# A Soft Decision Rule for Sparse Signal Modeling via Dempster-Shafer Evidential Reasoning

Ganggang Dong, Siqian Zhang, and Gangyao Kuang, *Member, IEEE*

***Abstract*—Recently, the problem of recovering sparse linear representation of a query in terms of a redundant dictionary has received great interest. The query sample is represented as a linear combination of the atoms of dictionary. The unique representation is obtained via sparsity constraint, and the decision is made in terms of the reconstruction error. The decision rule for sparse signal modeling can be viewed as a typical application of Bayesian estimation, where the likelihood function is inversely proportional to the reconstruction error. Different from the conventional algorithm, where the decision is directly made according to the overall reconstruction error committed to each class, this paper proposes a soft decision rule via Dempster-Shafer theory of evidence. To model the imprecision on uncertainty measurement, we introduce the sample-wise ambiguity and the class-wise ambiguity during the quantification of probability mass. Each sample that participates in the recovery of the query is considered as an item of evidence that supports certain hypothesis in regard to the class membership of the query. The amount of evidence is quantified by a function of the distance between the query and the weighted training sample, where the weights result from sparse representation coefficient. Then, various pieces of evidence deriving from the candidate samples are pooled by means of Dempster's rule of combination, from which a soft decision can be reached.**

***Index Terms*—Dempster-Shafer theory of evidence, Sparse representation, soft decision, classification, SAR target recognition.**

## I. Introduction

SAR could complement the photographic and other optical imaging sensors for the capability to work under inclement weather and 24-hour a day, thus it has been widely used in many fields. Automatic target recognition is a basic research topic for SAR image interpretation. It is an application of signal processing technique to recognize the unknown targets. The current approaches to SAR ATR include template matching, statistical model-based scheme, and feature-based strategy. The performance of template matching is significantly affected by the variability of target configuration, the revetments, and the resolution of the image. Statistical model-based schemes easily fail when there does not exist strong statistical relationship between the training and the query. Feature-based methods are enslaved to the multiple preprocessing.

Recently, with the development of the theory of compressive sensing, a great resurgence of sparse signal modeling over a redundant dictionary has been witnessed. The representative application includes sparse representation-based classification (SRC) [1], where the training samples are used to build an overcomplete dictionary to represent the query as a linear combination of them. The unique representation is obtained via sparsity constraint, and the decision is reached by evaluating which class of samples could produce the minimal reconstruction error. Due to its promising performance, sparse signal modeling has also been introduced into SAR target recognition. Ref [2] presents a sparse representation approach for classifying different targets in SAR images. It explains the applicability of sparse signal modeling in SAR target recognition from the perspective of class manifolds. Compared to the conventional methods, it does not require explicit pose estimation and any preprocessing. Then, Ref [3] proposes a joint sparse representation model for multi-view target recognition. Ref [4] jointly considers the monogenic signal and sparse representation in a unifying framework for SAR target recognition. In the preceding works, the decision is made by evaluating which class of samples could produce the minimal reconstruction error. This is a typical application of Bayesian inference. If we define the likelihood function to be inversely proportional to the reconstruction error, the decision rule is actually maximum likelihood estimation (or maximum a posterior probability), and hence is a hard decision. Different from the above works, where the decision is directly made by the overall reconstruction error associated with each class, this letter presents a soft decision for sparse signal modeling via Dempster-Shafer theory of evidence [5]–[8]. Then, it is capable to model the imprecision on uncertainty measurement. Over the past decades, the evidential reasoning by Dempster-Shafer theory has gained popularity. Ref [9] addresses the problem of classifying a query pattern on the basis of its nearest neighbors in a recorded dataset via Dempster-Shafer theory. Ref [10] applies Dempster-Shafer theory to unsupervised classification in multisource remote sensing. Ref [11] achieves SAR image classification by combining several structure detectors via Dempster-Shafer theory. Ref [12] recommends a method for the detection of buildings in densely built-up urban areas by fusing the first and last pulse laser scanner data and multi-spectral images via Dempster-Shafer theory. Ref [13] proposes a dynamic evidential reasoning system for the sequential fusion of multitemporal images by Dempster-Shafer theory. Ref [14] adopts Dempster-Shafer theory to change detection, where two change indicators, stereo imagery and digital surface models generated with stereo matching methodology, are combined. Ref [15] aggregates multiple kinds of information for target recognition via Dempster-Shafer theory.

