ circumvents the problem by computing a covariance descriptor, whose elements are the correlation of the components. Fourth, the recognition rate of KLSR is 0.9488, 2.57%, 1.03%, 1.96%, 3.63%, 4.94%, 9%, 7.47% better than KSVM, SRC$^{\mathbf{R}}$, MSRC, SRC, SVM$^{\mathbf{R}}$, SVM, and kNN. The results show that the logDet divergence metric could accurately measure the similarity between a pair of covariance matrices.

V. Conclusion

This paper proposes a novel technique, the covariance descriptor, to combine the multiple components of the monogenic signal. Then, two classification schemes are presented to predict the class identity. The first maps the covariance matrix into the vector space by the logarithm operator, followed by SRC to make decision. The other embeds the Riemannian manifolds into the RKHS, and recover the test by least square fitting. Several experiments on MSTAR database demonstrate that the covariance descriptor could exploit the inter-correlation of the monogenic components, and hence combine various information in monogenic signal. Moreover, it is discriminative, and hence achieves significant improvement in accuracy than the standard ones.

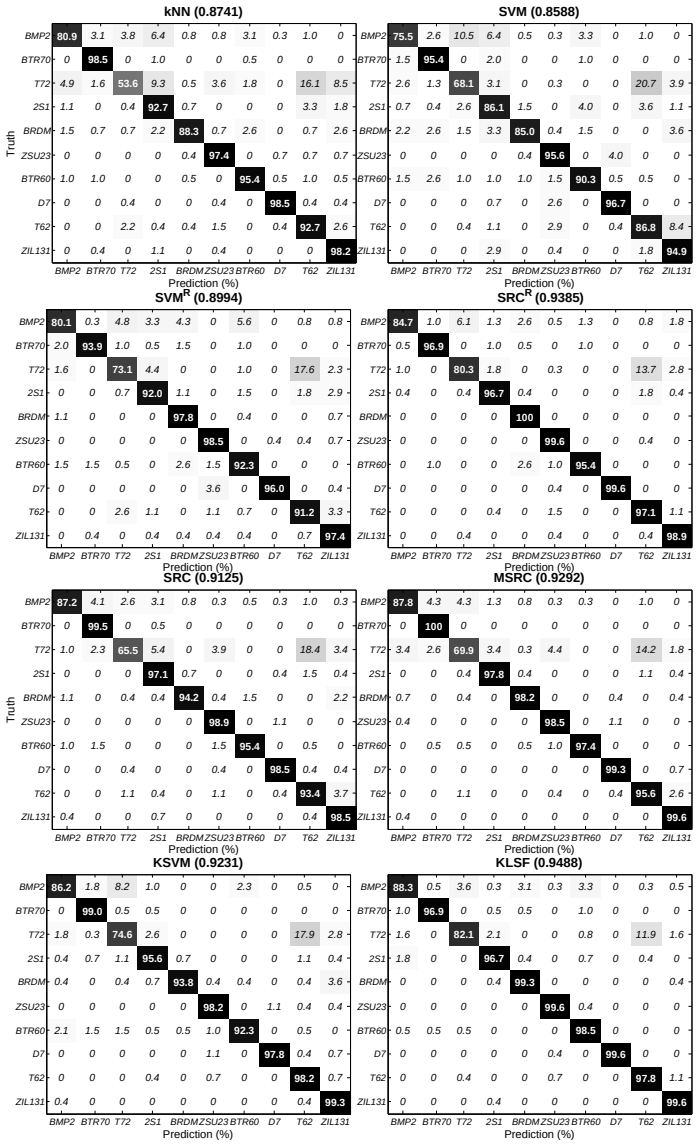

Fig. 4. The confusion matrices and overall accuracy of the methods to be studied.

References

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

[2] J. 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.

[3] M. D. DeVore and J. A. 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.

[4] J. Thiagarajan, N. Karthikeyan, K. Peter, and et al., "Sparse representation for automatic target classification in SAR images," in *Int'l Sym. Communcitaion, Control and Signal Processing*, 2010, pp. 1–4.

[5] 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, pp. 1–13, Aug. 2014.

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

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

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

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

[10] 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.

[11] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imag. Vis.*, vol. 25, pp. 127–154, Jul. 2006.

[12] V. Arsigny, P. Pennec, and X. Ayache, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magn. Reson. Med.*, vol. 56, no. 2, pp. 411–421, Aug. 2006.

[13] A. Cherian, S. Sra, A. Banerjee, and et al., "Jensen-Bregman logdet divergence with application to efficient similarity search for covariance matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2162–2175, Sep. 2013.

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

[15] S. Sra, "A new metric on the manifold of kernel matrices with application to matrix geometric means," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012.

[16] J. Wang, J. Yang, K. Yu, and *et al.*, "Locality-constraint linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3360–3367.