Target Recognition via Information Aggregation Through Dempster-Shafer's Evidence Theory

Ganggang Dong and Gangyao Kuang, *Member, IEEE*

***Abstract*—In this paper, a novel classification via information aggregation through Dempster-Shafer's evidence theory (DS) has been presented to target recognition in SAR image. Though DS theory of evidence has been widely studied over the decades, less attention has been paid to the application for target recognition. To capture the characteristics of SAR image, this paper exploits a new multi-dimensional analytic signal named monogenic signal. Since the components of monogenic signal are of high-dimension, it is unrealistic to be directly used. To solve the problem, an intuitive idea is to derive a single feature by these components. However, this strategy usually results in some information loss. To boost the performance, this paper presents a classification framework via information aggregation. The monogenic components are individually fed into a recently developed algorithm, sparse representation-based classification, from which the residual with respect to each target class can be produced. Since the residual from a query sample reflects the distance to the manifold formed by the training samples of a certain class, it is reasonable to be used to define the probability mass. Then, the information provided by monogenic signal can be aggregated via Dempster's rule, and hence the inference can be reached.**

***Index Terms*—Dempster-Shafer theory of evidence, Sparse representation, the monogenic signal, SAR target recognition.**

I. Introduction

DUE to the capability to work under inclement weather and 24-hour a day, synthetic aperture radar (SAR) has been widely used. Automatic target recognition (ATR) is an essential topic for SAR image interpretation. The baseline SAR ATR comprises three separate stages [1]. First, a detector roughly locates the candidate targets in a SAR image. Next, a target-size matched filter is used to accurately locate the candidate targets; several textural discriminants are then used to eliminate the natural-clutter false alarms, followed by a classifier to classify the remaining detections.

At present, SAR ATR mainly relies on template matching strategy, model-based scheme, and feature-based method. Template matching strategy refers to generating the templates and comparing with the test exemplar. It usually suffers ubiquitous speckle and inaccurate prediction. Model-based method represents the pixels in SAR image via a statistical distribution and seeks which distribution could maximize the probability. Since only limited amount of samples are accessed, it easily produces inaccurate empirical distributions. Feature-based method is enslaved to the accuracy of preprocessing, *e.g.*, image segmentation, which is still an open problem.

Recently, a novel method, sparse representation-based classification (SRC), is presented [2], where the training samples are used to build a redundant dictionary to encode the query as a linear combination of them. The representation can be obtained via sparsity constraint. Due to these advantages, *e.g.*, no need for preprocessing, pose and parameter estimation, it has been introduced to SAR target recognition [3]. To improve the performance, a novel multi-dimensional analytic signal named monogenic signal [4] has been incorporated into sparse representation technique [5], [6]. To combine multiple kinds of information, a single feature descriptor is derived from the monogenic components. The resulting descriptor is inputted into sparse representation-based classification. This approach can be viewed as the data fusion in feature-level. Since various kinds of information are roughly combined, it usually produces some information loss. To boost the performance further, this paper presents a novel classification via information aggregation. At present, many approaches to information aggregation have been proposed, among which Dempster-Shafer's evidence theory [7] has gained popularity. In [8], the problem of classifying an unseen pattern on the basis of its nearest neighbors in a recorded dataset is solved from the viewpoint of DS theory. In [9], DS evidence theory is applied to unsupervised classification in multisource remote sensing, where both imprecision and uncertainty are modeled. In [10], images of the same site acquired by different sensors have been analyzed by combining the information available with Dempster's rule. In [11], the contextual information of multispectral image is aggregated for land cover classification through DS theory. In [12], a method for building detection by the fusion of first and last pulse laser scanner data and multi-spectral images is presented. In [13], an application of data-driven DS theory is presented to fuse multisensor data for land-cover feature extraction.

Inspired by the preceding works [11]–[13], a novel classification scheme via information aggregation through DS theory of evidence is presented for target recognition in SAR image. First, the monogenic signal is exploited to capture the characteristics of SAR image. The components of monogenic signal at various scales are independently input into sparse representation-based classifications. Specifically, the components of training samples are used to build an overcomplete dictionary to represent the counterpart of the query as a linear combination of them. Via sparse signal representation, the unique representation can be obtained, from which the reconstruction error with respect to each target class can be

This works was supported by the National Natural Science Foundation of China under Grant 61201338.