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

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

Inspired by these works, this paper presents a soft decision rule for sparse signal modeling via evidential reasoning. Different from the hard decision rule, the singleton hypotheses as well as the uncertainty states, including the sample-wise uncertainty and the class-wise uncertainty, are considered simultaneously. By introducing the ambiguous propositions, the imprecision about uncertainty measurement can be modeled. Specifically, the sample that participates in the representation of the query is considered as an item of evidence that supports certain hypotheses in regard to the class membership of the query. The amount of evidence is quantified by a function of the reconstruction error resulting from the candidate sample. Then, various pieces of evidence derived from the candidate samples are pooled via Dempster's rule of combination. The evidence aggregation is realized by two sequential steps, intra-class combination and inter-class ones. The former is devoted to the combination of evidence from the candidate samples of the same class, while the latter contributes to the combination of evidence committed to certain target class and the ambiguity. A soft decision is reached by maximum belief (or maximum plausibility) [10], [16]. The proposed method is pictorially shown in Fig. 1.

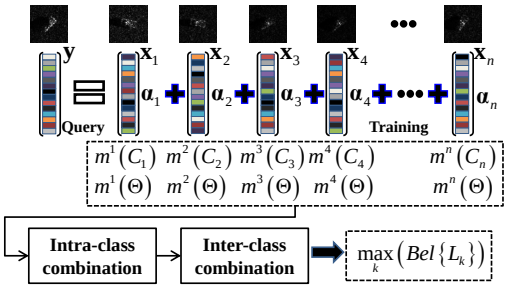


Fig. 1. The mechanism of soft decision rule.

## II. The Proposed Method

### A. Dempster-Shafer's Theory of Evidence

This paper considers a recognition problem where the query is to be classified into $c$ mutually exclusive target classes. Denote by $\Theta$ the universal set and $2^\Theta$ is its power set. It contains both the single classes and all the possible unions. A basic probability assignment (BPA) $m(A)$, is assigned to every proposition $A \in 2^\Theta$ such that,

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

with $\varnothing$ denoting the empty set. BPA $m(A)$ provides the body of confidence that proposition $A$ is true. Any set $A \subset 2^\Theta$ with nonzero BPA $m(A) > 0$ is called a focal element. In the formalism of evidence theory, imprecision about uncertainty measurement can be handled by assigning a non-zero probability mass to the union of classes. The belief and plausibility functions, derived from $m(\cdot)$, are respectively defined,

$$\begin{aligned} Bel(A) &= \sum_{B \subseteq A} m(B) \\ Pl(A) &= \sum_{A \cap B \neq \varnothing} m(B) = 1 - Bel(\bar{A}) \end{aligned} \tag{2}$$

They can be referred to as lower and upper probability, and hence have the properties, $Bel(A) \leq Pl(A)$ and $Pl(A) = 1 - Bel(\bar{A})$ for $\bar{A}$ noting the complementary.

A peculiar property for the theory of evidence is the ability to combine the evidence from the distinct sources. If express by $m^1, m^2$ two BPAs on the same power set $2^\Theta$, a joint BPA $m$ can be obtained by Dempster's rule of combination,

$$m(A) = m^1 \oplus m^2(A) = \frac{\sum_{B \cap C = A} m^1(B) m^2(C)}{1 - Conf}, \tag{3}$$

for $Conf = \sum_{B \cap C = \varnothing} m^1(B) m^2(C)$. Since (3) is commutative and associative, evidence from $K$ distinct sources can be combined by repetitive application, $m = m^1 \oplus m^2 \oplus \cdots \oplus m^K$. Having computed BPA, belief, and plausibility for each hypothesis, the decision can be made by evaluating which hypothesis could provide the maximum amount of evidence, *e.g.*, maximum of plausibility, maximum of belief, maximum of belief without overlapping of belief intervals [10].

### B. Sparse Representation-based Classification (SRC)

Sparse representation aims to recovering the query as a linear combination of a few atoms selected from a redundant dictionary formed by the training samples themselves. Given sufficient samples for the $k$-th class, $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, ..., \mathbf{x}_{k,n_k}]$, any new query sample $\mathbf{y} \in \mathbb{R}^m$ from the same class can be represented using the training samples of the $k$-th class, $\mathbf{y} = \mathbf{x}_{k,1}\alpha_{k,1} + \mathbf{x}_{k,2}\alpha_{k,2} + \cdots + \mathbf{x}_{k,n_k}\alpha_{k,n_k}$, where $\alpha_{k,j}$ denotes the weighted contribution of the $j$-th sample.

Since the class membership of the query is unknown initially, the representation of $\mathbf{y}$ can be rewritten using $n$ training samples from all $K$ classes, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K] \in \mathbb{R}^{m \times n}$,

$$\mathbf{y} = \mathbf{x}_{1,1}\alpha_{1,1} + \mathbf{x}_{1,2}\alpha_{1,2} + \cdots + \mathbf{x}_{K,n_K}\alpha_{K,n_K} \tag{4}$$

where $\alpha = [\alpha_{1,1}, ..., \alpha_{K,n_K}] \in \mathbb{R}^n$ is the representation.

Considering the underdetermined system ($m < n$), (4) is ill-conditioned, and hence does not have unique solution. The popularly used strategy is to find the most parsimonious representation via sparesity constraint,

$$\min_{\alpha} \|\alpha\|_1 \ \textit{subject to} \ \|\mathbf{y} - \mathbf{X}\alpha\|_2 < \varepsilon \tag{5}$$

where $\|\cdot\|_1$ sums the absolute value of entries. Since the objective function of (5) is convex, it can be solved by optimization skills. Having obtained the optimal representation $\hat{\alpha}$, the inference can be reached by evaluating which class of samples could result in the minimal reconstruction error,

$$\min_{1 \leq k \leq K} \{\mathbf{e}_{SRC,k} = \|\mathbf{y} - \mathbf{X}_k \hat{\alpha}_k\|_2^2\} \tag{6}$$

where $\mathbf{e}_{SRC,k}$ is the reconstruction error committed to the $k$-th target class via sparse signal modeling.

### C. A Soft Decision via Evidential Reasoning

In sparse signal modeling, the reconstruction error $\mathbf{e}_k$ reflects the distance between the query and the manifold formed by samples from the $k$-th class [2], and hence is used to make the decision. In the perspective of Bayesian inference, this decision rule can be regarded as an application of maximum

likelihood estimation (or maximum a posterior probability), where the likelihood probability is defined to be inversely proportional to the residual. The smaller the residual, the more probable the hypothesis. Thus, the decision reached using (6) is a hard one. It produces the final inference by the overall reconstruction error committed to each class. Unlike the hard decision rule, this paper proposes a novel decision framework for sparse signal modeling via DS theory. Samples that participate in the representation of the query are used to produce the amount of evidence. By pooling the pieces of evidence, the global evidence can be obtained, from which the inference can be drawn. Since the proposed method considers the uncertainty measurement, it is a soft decision.