Dr. Ganggang Dong and Prof. Gangyao Kuang are with the College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China, 410073 (e-mail: dongganggang@nudt.edu.cn)

produced. Since the residual resulting from a new test sample reflects the distance to the manifold formed by the training samples of a certain class [3], [14], it is reasonable to be used to define the probability mass; hence the confidence of each classifier with respect to each proposition can be quantified. Finally, evidence from different classifications is combined using Dempster's rule of combination, and the inference can be reached accordingly. The proposed method is pictorially summarized in Fig. 1.

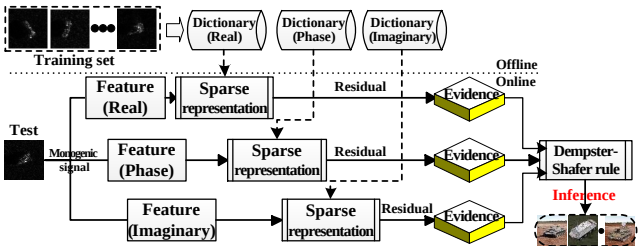

Fig. 1. The block diagram of proposed method. Three sparse classifiers are formed on the monogenic components. The residual of each classifier is used to generate a body of evidence. Via Dempster's rule of combination, the global evidence is obtained, from which the inference can be reached.

II. The Proposed Method

A. Dempster-Shafer Theory of Evidence

Denote by Θ the universal set and 2^Θ be its power set. A function $m : 2^\Theta \mapsto [0, 1]$ named basic probability assignment (BPA), is defined to quantify the candidate proposition,

$$m(\varnothing) = 0, \sum_{A \subseteq 2^\Theta} m(A) = 1. \tag{1}$$

A probability mass $m(A)$ provides the body of confidence that the proposition A is true. Any set $A \subset 2^\Theta$ that possesses nonzero BPA $m(A) > 0$ is called a focal element. In the formalism of evidence theory, imprecision of knowledge can be handled by assigning a non-zero probability mass to the union of two or more classes. Two parameters, support $Sup(A) = \sum_{B \subseteq A} m(B)$ and plausibility $Pls(A) = \sum_{A \cap B \neq \varnothing} m(B)$ can be defined for all $A \subset 2^\Theta$.

A peculiar property for Dempster-Shafer's evidence theory consists in the capability to combine multiple distinct sources. If express by m^1, m^2 two BPAs on the same power set 2^Θ, a joint BPA m can be obtained by Dempster's rule of combination,

$$m(A) = m^1 \oplus m^2(A) = \frac{\sum_{B \cap C = A} m^1(B) m^2(C)}{1 - K}, \tag{2}$$

for $K = \sum_{B \cap C = \varnothing} m^1(B) m^2(C)$. Rule (2) is commutative $m^1 \oplus m^2(A) = m^2 \oplus m^1(A)$ and associative. Thus, evidence from K distinct sources can be combined by repetitive application of (2), $m = m^1 \oplus m^2 \oplus \cdots \oplus m^K$. Once the combined evidence is generated, the decision can be made by seeking the proposition that provides the maximum amount of evidence, *e.g.*, maximum support, maximum plausibility, maximum support without overlapping of uncertainty intervals [12].

B. The Monogenic Signal

The monogenic signal is a 2-D extension of analytic signal [4]. It has been built around the Riesz transform. For signal $f(\mathbf{z}), \mathbf{z} = (x, y)^T$ and its Fourier spectra $F(\mathbf{u}), \mathbf{u} = (u, v)^T$, the expression of Riesz-transformed signal in the frequency domain is $F_R(\mathbf{u}) = \frac{i\mathbf{u}}{|\mathbf{u}|} F(\mathbf{u})$, while the corresponding one in the spatial domain is $f_R(\mathbf{z}) = -\frac{\mathbf{z}}{2\pi|\mathbf{z}|^3} * f(\mathbf{z}) = [f_R^x(\mathbf{z}), f_R^y(\mathbf{z})]^T$.