Given the optimal representation $\hat{\alpha}$, whose entries are zeros except for those which partake in the recovery of the query, we select those samples with the nonzero weights to form a set $\Phi$. For any sample $\mathbf{x}_j \in \Phi$, the knowledge that $C_j = q$ can be viewed as a piece of evidence that enhances the belief that the query $\mathbf{y}$ belongs to the $q$-th class, where $C_j$ is the class membership of $\mathbf{x}_j$. However, it could not provide 100% certainty. In the formalism of DS theory, we could consider that only part of our belief is committed to the proposition that the query belongs to the $q$-th class. Since it does not refer to the other hypothesis, the rest of our belief is distributed to the local ambiguity $\Theta$. Thus, the focal elements are $\{C_j\}$ and $\Theta$. This item of evidence can therefore be formulated by

$$\begin{aligned} m^j(\{C_j\}) &= \eta \\ m^j(\Theta) &= 1-\eta \\ m^j(A) &= 0 \ \forall A \in 2^\Theta, A \notin \{\{C_j\}, \Theta\} \end{aligned} \tag{7}$$

for $0 \leq \eta \leq 1$. Since each representation coefficient is independent of another, they are separate and not connected. Each coefficient is not affected or influenced by the second, therefore the representation coefficients are statistically independent of the remaining, and hence can be used to generate BPA.

According to the above analysis, the smaller the reconstruction error $\mathbf{e}^j = \|\mathbf{y} - \mathbf{x}_j\alpha_j\|$, the more probable the query belongs to the $C_j$-th class, and hence the bigger $\eta$. On the contrary, the bigger $\mathbf{e}^j$, the less information $\mathbf{x}_j$ provides, and hence the smaller $\eta$. Thus, we postulate that $\eta$ should be a decreasing function of $\mathbf{e}^j$, $\eta = \eta_0 \cdot g(\mathbf{e}^j)$, where

$$g(0) = 1, \lim_{x\to\infty} g(x) = 0 \tag{8}$$

Note that we never assume the information provided from a single sample to be 100% certain, even though the reconstruction error is zero. Thus, the introduction of $\eta_0$ is necessary for $0 < \eta_0 < 1$. Obviously, there are a large number of functions verifying (8), and it is intractable to differentiate which one is better than the others. Inspired by the preceding works [9], this paper proposes to choose $g(\cdot)$ as

$$g(x) = e^{-\lambda\|x\|_1} \tag{9}$$

with $\lambda > 0$. For any sample $\mathbf{x}_j \in \Phi$, the amount of evidence, depending on the reconstruction error as well as its class membership, can be produced. Then, the remaining problem is how to pool these pieces of evidence. In this paper, it is achieved by two sequential steps, intra-class combination and interclass one.

*1) Intraclass combination:* According to the rule (3), for any two samples $\mathbf{x}_i, \mathbf{x}_j \in \Phi$ belonging to the $q$-th class (note by $L_q$), the BPA $m^{(i,j)} = m^i \bigoplus m^j$ deriving from the combination of $m^i$ and $m^j$ can be provided by

$$\begin{aligned} m^{(i,j)}(\{q\}) &= 1 - (1-\eta_0 \cdot g(\mathbf{e}^i)) \cdot (1-\eta_0 \cdot g(\mathbf{e}^j)) \\ m^{(i,j)}(\Theta) &= (1-\eta_0 \cdot g(\mathbf{e}^i)) \cdot (1-\eta_0 \cdot g(\mathbf{e}^j)) \end{aligned} \tag{10}$$

If denoted by $\Phi_q$ the set of samples from the $q$-th class that participate in the recovery of the query, the corresponding BPAs resulting from the combination of samples belonging $\Phi_q$ can be obtained by $m_q = \bigoplus_{\mathbf{x}_j \in \Phi_q} m^j$, as formulated by

$$\begin{aligned} m_q(\{L_q\}) &= 1 - \prod_{\mathbf{x}_j \in \Phi_q} (1-\eta_0 \cdot g(\mathbf{e}^j)) \\ m_q(\Theta) &= \prod_{\mathbf{x}_j \in \Phi_q} (1-\eta_0 \cdot g(\mathbf{e}^j)) \end{aligned} \tag{11}$$

*2) Interclass combination:* The next task is to generate the global BPA from the class-specific ones shown in (11). By combining all the BPAs $m_q(\cdot)$ for each class, a global BPA $m = \bigoplus_{q=1}^K m_q$ can be obtained,

$$\begin{aligned} m(\{L_q\}) &= \frac{m_q(\{L_q\}) \prod_{r\neq q} m_r(\Theta)}{1 - Conf} \\ m(\Theta) &= \frac{\prod_{q=1}^K m_q(\Theta)}{1 - Conf} \end{aligned} \tag{12}$$

where $Conf = \sum_{j=1}^K \prod_{q\neq j} m_q(\{L_q\})$ is the conflict measurement.

Since this paper only considers the global uncertainty proposition ($\Theta$), the belief and plausibility of the $q$-th class are

$$\begin{aligned} Bel(\{q\}) &= m(\{L_q\}) \\ Pl(\{q\}) &= m(\{L_q\}) + m(\Theta). \end{aligned} \tag{13}$$

Thus, the rules presented in Section II-A produce the same rank, and hence draw the same inference. The proposed framework for the combination of evidence is shown in TABLE I. As can be easily seen, the main difference between the hard decision and the soft decision consists in the quantification of the ambiguous states, *i.e.*, the sample-wise uncertainty $m^{j_i}(\Theta)$ and the class-wise one $m_j(\Theta)$.

TABLE I
THE GENERATION OF GLOBAL BPA ($\Theta = L_1 \cup L_2 \cup \cdots \cup L_K$)

| | $L_1$ | $L_2$ | $\cdots$ | $L_K$ | $\Theta$ |
|---|---|---|---|---|---|
| *Intraclass* | $m^{j_1}(L_1)\ldots$ $m^{j_1}(\Theta)\ldots$ | $m^{j_2}(L_2)\ldots$ $m^{j_2}(\Theta)\ldots$ | $\cdots$ | $m^{j_p}(L_K)\ldots$ $m^{j_p}(\Theta)\ldots$ | |
| *Interclass* | $m_1(L_1)$ | | | | $m_1(\Theta)$ |
| | | $m_2(L_2)$ | | | $m_2(\Theta)$ |
| | | | $\cdots$ | | |
| | | | | $m_K(L_K)$ | $m_K(\Theta)$ |
| *Combined* | $m(L_1)$ | $m(L_2)$ | $\cdots$ | $m(L_K)$ | $m(\Theta)$ |

### D. Numerical Illustration

To demonstrate the proposed decision rule, two numerical examples are provided in TABLE II, where the hard decision rule (6) is employed as the reference. There is a total of four target classes (BMP2, T72, BTR60, and T62), notated by $L_1$, $L_2$, $L_3$, and $L_4$. The two query samples are from $L_1$. The details on the database can be addressed to Section III.