The combination of signal $f(\mathbf{z})$ and the Riesz transformed one forms a 2-D analytic signal named monogenic signal,

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

where i and j are the imaginary units. Here, f_R composes of two distinct imaginary units f_R^x and f_R^y, because (3) is meant in a Clifford algebra, rather than in the classical field of complex numbers. Then, it is natural to represent the signal by three components, the real part $Re\{f_M\}$, local phase $\varphi(\mathbf{z}) = atan2(f_R(\mathbf{z}), f(\mathbf{z}))$, and imaginary part $Im\{f_M\}$.

Since the practical signal is of finite length, it has broad spectra in the frequency domain, and hence needs infinite extension by a band-pass filter before Riesz transform. Here, log-Gabor filters are used due to the ability to capture the broad spectral information, as contended in the preceding works [5], [6]. Then, (3) can be rewritten as

$$f_M(\mathbf{z}) = (h(\mathbf{z}) * f(\mathbf{z})) - (i, j)(h(\mathbf{z}) * f_R(\mathbf{z})), \tag{4}$$

where $h(\cdot)$ denotes the log-Gabor kernel. Suppose S-scale log-Gabor filters are performed, the monogenic signal at different scales can be expressed as $f_M^1, f_M^2, \cdots, f_M^S$, and the corresponding components are

$$\big[\underbrace{Re_1, \varphi_1, Im_1}_{f_M^1}, \underbrace{Re_2, \varphi_2, Im_2}_{f_M^2}, \cdots, \underbrace{Re_S, \varphi_S, Im_S}_{f_M^S}\big]. \tag{5}$$

C. Information Aggregation Through DS Theory of Evidence

To apply the monogenic signal into classification system, the preceding works define a single feature descriptor by the components [5], [6]. However, this strategy usually results in some information loss. To boost the performance, this paper proposes a classification via information aggregation through DS theory of evidence. Unlike the preceding works, where the monogenic components are simply combined, three distinct feature descriptors are derived from real part, local phase, and imaginary part of monogenic signal at various scales. The monogenic components at different scales are first downsampled by a factor of ρ, followed by normalization and concatenation to form a feature vector

$$\begin{aligned} \chi_R &= [vec_\rho(Re_1), vec_\rho(Re_2), \cdots, vec_\rho(Re_S)]^T \\ \chi_\varphi &= [vec_\rho(Im_1), vec_\rho(Im_2), \cdots, vec_\rho(Im_S)]^T. \\ \chi_I &= [vec_\rho(\varphi_1), vec_\rho(\varphi_2), \cdots, vec_\rho(\varphi_S)]^T \end{aligned} \tag{6}$$

where $vec(\cdot)$ reshapes a matrix to be a single vector. Then, three sparse representation-based classifications can be formed by features (6). Specifically, the features of training samples are used to build a redundant dictionary to represent the counterpart of the query as a linear combination of them.

Given n training samples $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ from c distinct classes. The feature descriptors of training samples, obtained

using (6), are used to build a redundant dictionary, for example $\mathcal{D}_R = [\chi_R(\mathbf{x}_1); \chi_R(\mathbf{x}_2); ...; \chi_R(\mathbf{x}_n)]$. For a new test sample $\mathbf{y}$, the counterpart $\chi_R(\mathbf{y})$ can be represented as a linear combination of atoms in dictionary $\mathcal{D}_R$,

$$\chi_R(\mathbf{y}) = \chi_R(\mathbf{x}_1)\alpha_1 + \chi_R(\mathbf{x}_2)\alpha_2 + \cdots + \chi_R(\mathbf{x}_n)\alpha_n, \quad (7)$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n] \in \mathbb{R}^n$ is the representation.