TABLE II
CLASSIFICATION EXAMPLES

| | *Bayesian* | | *Dempster-Shafer* | | | |
|---|---|---|---|---|---|---|
| *Lab* | $\mathbf{e}^j$ | $\mathbf{e}_{SRC}$ | $m^j(C_j)$ | $m^j(\Theta_j)$ | $m(L_q)$ | $m(\Theta)$ |
| $L_1$ | 0.7419 | | 0.4524 | 0.5476 | | |
| $L_1$ | 0.7056 | | 0.4691 | 0.5309 | | |
| $L_1$ | 0.7629 | | 0.4430 | 0.5570 | | |
| $L_1$ | 0.7503 | 0.1707 | 0.4486 | 0.5514 | 0.9713 | |
| $L_1$ | 0.7334 | | 0.4563 | 0.5437 | | |
| $L_1$ | 0.7091 | | 0.4675 | 0.5325 | | 0.0081 |
| $L_1$ | 0.8030 | | 0.4256 | 0.5744 | | |
| $L_1$ | 0.7416 | | 0.4525 | 0.5475 | | |
| $L_2$ | 0.6973 | 0.7416 | 0.4730 | 0.5270 | 0.0069 | |
| $L_3$ | 0.6929 | 0.7566 | 0.4751 | 0.5249 | 0.0075 | |
| $L_4$ | 0.7873 | 0.9527 | 0.4323 | 0.5677 | 0.0062 | |
| $L_1$ | 0.6465 | | 0.4977 | 0.5023 | | |
| $L_1$ | 0.6763 | 0.4899 | 0.4831 | 0.5169 | 0.4333 | |
| $L_1$ | 0.6908 | | 0.4761 | 0.5239 | | |
| $L_1$ | 0.6835 | | 0.4796 | 0.5204 | | |
| $L_2$ | 0.6217 | 0.3745 | 0.5102 | 0.4898 | 0.1134 | 0.0330 |
| $L_2$ | 0.5654 | | 0.5397 | 0.4603 | | |
| $L_3$ | 0.7236 | | 0.4608 | 0.5392 | | |
| $L_3$ | 0.6853 | 0.6729 | 0.4787 | 0.5213 | 0.3922 | |
| $L_3$ | 0.6980 | | 0.4727 | 0.5273 | | |
| $L_3$ | 0.6907 | | 0.4762 | 0.5238 | | |
| $L_4$ | 0.7254 | 0.9732 | 0.4599 | 0.5401 | 0.0281 | |

In the first example, there is a total of 11 samples that partake in the representation of the query, and 8 samples of which are from $L_1$. The maximum likelihood probability (corresponding to the minimum reconstruction error 0.1707) is committed to the hypothesis $\mathbf{y} \in L_1$ by the hard decision, while the maximum of mass of belief 0.9713 is assigned to the proposition $\mathbf{y} \in L_1$ by the soft one. Though the two decision rules reach the correct inference, the confidence level (difference between the probability or belief for each class) for the hard decision rule is not as high as the soft one.

In the second example, there is also a total of 11 samples that participate in the recovery of the query, among which 4 samples are from $L_1$, and 4 samples are from $L_3$. Distinct inferences are generated using the two rules. The minimum reconstruction error (0.3745) is derived from $L_2$, while the maximum of mass of belief (0.4333) is produced by the proposition $\mathbf{y} \in L_1$, *i.e.*, the decision rule via Bayesian inference draws the incorrect inference ($\mathbf{y} \in L_2$), while the one via evidential reasoning achieves the correct one ($\mathbf{y} \in L_1$).

## III. EXPERIMENTS AND DISCUSSIONS

To verify the proposed rule (SRC$^{\mathbf{DS}}$), extensive experiments are conducted on MSTAR database, a gallery collected using a 10 GHz SAR sensor with $1ft \times 1ft$ resolution in azimuth and range. Images are captured at several depressions. The depression refers to angle between the line of sight from radar to target and the horizontal plane at the radar. Images are of around $128 \times 128$ pixels in size, and cropped to $80 \times 80$ pixels to exclude the background clutter. The parameters $\eta_0, \lambda$ are experimentally set as $\eta_0 = 0.95$ and $\lambda = 1$. The reference methods include SRC via Bayesian inference (SRC$^{\mathbf{BY}}$) [1], [2], kNN, and SVM.

### A. Configuration and Depression Variation

We first consider target recognition under depression and configuration variations. The configuration refers to physical difference, as summarized in TABLE III. Four vehicle targets are used, among which BMP2 and T72 consist of several variants with small structural modifications, *i.e.*, SN_9563, SN_9566, and SN_c21 for BMP2, SN_132, SN_812, and SN_s7 for T72. The standards (SN_9563 and SN_132) taken at an operating condition of 17° depression are used for training, while the remaining collected at an operating condition of 15° depression are used for testing. The details on the aspect views can be found in TABLE IV.