Considering the underdetermined system ($card(\chi_R) < n$), a great many solutions to (7) can be provided. The popularly used strategy is to find the most compact representation via sparsity constraint [2],

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|\chi_R(\mathbf{y}) - \mathcal{D}_R\alpha\|_2 < \epsilon, \quad (8)$$

where ϵ is the allowed error tolerance. By the standard optimization method, the optimal representation $\hat{\alpha}$ can be obtained, from which the residual with respect to each class can be computed,

$$\{e_k^R\}_{k=1,...,c} = \|\chi_R(\mathbf{y}) - \mathcal{D}_R\delta_k(\hat{\alpha})\|_2^2, \quad (9)$$

where $\delta_k(\cdot)$ preserves the entries associated with the k-th class and sets the remaining to be zeros. Similarly, another two classification systems can be built on χ_φ and χ_I, and hence the residual energy $\mathbf{e}^\varphi$ and $\mathbf{e}^I$ can be generated, accordingly.

In (8), the recognition problem is cast as one of classifying among multiple linear regression models, while the training and the query samples play the role of regressors and response. Since the residual from a new test sample reflects the distance to the manifold formed by the training samples of a certain class [3], it is used to define the probability mass in this paper.

Denote by $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_c$ the c target class to which the query belongs. Then, the framework of discernment is $\{\mathcal{C}_1\}$, $\{\mathcal{C}_2\}, ..., \{\mathcal{C}_1 \cup \mathcal{C}_2\}, ...$, where $\{\mathcal{C}_1 \cup \mathcal{C}_2\}$ denotes the union of $\mathcal{C}_1$ and $\mathcal{C}_2$. Since the residual energy reflects the signal-to-manifold distance, it is reasonable to derive the probability mass from the residual energy.

In the framework of SRC, the query is assumed to belong the class with the minimal residual energy. The smaller the minimal residual is, the more believable the decision is. Given a residual vector $\mathbf{e} = [e_1, e_2, ..., e_c] \in \mathbb{R}^c$, we assume the hypothesis $\mathbf{y} \in \mathcal{C}_i$ is very unlikely if $\mathbf{e}_i > \mathcal{T}_0$ for $i = 1, ..., c$, where $\mathcal{T}_0$ is a threshold to judge whether a candidate proposition is reliable. Thus, it does not make sense for assigning the nonzero probability mass to $\{\mathcal{C}_i\}$, *i.e.*, $m(\{\mathcal{C}_i\}) = 0$. Vice versa, we assume the hypothesis $\mathbf{y} \in \{\mathcal{C}_i\}$ is likely if $\mathbf{e}_i < \mathcal{T}_0$. In other words, this paper assigns the nonzero probability mass only to those target class with a tolerant residual energy. Next, this paper will pay more attention to the generation of nonzero probability mass when $\mathbf{e}_i < \mathcal{T}_0$.

For the residual with respect to the i-th target class e_i, the smaller the residual, the more likely the hypothesis that the query belongs to the i-th target class $\mathbf{y} \in \mathcal{C}_i$. Thus, it is modeled by a big probability mass. On the contrary, the bigger the residual $\mathbf{e}_i$, the less likely the hypothesis $\mathbf{y} \in \mathcal{C}_i$, and hence modeled by a small probability mass. According to the thoughts above, this paper presents the BPA function as

$$m(\{\mathcal{C}_i\}) = \begin{cases} C_i \cdot (1 - \mathbf{e}_i) & \mathbf{e}_i \le \mathcal{T}_1 \\ C_i \cdot g(\mathbf{e}_i) & \mathcal{T}_1 < \mathbf{e}_i < \mathcal{T}_0 \\ 0 & \mathbf{e}_i \ge \mathcal{T}_0 \end{cases} \quad (10)$$

where C_i is a weight coefficient; $\mathcal{T}_0$ and $\mathcal{T}_1$ are two critical thresholds ($\mathcal{T}_0 > \mathcal{T}_1$); $g(\cdot)$ is an auxiliary function. Here, the weight ($0 < C_i < 1$) is used to introduce the uncertain states, *i.e.*, union of classes. Via (10), the simple hypotheses (*i.e.*, singletons) can be quantified. The remaining key problem is how to model the compound hypotheses. In this paper, we assume that the nonzero probability mass is assigned only to those classes whose residuals are less than a threshold, otherwise it contributes to the uncertain state. Thus, the remaining probability mass $1 - \sum_{i=1}^{c} m(\{\mathcal{C}_i\})$ is assigned to the uncertain state $m(\{\mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_c\})$. Then, the actual framework of discernment is $\{\mathcal{C}_1\}, ..., \{\mathcal{C}_c\}, \{\mathcal{C}_1 \cup \mathcal{C}_2 \cup ... \cup \mathcal{C}_c\}$. The support and plausibility can be obtained, accordingly. To satisfy the constraint terms shown in (1), this paper derives the weights C_i from residual, $C_i = \frac{1/e_i}{sum(1/\tilde{\mathbf{e}})}$, where $\tilde{\mathbf{e}}$ denotes the entries of $\mathbf{e}$ whose values are less then $\mathcal{T}_0$.

In (10), $g(e_i)$ is a decreasing function of e_i. Moreover, it is varied in a range of $g(\cdot) \in (0, 1)$. Obviously, there is an infinitely large number of decreasing function verifying the terms above, and it is very difficult to find any a prior argument in favor of one particular function or another. Inspired by the preceding works [12], this paper suggests to choose it as

$$g(\mathbf{e}_i) = P_1 + (P_2 - P_1)\left[3(\frac{\mathcal{T}_0 - \mathbf{e}_i}{\mathcal{T}_0 - \mathcal{T}_1})^2 - 2(\frac{\mathcal{T}_0 - \mathbf{e}_i}{\mathcal{T}_0 - \mathcal{T}_1})^3\right] \quad (11)$$

where P_1 and P_2 are two probabilities committed to the critical residuals ($P_1 < P_2$). The proposed function is pictorially shown in Fig. 2.

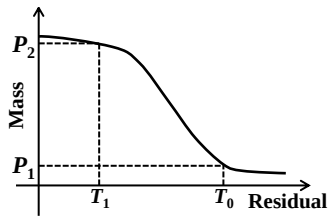

Fig. 2. The probability mass function.

Given the function (11), the BPAs resulting from various classifications can be fused using Dempster-Shafer's rule of combination, and the inference can be reached accordingly. Since the uncertainty intervals $[Sup(\{\mathcal{C}_i\}), Pls(\{\mathcal{C}_i\})]$ are same for all singleton hypotheses ($i = 1, ..., c$), the decision rules shown in Section II-A will produce the same rank. The procedure of proposed scheme can be shown by TABLE I.

TABLE I
THE PROCEDURE OF EVIDENTIAL REASONING ($\Theta = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_c$)

Class	$\mathcal{C}_1$	$\mathcal{C}_2$	$\cdots$	$\mathcal{C}_c$	Θ
Real	$m^R(\mathcal{C}_1)$	$m^R(\mathcal{C}_2)$	$\cdots$	$m^R(\mathcal{C}_c)$	$m^R(\Theta)$
Phase	$m^\varphi(\mathcal{C}_1)$	$m^\varphi(\mathcal{C}_2)$	$\cdots$	$m^\varphi(\mathcal{C}_c)$	$m^\varphi(\Theta)$
Imaginary	$m^I(\mathcal{C}_1)$	$m^I(\mathcal{C}_2)$	$\cdots$	$m^I(\mathcal{C}_c)$	$m^I(\Theta)$
Combined	$m(\mathcal{C}_1)$	$m(\mathcal{C}_2)$	$\cdots$	$m(\mathcal{C}_c)$	—

$m(\mathcal{C}_1) = \frac{(m^R(\mathcal{C}_1)+m^R(\Theta))*(m^\varphi(\mathcal{C}_1)+m^\varphi(\Theta))*(m^I(\mathcal{C}_1)+m^I(\Theta))}{1-K}$
$K = \sum_{\mathcal{C}_i \cap \mathcal{C}_j = \varnothing} m(\mathcal{C}_i)m(\mathcal{C}_j)$ is the conflict defined in (2).