TABLE III
EXAMPLE TARGET VARIABILITY CATEGORIES.

| | |
|---|---|
| *Version Variant* | Smoke Grenade Launchers |
| *Configuration Variant* | Two Cables, Fuel Barrels |
| *Incidental Structural Modifications* | Broken Antenna Mount |

TABLE IV
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*)<br>196 (*SN_c21* ) | 195 (*SN_812*)<br>191 (*SN_s7* ) | 195 | 273 |

To make the method computationally tractable, images are downsampled by a factor of $\rho$. We choose $\rho$ from $\left\{\frac{1}{10}, \frac{1}{8}\right\}$, corresponding to 8×8 pixels and 10×10 pixels downsampled image. To make the comparison fair, the input of the references are chosen as same as the proposed method. TABLE V lists the accuracy of proposed method, as well as the references. Under the downsampled image of 8×8 pixels, the recognition rates for SRC$^{\mathbf{BY}}$ and SRC$^{\mathbf{DS}}$ are 0.8475 and 0.8604, compared to 0.8058 for SVM, 0.8122 for kNN. The similar performance can be obtained under the downsampled image of 10×10 pixels. The results prove that the classification via sparse representation technique could improve the recognition accuracy in comparison to the conventional skills. Under 8×8 downsampled image, the recognition rate for SRC$^{\mathbf{DS}}$ is 0.8604, compared to 0.8475 for SRC$^{\mathbf{BY}}$. The proposed rule is 1.29% better than the hard one. Correspondingly, the improvement of 1.04 percentage is achieved by soft decision in comparison to the hard one under the downsampled image of 10×10 pixels. The results demonstrate that the soft decision via evidential reasoning performs better in this specific experiment than the hard one via Bayesian inference.

### B. 10-Target Recognition

A more challenging problem, 10-target recognition, is considered subsequently. For BMP2 and T72, the standards taken

TABLE V
RECOGNITION RATES IN TWO DOWNSAMPLING FACTORS.

| *Dimension* | kNN | SVM | $\text{SRC}^{\mathbf{BY}}$ | $\text{SRC}^{\mathbf{DS}}$ |
|---|---|---|---|---|
| 8×8 | 0.8122 | 0.8058 | 0.8475 | 0.8604 |
| 10×10 | 0.8339 | 0.8379 | 0.8708 | 0.8812 |

at an operating condition of 17° are used to train the algorithms, while images collected at an operating condition of 15° are used for testing. The details on aspect views available for training and testing are listed in TABLE VI. The downsampling factor is chosen as $\frac{1}{8}$. The confusion matrices as well as the overall recognition rates obtained using the methods to be studied are given in Fig. 2.

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

| *Target* | BMP2 | BTR70 | T72 | BTR60 | 2S1 |
|---|---|---|---|---|---|
| 17° | 233 **(SN_9563)** | 233 | 232 **(SN_132)** | 256 | 299 |
| 15° *Test* | 195 (SN_9563) 196 (SN_9566) 196 (SN_c21) | 196 | 196 (SN_132) 195 (SN_812) 191 (SN_s7) | 195 | 274 |
| *Target* | BRDM2 | D7 | T62 | ZIL131 | ZSU23/4 |
| 17° | 298 | 299 | 299 | 299 | 299 |
| 15° | 274 | 274 | 273 | 274 | 274 |