III. Experiments and Discussions

This section evaluates the proposed method on MSTAR database, a gallery collected using a SAR sensor in spotlight mode. For each target, images are captured at various depressions over $0 \sim 359°$ range of aspect view. The depression refers to the angle between the line of the sight from the radar to the target and the horizontal plane at the radar. The images are of around 128×128 pixels in size, and cropped to 80×80 pixels region of interest. The problem (8) is numerically solved by least angle regression [15] due to the computational efficiency. The baseline methods include the nearest neighbor classifier (kNN), SVM (with nonlinear kernel), SRC [2], [3], Bayesian decision [16], and our preceding work [5] (MSRC).

A. Illustrative Examples

To demonstrate the proposed method, several experiments are performed. There is a total of four target classes (BMP2, T72, BTR60, and T62), whose notations are $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ and $\mathcal{C}_4$. The experimental results are listed in TABLE II, where the classification via Dempster-Shafer theory is compared with the one via Bayesian decision ($Bayes$). To assure fair comparison, the probabilities for Bayesian decision are generated from the residual energy too, $P(\mathcal{C}_i) = \frac{1/e_i}{\sum_j 1/e_j}$. Following the preceding works [16], sum rule is used to draw the inference.

TABLE II
Classification Examples

	Bayesian				*Dempster-Shafer*				
	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$m(\mathcal{C}_1)$	$m(\mathcal{C}_2)$	$m(\mathcal{C}_3)$	$m(\mathcal{C}_4)$	$m(\Theta)$
R	0.0920	0.3716	0.1659	0.3705	0.5820	0	0.1458	0	0.2722
φ	0.0404	0.3553	0.2489	0.3553	0.9596	0	0	0	0.0404
I	0.1022	0.3618	0.1188	0.4171	0.4679	0	0.3607	0	0.1714
	0.5375	0.1158	0.2363	0.1103	0.5461	0.0019	0.0090	0.0019	—
R	0.0263	0.3332	0.3059	0.3346	0.9737	0	0	0	0.0263
φ	0.3060	0.1945	0.3088	0.1907	0	0.0985	0	0.1126	0.7889
I	0.3672	0.2103	0.2971	0.1253	0	0.0446	0	0.4613	0.4941
	0.2638	0.2501	0.2024	0.2837	0.3898	0.0126	0.0102	0.0226	—
R	0.1877	0.3475	0.3777	0.0872	0.0771	0	0	0.6244	0.2985
φ	0.3383	0.1226	0.3289	0.2101	0	0.4768	0	0.0443	0.4789
I	0.2626	0.1172	0.4011	0.2190	0	0.5149	0	0.0324	0.4528
	0.2183	0.2931	0.1554	0.3333	0.0814	0.2761	0.0647	0.2342	—
R	0.2998	0.2340	0.2830	0.1832	0	0	0	0.2741	0.7259
φ	0.2921	0.1373	0.3170	0.2537	0	0.6461	0	0	0.3539
I	0.2008	0.2285	0.4137	0.1570	0.0517	0.0223	0	0.1958	0.7302
	0.2253	0.2978	0.1762	0.3007	0.2009	0.5463	0.1876	0.3277	—

In the first example, the test sample comes from $\mathcal{C}_1$. Both two methods reach the correct decisions. Bayesian produces the maximum probability to $\mathcal{C}_1$, while Dempster-Shafer generates the maximum support (or plausibility) to $\mathcal{C}_1$ too. However, the confidence level (difference between the probability for each class) for $Bayes$ is not as high as DS. In the second example, the test sample is from $\mathcal{C}_1$. Bayesian assigns the maximum probability (0.2837) to $\mathcal{C}_4$, while DS produces the maximum support (0.3898) to $\mathcal{C}_1$. In the third and fourth examples, the query samples are from $\mathcal{C}_2$. The classification via Bayesian theory provides the error inferences, $\mathcal{C}_4$, while the proposed method reaches the correct decisions, $\mathcal{C}_2$.

TABLE III
The Number of Aspect Views Available for Different Targets

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

TABLE IV
Recognition Rates for Configuration Variation.

Algorithm	kNN	SVM	SRC	MSRC	*Bayes*	*DS*
Accuracy	0.8339	0.8331	0.8708	0.8732	0.8804	0.8925

B. Configuration variation

This subsection considers target recognition under different configurations. The configuration refers to physical difference, *e.g.*, auxiliary fuel barrels, side skits, smoke grenade launchers, as an example shown in Fig. 3. Four vehicle targets are used, among which BMP2 and T72 comprise several variants with structural modifications. The standards (SN_9563 for BMP2, SN_132 for T72) taken at a 17° depression are used for training, while the remaining collected at a 15° depression are used for testing. The number of aspect views available for different targets are listed in TABLE III. The parameters in (10) are set as follows, $\mathcal{T}_0 = 0.24, \mathcal{T}_1 = 0.08$ and $P_1 = 0.1, P_2 = 0.9$.

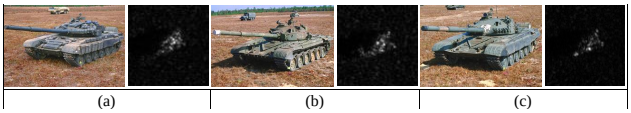

Fig. 3. Illustration of configuration. (a), (b), and (c) show three variants of T72 tank (optical image and SAR image with 18° aspect view).

TABLE IV gives the recognition rates of the methods to be studied. The classification via information aggregation through Dempster-Shafer theory of evidence is notated as DS, while the one via Bayesian is marked by $Bayes$. From TABLE IV, we are capable to come the conclusions as follows.

- The recognition rates for MSRC, $Bayes$, and DS are 0.8732, 0.8804, and 0.8925, outperforming the remaining algorithms, SRC, SVM, and kNN. The results prove that the performance in accuracy can be significantly improved with the monogenic signal representation.
- Two classifications via information aggregation, $Bayes$ and DS, achieve the recognition rates of 0.8804 and 0.8925, 0.72%, and 1.93% better than the preceding work, MSRC. The results corroborate that the preceding algorithm usually result in some information loss.
- By aggregating various kinds of information with Dempster-Shafer theory of evidence, the performance for DS significantly exceeds the ones of $Bayes$.

C. Depression Variation

This subsection deals with target recognition under different depressions. Three targets, 2S1, BRDM2, and ZSU23/4 are used, and two of which (BRDM2 and ZSU23/4) have the articulated variants. The articulated refers to physical change to the standard such as open hatch or rotated gun turret.

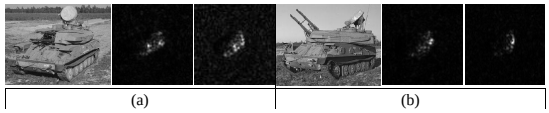

Fig. 4. Illustration of articulation; (a) and (b) show the standard and the articulate of ZSU23/4 taken at 30° and 45° depression with aspect view of 25.9°.

Examples of the standards and the articulated variants for ZSU23/4 are shown in Fig. 4. Images collected at a 17° depression are used to train the algorithm, while the ones taken at 30° and 45° depressions are utilized for testing. The details can be found in TABLE V. The parameters to generate the probability mass are set as, $\mathcal{T}_0 = 0.33, \mathcal{T}_1 = 0.1$ and $P_1 = 0.1, P_2 = 0.9$. The overall recognition rates as well as the confusion matrices are shown in Fig. 5.

TABLE V
THE NUMBER OF ASPECT VIEWS FOR DEPRESSION VARIATION.