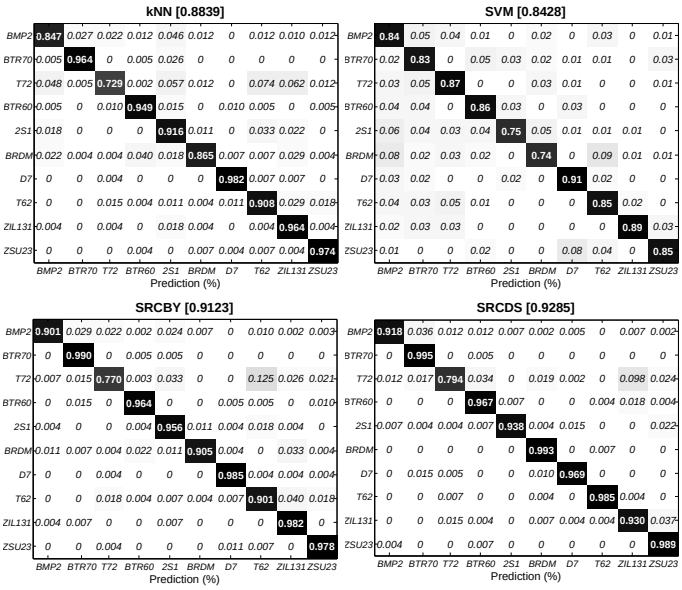


Fig. 2. The confusion matrices for the methods to be studied. In each subfigure, the row notes the ground-truth of the query, while the column gives the identity inferred by the algorithms. The diagonal entries show the recognition rate of each class, while the non-diagonal ones display the misclassification rate. The top entries in bracket are the overall recognition rates.

The experimental results shown in Fig. 2 are similar to the ones above. The classification accuracy obtained using sparse signal modeling is slightly better than the remaining algorithms. The recognition rate for $\text{SRC}^{\mathbf{BY}}$ is 0.9123, compared to 0.8428 for SVM, and 0.8839 for kNN, 6.95%, and 2.84% better than the reference methods. The overall recognition rate for the proposed soft decision is 0.9285, 1.62% better than the competitor, the hard decision.

## IV. CONCLUSION

This paper proposes a new decision rule for sparse signal modeling. Unlike the conventional method, where the decision is directly made by the overall reconstruction error associated with each class, it is dealt with DS evidence theory. The imprecision about the uncertainty measurement is modeled by introducing the sample-wise ambiguity and the class-wise one. Each sample that participates in the reconstruction of query is viewed as an item of evidence regarding the class membership of the query. The global evidence is obtained by pooling the pieces of evidence via Dempter's rule of combination. Several specific examples demonstrate that the proposed scheme better than the conventional rule. To make the proposed rule calculable, only the global ambiguous hypothesis is considered. The future researches will be paid to the local hypotheses, and the new rule of combination.

## REFERENCES

[1] 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.

[2] J. Thiagarajan, K. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representation for automatic target classification in SAR images," in *Int'l Sym. Communication, Control and Signal Processing*, Mar. 2010.

[3] H. Zhang, N. Nasrabadi, Y. Zhang, and T. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2481–2495, Jul. 2011.

[4] 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.

[5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

[6] Z. Liu, Q. Pan, and J. Dezert, "A new belief-based K-nearest neighbor classification method," *Pattern Recognition*, vol. 46, no. 3, pp. 834–844, 2013.

[7] Z. Liu, Q. Pan, J. Dezert, and G. Mercier, "Credal classification rule for uncertain data based on belief functions," *Pattern Recognition*, vol. 47, no. 7, pp. 2532–2541, 2014.

[8] T. Denoeux, O. Kanjanatarakul, and S. Sriboonchitta, "EK-NNclus: a clustering procedure based on the evidential K-nearest neighbor rule," *Knowledge-Based Systems*, vol. 88, pp. 57–69, 2015.

[9] T. Denœux, "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.

[10] 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, pp. 1018–1031, Jul. 1997.

[11] F. Tupin, I. Bloch, and H. Maitre, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1327–1343, May 1999.

[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] Z. Liu, J. Dezert, G. Mercier, and Q. Pan, "Dynamic evidential reasoning for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1955–1967, May 2012.

[14] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.

[15] G. Dong and G. Kuang, "Target recognition via information aggregation through Dempster-Shafer's evidence theory," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, 2015.

[16] I. Bloch, "Some aspect of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," vol. 17, pp. 905–916, 1996.