Depression		2S1($\mathcal{C}_1$)	BRDM2($\mathcal{C}_2$)	ZSU23/4($\mathcal{C}_3$)	SUM
Training	(17°)	**299**	**298**	**299**	**896**
Testing	(30°)	288	287(133)	288(118)	1114
	(45°)	303	303(120)	303(119)	1148

30 °	**SVM** (0.9443)			**SRC** (0.9614)			**MSRC**(0.9829)			**Bayes**(0.9829)			**DS** (0.9838)		
Truth	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
C_1	263	25	0	280	8	0	282	5	1	287	1	0	287	1	0
C_2	8	404	8	13	398	9	1	413	6	0	408	12	0	409	11
C_3	17	4	385	6	7	393	1	5	400	2	4	400	2	4	400
45 °	**SVM** (0.5261)			**SRC** (0.5366)			**MSRC**(0.6394)			**Bayes**(0.7456)			**DS** (0.7613)		
Truth	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
C_1	236	61	6	107	195	1	262	25	16	221	69	13	224	69	10
C_2	179	239	5	33	383	7	112	195	116	23	306	94	19	318	86
C_3	283	10	129	55	241	126	78	67	277	26	77	319	24	75	323

Prediction

Fig. 5. The experimental results for depression variation.

The first scenario considers a minor difference on depression. The algorithms are trained at an operating condition of 17° depression, and tested at an operating condition of 30° one, *i.e.*, a variation of 13° from 17° to 30° exists between the training and the query. As can be seen, the performance is satisfied on the whole. The recognition rate for DS is 0.9838, compared to 0.9829 for $Bayes$, 0.9829 for MSRC, 0.9614 for SRC, and 0.9414 for SVM.

To further verify the algorithms under nonliteral experimental setup, the second scenario deals with a major difference on depression. The methods are further tested at an operating condition of 45° depression, *i.e.*, a change of 28° from 17° to 45° exists. The recognition accuracies are sharply degraded. It is easy to confuse these targets due to the abrupt variation in the signatures of the same target from different depressions. The overall accuracy for SVM declines from 0.9443 to 0.5261, while the one for SRC drops from 0.9614 to 0.5366. $Bayes$ obtains a drop of 23.72% for accuracy. However, the proposed method achieve the rate of 0.7613, 1.57%, 12.19% and 22.47% better than the competitors, $Bayes$, MSRC, and SRC.

IV. CONCLUSION

In this paper, a new classification via information aggregation through Dempster-Shafer theory of evidence is presented for target recognition in SAR image. The monogenic signal is exploited to capture the characteristics of SAR image. The residuals derived from sparse representation-based classifier are used to produce the probability mass. Dempster-Shafer theory of evidence is employed to aggregate multiple kinds of information. Extensive experiments, including variations in pose, configuration, and depression, are performed on MSTAR database. The results prove that the proposed method is much more effective to deal with the extended operating conditions, *e.g.*, different configurations and depressions, than all the reference methods.

REFERENCES

[1] L. Novak, G. Owirka, and C. Netishen, "Performance of a high-resolution polarimetric SAR automatic target recognition system," *Lincoln Lab J.*, vol. 6, no. 1, pp. 11–23, 1993.
[2] J. Wright, A. Yang, A. Ganesh, S. Satstry, and Y. Ma., "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[3] 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.
[4] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, 2001.
[5] 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.
[6] G. Dong and G. Kuang, "Target recognition in SAR images via classification on Riemannian manifolds," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, 2015.
[7] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[8] T. Denceux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.
[9] S. Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, p. 10181031, Jul. 1997.
[10] A. Sarkar, A. Banerjee, N. Banerjee, and *et al.*, "Landcover classification in MRF context using Dempster-Shafer fusion for multisensor imagery," *IEEE Trans. Image Process.*, vol. 14, no. 5, pp. 634–644, May 2005.
[11] A. Laha, N. Pal, and J. Das, "Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1633–1641, Jun. 2006.
[12] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection," *Information Fusion*, vol. 6, pp. 283–300, 2005.
[13] V. Saeidi, B. Pradhan, M. O. Idrees, and Z. A. Latif, "Fusion of airborne LiDAR with multispectral spot 5 image for enhancement of feature extraction using Dempster-Shafer theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6017–6028, Oct. 2014.
[14] V. Berisha, N. Shah, D. Waagen, and *et al.*, "Sparse manifold learning with applications to SAR image classification," in *Proc. IEEE Conf. Acoustic, Sonar and Signal Processing (ICASSP)*, 2007, pp. 1089–1092.
[15] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
[16] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.